Journal of Statistical Software

April 2014, Volume 57, Code Snippet 2. *http://www.jstatsoft.org/*

DBKGrad: An R Package for Mortality Rates Graduation by Discrete Beta Kernel Techniques

Angelo Mazza
University of Catania

Antonio Punzo
University of Catania

Abstract

We introduce the R package **DBKGrad**, conceived to facilitate the use of kernel smoothing in graduating mortality rates. The package implements univariate and bivariate adaptive discrete beta kernel estimators. Discrete kernels have been preferred because, in this context, variables such as age, calendar year and duration, are pragmatically considered as discrete and the use of beta kernels is motivated since it reduces boundary bias. Furthermore, when data on exposures to the risk of death are available, the use of adaptive bandwidth, that may be selected by cross-validation, can provide additional benefits. To exemplify the use of the package, an application to Italian mortality rates, for different ages and calendar years, is presented.

Keywords: kernel smoothing, graduation, beta distribution, cross-validation.

1. Introduction

Mortality rates (or probabilities of dying) are indicators commonly used in demography and actuarial practice. They are usually referred to one or more variables, the most common being age, calendar years and duration; for ease of presentation, we will start focusing on age only, while bivariate extensions will be addressed afterward. To be specific, the d_x deaths at age x can be seen as arising from a population, initially exposed to the risk of death, of size e_x. This can be summarized via the model $d_x \sim \mathrm{Bin}\left(e_x, q_x\right)$, where q_x represents the true, but unknown, mortality rate at age x. The crude rate $\mathring{q}_x$ is the observed counterpart of q_x. Graduation is necessary because crude data usually presents abrupt changes, which do not agree with the dependence structure supposedly characterizing the true rates (London 1985). Nonparametric models are the natural choice if the aim is to reflect this belief. Furthermore, a nonparametric approach can be used to choose the simplest suitable parametric model, to provide a diagnostic check of a parametric model, or to simply explore the data (see Härdle

1992, Section 1.1, for a detailed discussion on the chief motivations that imply their use, and Debòn, Montes, and Sala 2006 for an exhaustive comparison of nonparametric methods in the graduation of mortality rates).

Due to its conceptual simplicity and practical and theoretical properties, kernel smoothing is one of the most popular statistical methods for nonparametric graduation. Among the various alternatives existing in literature (see Copas and Haberman 1983, Bloomfield and Haberman 1987, Gavin, Haberman, and Verrall 1993, 1994, 1995 and Peristera and Kostaki 2005), the attention is focused on the discrete beta kernel estimator proposed by Mazza and Punzo (2011). 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 $\mathcal{X} = \{0, 1, \ldots, \omega\}$, ω being the highest age of interest. Discretization could also come handy to actuaries who have to produce "discrete" graduated mortality tables. Discrete beta kernels are considered to overcome the problem of boundary bias, commonly arising from the use of symmetric kernels (Chen 2000); their support $\mathcal{X}$, in fact, matches the age range and this, when smoothing is made near the boundaries, allows avoiding allocation of weight outside the support (for example negative or unrealistically high ages). Adaptive variants, based on the reliability of the data at each age, are introduced in Mazza and Punzo (2013a,b).

In this paper we present the R (R Core Team 2013) package **DBKGrad**, available from the Comprehensive R Archive Network at `http://CRAN.R-project.org/package=DBKGrad`, which implements all the methods described above and offers some related functionalities like bivariate graduation and diagnostic checks. Although R is well-provided with functions performing kernel smoothing techniques (see, e.g., Hayfield and Racine 2008), there aren't any offering discrete beta kernel smoothing or kernel smoothing in the field of graduation. However, there are packages that allow for nonparametric graduation using spline-based methods. For example, the **demography** package (Hyndman, Booth, Tickle, and Maindonald 2014) does partially monotonic penalized spline smoothing for mortality rates and other demographic indicators. Furthermore, the **MortalitySmooth** package of Camarda (2012) treats mortality data as Poisson counts and smooths using P-splines; similarly to **DBKGrad**, this package also allows for bivariate graduation.

The paper is organized as follows. Section 2 retraces the fixed discrete beta kernel estimator. Its adaptive variants are recalled in Section 3 while some cross-validation approaches for the selection of both the fixed and the adaptive bandwidth are discussed in Section 4. Further related aspects, such as the adoption of a preliminary transformation of the rates and the extension of the estimator to the bivariate case, are given in Section 5. The relevance of the **DBKGrad** package is shown, via a real dataset included in the package, in Section 6, and conclusions are finally given in Section 7.

2. Discrete beta kernel graduation

Given the crude rates $\mathring{q}_y$, for each age (at last birthday) $y \in \mathcal{X}$, the Nadaraya-Watson kernel estimator of the true but unknown mortality rate q_x, at the evaluation age x, is

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

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 (fixed) *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). In (1), the discrete beta kernels (Mazza and Punzo 2011)

$$k_h(x;m) = \left(x + \frac{1}{2}\right)^{\frac{m+\frac{1}{2}}{h(\omega+1)}} \left(\omega + \frac{1}{2} - x\right)^{\frac{\omega+\frac{1}{2}-m}{h(\omega+1)}}, \quad x \in \mathcal{X}, \tag{2}$$

are adopted. Their normalized version,

$$K_h(x;m) = \frac{k_h(x;m)}{\sum\limits_{y\in\mathcal{X}} k_h(y;m)}, \quad x \in \mathcal{X},$$

corresponds to the discrete beta probability mass function defined in Punzo and Zini (2012) and parameterized, as in Punzo (2010, see also Bagnato and Punzo 2012b), according to the mode m and another parameter h that is closely related to the distribution variability. In particular, for $h \to 0^+$, $K_h(x;m)$ tends to a Dirac delta function in $x = m$, while for $h \to \infty$, $K_h(x;m)$ tends to a discrete uniform distribution. Thus h can be considered as the bandwidth of the estimator (1).

Roughly speaking, discrete beta kernels possess two peculiar characteristics. Firstly, their shape, for 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); see also Chen (2000) to find out more on the properties of the discrete beta kernel estimator in its continuous counterpart.

3. Making the bandwidth adaptive

Rather than restricting h to a fixed value, a more flexible approach is to allow the bandwidth to vary according to the reliability of the data measured in a convenient way. Thus, for ages in which the reliability is relatively large, a lower value for h results in an estimate that more closely reflects the crude rates. For ages in which the reliability is smaller, such as at old ages, a higher value for h allows the estimate of the true mortality rates to progress more smoothly; this means that at older ages we are calculating local averages over a greater number of observations. This technique is often referred to as a variable or *adaptive (bandwidth) kernel estimator*.

As well-documented in Gavin *et al.* (1995):

(a) a different bandwidth, say h_x, can be used for each evaluation age x at which the rates are estimated;

(b) a different bandwidth, say h_y, can be used for each age y;

(ab) a different bandwidth, say $h_{x,y}$, can be selected for each evaluation point x and for each age $y \in \mathcal{X}$.

According to Gavin *et al.* (1995), the adaptive bandwidth could have the following multiplicative formulation

$$\text{(a) } h_x(s) = hl_x^s \qquad \text{(b) } h_y(s) = hl_y^s \qquad \text{(ab) } h_{x,y}(s) = hl_{x,y}^s \qquad x, y \in \mathcal{X}, \tag{3}$$

where h is the *global bandwidth*, l_x (or l_y or $l_{x,y}$) is the *local factor*, and $s \in [0,1]$ is the *sensitivity parameter* inserted as a power of the local factor only. Reliability decides the shape of the local factors, while s is necessary to dampen the possible extreme variations in reliability that can arise between young and old ages. Obviously, the case $s = 0$ results in a fixed bandwidth estimator. By applying to model (1) formulation (a), we have

$$\widehat{q}_x = \sum_{y\in\mathcal{X}} \frac{k_{h_x}(y; m = x)}{\sum_{j\in\mathcal{X}} k_{h_x}(j; m = x)} \mathring{q}_y = \sum_{y\in\mathcal{X}} K_{h_x}(y; m = x)\, \mathring{q}_y, \quad x \in \mathcal{X}, \tag{4}$$

where the notation h_x is used to abbreviate $h_x(s)$.

As concerns (a), but similar reasoning also holds for (b), Mazza and Punzo (2013a) consider the reliability a function only of the amount of exposure, according to the formulation

$$l_x = \frac{e_x^{-1}}{\max_{y\in\mathcal{X}} \left\{e_y^{-1}\right\}}, \qquad x \in \mathcal{X}. \tag{5}$$

According to the model $d_x \sim \text{Bin}(e_x, \mathring{q}_x)$, where $\mathring{q}_x$ is the maximum likelihood estimate of q_x, a natural index of reliability is represented by the reciprocal of a relative measure of variability. As relative measure of variability, Mazza and Punzo (2013b) adopt the variation coefficient (VC) which, in this context, can be computed as

$$\text{VC}_x = \frac{\sqrt{e_x \mathring{q}_x (1 - \mathring{q}_x)}}{e_x \mathring{q}_x}, \qquad x \in \mathcal{X},$$

and it is normalized, so that $l_x^s \in [0,1]$, according to the formulation

$$l_x = \frac{\text{VC}_x}{\sum_{y\in\mathcal{X}} \text{VC}_y}, \qquad x \in \mathcal{X}. \tag{6}$$

Finally, as concerns (ab), one way of defining the local factor is $l_{x,y} = l_y / l_x$, $x, y \in \mathcal{X}$.

4. The choice of h and s

Choosing h by trial and error is informative, but it is also convenient to have an objective selection method. Among data-driven methods for bandwidth selection, cross-validation is the simplest and most commonly used; for an application in graduation, see Gavin *et al.* (1995). In detail, h is obtained by minimizing the cross-validation statistic

$$\text{CV}(h) = \sum_{x\in\mathcal{X}} r^2 \left(\mathring{q}_x, \widehat{q}_x^{(-x)}\right), \tag{7}$$

where $r\left(\mathring{q}_x, \widehat{q}_x^{(-x)}\right)$ denotes the residual (at age x) and

$$\widehat{q}_x^{(-x)} = \sum_{\substack{y \in \mathcal{X} \\ y \neq x}} \frac{K_h\left(y; m = x\right)}{\sum_{\substack{j \in \mathcal{X} \\ j \neq x}} K_h\left(j; m = x\right)} \mathring{q}_y$$

is the estimated value at age x computed by removing the crude rate $\mathring{q}_x$ at that age. As residuals, Mazza and Punzo (2011, 2013a) consider the classical residuals

$$r\left(\mathring{q}_x, \widehat{q}_x^{(-x)}\right) = \widehat{q}_x^{(-x)} - \mathring{q}_x, \tag{8}$$

while Mazza and Punzo (2013b), because of the high differences in mortality rates among ages, adopt the proportional differences

$$r\left(\mathring{q}_x, \widehat{q}_x^{(-x)}\right) = \frac{\widehat{q}_x^{(-x)}}{\mathring{q}_x} - 1 \tag{9}$$

which are commonly used in the graduation literature, since we want the mean relative square error to be low (see Heligman and Pollard 1980).

Similarly, in the adaptive frame, we minimize the two-dimensional cross-validation statistic $\mathrm{CV}\left(h, s\right)$. Alternatively, it is possible to select subjectively one of the two parameters and let cross-validation select the other. In literature, actually, s is often chosen subjectively (see Gavin *et al.* 1995 and Mazza and Punzo 2011, 2013a,b) and cross-validation is used to select h by minimizing the conditional cross-validation statistic $\mathrm{CV}\left(h|s\right)$.

In **DBKGrad**, the Levenberg-Marquardt algorithm (Moré 1978) in the **minpack.lm** package (Elzhov, Mullen, Spiess, and Bolker 2013) is used to minimize the cross-validation statistic.

5. Further aspects

5.1. Transforming mortality rates

Before applying any model, it is always worth considering a transformation of the data into a more tractable form, that better reflects the strengths of the model or that more clearly reveals the structure of the data. In parametric graduation, for example, it may be easier to transform the rates and work with a linear model than to graduate the crude rates using a more mathematically demanding nonlinear model. The same philosophy applies in nonparametric graduation. The **DBKGrad** package allows for log, logit, and Gompertz transformations and, once the transformed data are graduated, a back-transformation is applied. However, because the choice of a transformation remains subjective, and the relative success of a particular transformation seems to depend on the data set (Gavin *et al.* 1995), no transformation is applied by default in the **DBKGrad** package.

5.2. Pointwise confidence intervals and simultaneous confidence bands

In visual inspection and graphical interpretation of the estimated sequence of points, we may either be interested in $\widehat{q}_x$, evaluated at a specific age x, or we may be interested in the whole sequence of points $\widehat{q}_x$, $x \in \mathcal{X}$. The first case corresponds to pointwise confidence intervals, while the second case requires the construction of simultaneous confidence bands (see Härdle 1992, for details). Both provide relevant information because they indicate the extent to which the estimates are well defined. Moreover, they are useful when nonparametric and parametric models are compared.

Since $\widehat{q}_x$ is a linear function of the mortality rates, as can be easily seen from (1) and (4), and being $d_x \sim \text{Bin}\left(e_x, q_x\right)$

$$\mathsf{VAR}\left(\widehat{q}_x\right) = \sum_{y \in \mathcal{X}} \left[K_{h_x}\left(y; m = x\right)\right]^2 \frac{q_y\left(1 - q_y\right)}{e_y}.$$

The above formula holds if independence of the d_ys is assumed and requires the knowledge of the number e_y of exposed to risk at each age. Substituting q_y with $\widehat{q}_y$ yields the $(1-\alpha) \cdot 100\%$ pointwise confidence intervals

$$\widehat{q}_x \mp z_{1-\frac{\alpha}{2}} \sqrt{\sum_{y \in \mathcal{X}} \left[K_{h_x}\left(y; m = x\right)\right]^2 \frac{\widehat{q}_y\left(1 - \widehat{q}_y\right)}{e_y}}, \tag{10}$$

where $z_{1-\frac{\alpha}{2}}$ is such that $\Phi\left[z_{1-\frac{\alpha}{2}}\right] = 1 - \frac{\alpha}{2}$.

Pointwise confidence intervals, along with a correction of α, can be used to construct simultaneous confidence bands. The correction of Bonferroni, $\alpha / \left(\omega + 1\right)$, or that of Šidák (1967), $1 - \left(1 - \alpha\right)^{1/(\omega+1)}$, are common in this context.

5.3. Diagnostic checks

After graduating the crude rates, a common practice consists in analyzing the residuals behavior. **DBKGrad** method `residuals()` returns, according to argument `restype`, the types of residual listed in Table 1. Note that, types `"response"`, `"pearson"`, and `"deviance"`, can be only computed when the exposures are available. In particular, the last two (`"pearson"`, and `"deviance"`) are based on the binomial model.

The **DBKGrad** package also allows to graphically investigate the residuals dependence structure through the autocorrelogram, as implemented by the `acf()` function of the **TSA** package, and the autodependogram of Bagnato, Punzo, and Nicolis (2012), as implemented by the `ADF()` function of the **SDD** package (see also Bagnato and Punzo 2012a, 2013). The autodependogram looks like the autocorrelogram with the difference that the aucorrelations for each lag are substituted by the χ^2 statistics of (linear/nonlinear) dependence; in order to show possible "problematic" lags, a critical line is superimposed (for further details and developments on this diagram see, e.g., Bagnato and Punzo 2010 and Bagnato, De Capitani, and Punzo 2014, 2013a,b).

5.4. Bivariate graduation

In many cases of practical interest, mortality patterns are referred to both age and another variable Y. Demographers, for instance, may be interested in mortality patterns over age and calendar years. Another example are select and ultimate mortality problems, in which mortality of insured individuals is evaluated over age and duration from the last check by an

restype	Formulation
"working"	$\mathring{q}_x - \widehat{q}_x$
"proportional"	$\dfrac{\mathring{q}_x}{\widehat{q}_x} - 1$
"response"	$e_x\mathring{q}_x - e_x\widehat{q}_x$
"pearson"	$\dfrac{e_x\mathring{q}_x - e_x\widehat{q}_x}{\sqrt{e_x\widehat{q}_x\left(1-\widehat{q}_x\right)}}$
"deviance"	$\mathrm{sign}\left(\mathring{q}_x - \widehat{q}_x\right)\sqrt{2e_x\mathring{q}_x \ln\left(\dfrac{\mathring{q}_x}{\widehat{q}_x}\right) + 2e_x\left(1-\mathring{q}_x\right)\ln\left(\dfrac{1-\mathring{q}_x}{1-\widehat{q}_x}\right)}$

Table 1: Options available for the argument `restype` of the function `residuals()`

insurance company. The **DBKGrad** package allows for bivariate discrete beta kernel graduation; in detail, the estimated mortality rate $\widehat{q}_{x,y}$, $(x, y) \in \mathcal{X} \times \mathcal{Y}$, is obtained by using the following product discrete beta kernel estimator

$$\widehat{q}_{x,y} = \sum_{u\in\mathcal{X}}\sum_{v\in\mathcal{Y}} K_{h_X}\left(u; m = x\right) K_{h_Y}\left(v; m = y\right) \mathring{q}_{u,v}, \qquad (x, y) \in \mathcal{X} \times \mathcal{Y}, \tag{11}$$

where h_X and h_Y are the bandwidths referred to X and Y, respectively. Using this formulation, all the features described above, like the adaptive bandwidth, can be straightforwardly generalized to the bivariate case.

6. Package description and illustrative examples

In this section we provide a description of the main capabilities of **DBKGrad** along with illustrations via a real data set contained in the package.

6.1. Package description

Package **DBKGrad** is developed in an object-oriented design, using the standard S3 paradigm. Its main function, `dbkGrad()`, graduates one- and two-dimensional mortality rates using discrete beta kernel estimators; its arguments are listed in Table 2 and it returns a `dbkGrad` class object. The `plot()` method allows for a variety of exploratory plots; the type of the plot is selected through the argument `plottype` (see Table 3), while its appearance is governed by the `plotstyle` argument. Its options are: `"mat"` (default in univariate analyses) for classical plots, `"level"` for level plots (default in bivariate analyses), and `"persp"` for 3D plots. When plotting bivariate data, if `byage = TRUE` (default) the variable in rows (typically age) is displayed in abscissa and the variable in columns in ordinate while, if `byage = FALSE`, then axes are swapped.

Arguments	Description
`obsq`	Crude mortality rates with respect to one or two variables
`limx` (`limy`)	Interval of rows (columns) of `obsqx` to be considered in graduation
`ex`	Exposure with respect to the same variables of `obsqx`
`transformation`	Preliminary transformation to be applied. Options are: `"log"`, `"logit"`, `"Gompertz"`, and `"none"` to avoid transformations
`bwtypex` (`bwtypey`)	Type of bandwidth to be adopted by row (by column): `"FX"` for the fixed bandwidth, `"VC"` and `"EX"` for the adaptive bandwidth based on the local factors in (5) and (6), respectively
`adaptx` (`adapty`)	Type of adaptive bandwidth to be adopted by row (by column): `"a"` for type (a), `"b"` for type (b), and `"ab"` for (ab) of Section 3
`hx` (`hy`)	The global bandwidth used for the variable on the rows (columns)
`sx` (`sy`)	The sensitive parameter used for the variable on the rows (columns)
`cvres`	The type of residuals to be minimized for cross-validation: `"res"` for those in (8), and `"propres"` for those in (9)
`cvhx` (`cvhy`)	If `TRUE`, the global bandwidth for the variable on the rows (columns) is selected by cross-validation
`cvsx` (`cvsy`)	If `TRUE`, the sensitivity parameter for the variable on the rows (columns) is selected by cross-validation
`alpha`	Value of α used for pointwise confidence intervals and simultaneous confidence bands

Table 2: Arguments of the function `dbkGrad()`

Option	Description
`"observed"`	Plots observed mortality rates
`"fitted"`	Plots fitted mortality rates
`"obsfit"`	Plots observed and fitted mortality rates
`"exposure"`	Displays the number of exposures
`"residuals"`	Displays some plots related to residuals: density of residuals, residuals *versus* fitted values, and residuals *versus* the discrete variable of interest
`"checksd"`	Displays autocorrelogram and autodependogram of residuals (only for the unidimensional case)

Table 3: Options for argument `plottype` of the `plot()` method

6.2. The ItalyM dataset

This tutorial uses dataset `ItalyM` included in the **DBKGrad** package. Data come from the Human Mortality Database (2013) and consist of probabilities of dying and annual (January 1st) population (ages from 0 to 95 and years from 1906 to 2009, for a total of 104 years) for the Italian males. Data are loaded with

```
R> library("DBKGrad")
R> data("ItalyM")
```

The command `data("ItalyM")` loads two (ages $\times$ years) matrices, `obsq` and `population`, of observed probabilities of dying and annual population, respectively.

Univariate analysis

The first example is an unidimensional analysis referred to the year 2009. The following command

```
R> res1 <- dbkGrad(obsq = obsq, limx = c(6,71), limy = 104,
+    exposure = population, bwtypex = "VC", cvhx = TRUE, adaptx = "ab")

It.  0,  RSS =  2.85756,  Par. =  0.002        0.002   0.2   0.2
It.  1,  RSS =  2.82462,  Par. =  0.00291382   0.002   0.2   0.2
It.  2,  RSS =  2.81778,  Par. =  0.00363656   0.002   0.2   0.2
It.  3,  RSS =  2.8175,   Par. =  0.0034098    0.002   0.2   0.2
It.  4,  RSS =  2.81738,  Par. =  0.00353524   0.002   0.2   0.2
It.  5,  RSS =  2.81735,  Par. =  0.00347148   0.002   0.2   0.2
It.  6,  RSS =  2.81735,  Par. =  0.00350567   0.002   0.2   0.2
It.  7,  RSS =  2.81734,  Par. =  0.00348779   0.002   0.2   0.2
It.  8,  RSS =  2.81734,  Par. =  0.00349727   0.002   0.2   0.2
It.  9,  RSS =  2.81734,  Par. =  0.00349228   0.002   0.2   0.2
It. 10,  RSS =  2.81734,  Par. =  0.00349492   0.002   0.2   0.2
It. 11,  RSS =  2.81734,  Par. =  0.00349352   0.002   0.2   0.2
```

performs the discrete beta kernel graduation and provides an object of class `dbkGrad`. Argument `limy = 104` is used in order to select the last column in `obsq` and `exposure`, which corresponds to year 2009. In the same way, argument `limx = c(6, 71)` limits the graduation only to ages in $\{5, \ldots, 70\}$. Adaptive bandwidth, with local factor in (6), is selected using `bwtypex = "VC"`. As regards the type of adaptation, argument `adaptx = "ab"` applies formulation (ab) of Section 3. Since `cvhx = TRUE` and by default `cvsx = FALSE`, only `hx` is selected by cross-validation; note that, because by default `cvres = "propres"`, the proportional residuals (9) are used. Iterations from the cross-validation procedure are printed at video, with the last four columns showing the values of `hx`, `hy`, `sx`, and `sy`. As mentioned, only `hx` is being estimated, while `sx` is stuck to its default value and `hy` and `sy` are not used in unidimensional analysis. For more details about the choice of the default parameters values, see Mazza and Punzo (2013a). The bandwidth type is adaptive and it is based on the local factor in (6), since we put `bwtypex = "VC"`; also, being by default `transformation = "none"`, no preliminary transformation is applied.

We can compare observed and fitted probabilities of dying via the command

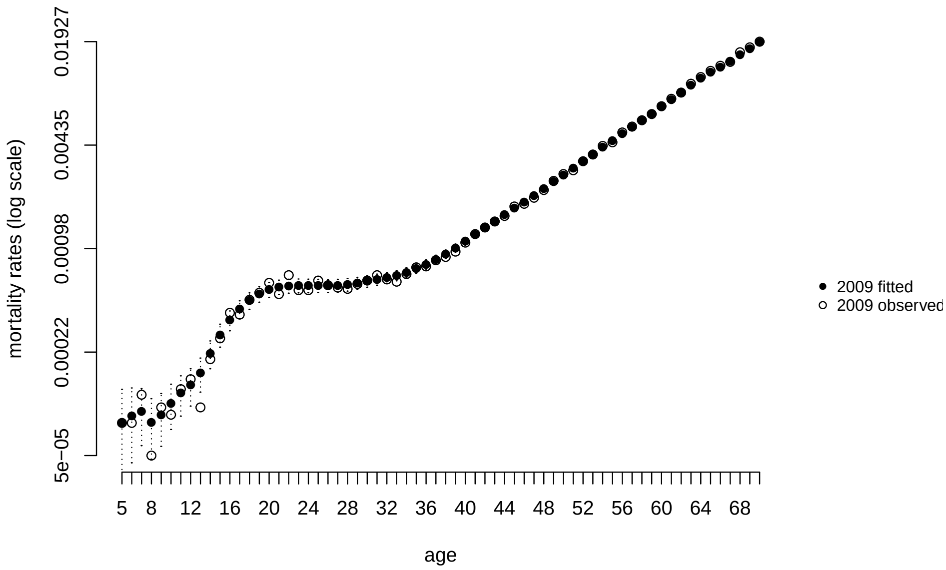


Figure 1: Observed and graduated male mortality rates, in logarithmic scale, of the `ItalyM` dataset (year = 2009). Graduation is made by the adaptive discrete beta kernel estimator where the bandwidth is estimated by minimizing the cross-validation statistic (7) with residuals defined by (9). Bonferroni's 95% confidence bands are also displayed.

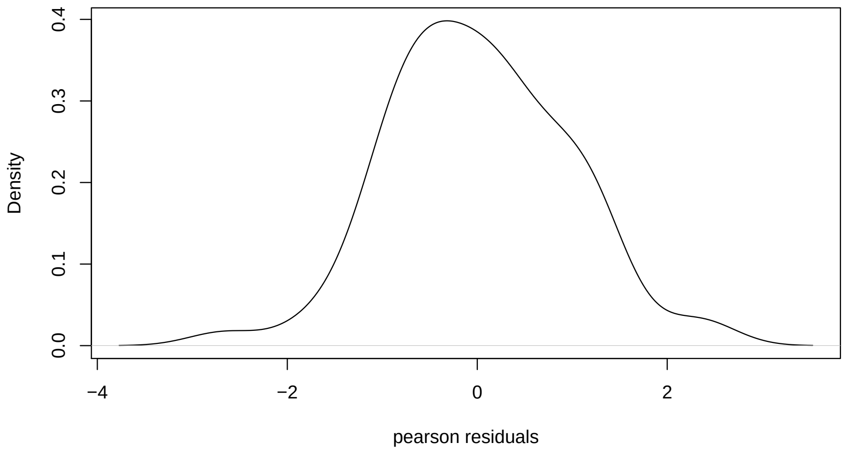
```
R> plot(res1, plottype = "obsfit", CI = FALSE, CBBonf = TRUE)
```

that produces the plot in Figure 1. Bonferroni's 95% confidence bands are also displayed (`CBBonf = TRUE`). As usual in the graduation literature, a logarithmic scale is used for the mortality rates. It is possible to note a small but prominent hump, peaking around 19 years of age. This excess mortality is known in literature as "accidental hump"; risk-taking and surplus mortality are signatures of the male human's early adult years (Heligman and Pollard 1980), and broadly coincides with a peak in male hormone production (Parkes 1976 and Goldstein 2011). The main causes of death at these ages are accidents, violence, and disease (Preston 1976). Although the statistical influence of the accident hump on survival and life expectancy is small, on a logarithmic-scale the hump is visible relative to the low mortality typical of late adolescence and early adulthood.

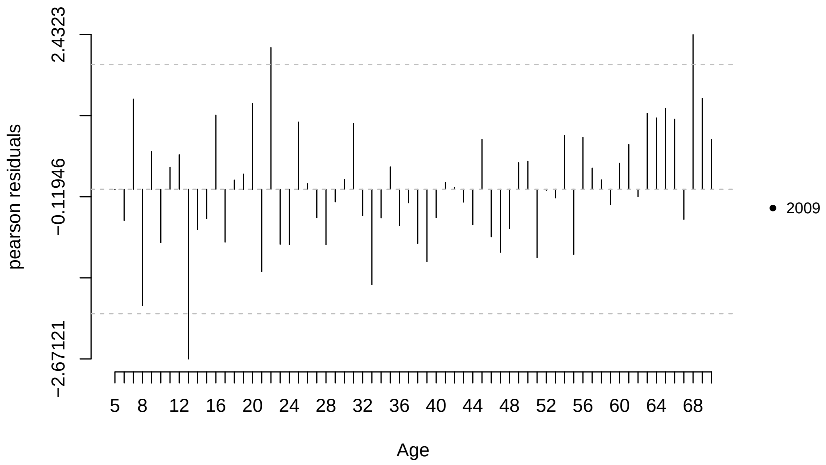
Once the model is fitted, we can conduct a residuals analysis. To begin, we can consider the command

```
R> plot(res1, plottype = "residuals", restype = "pearson")
```

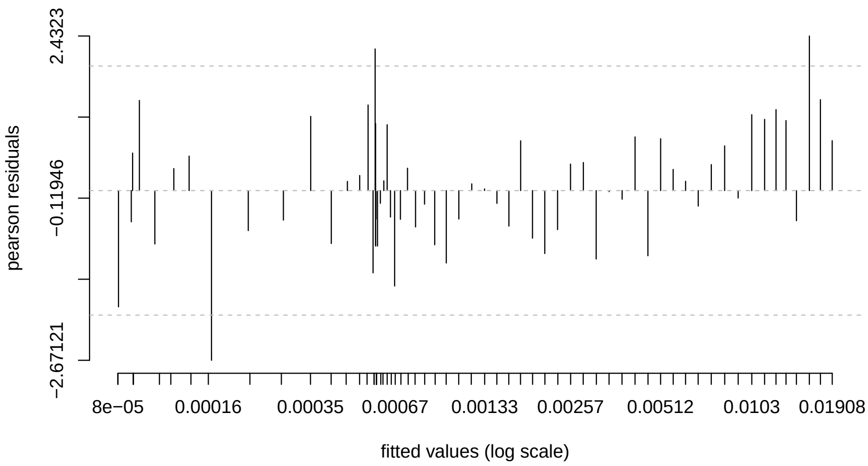
which generates the plots in Figure 2. They refer to the Pearson residuals (`restype = "pearson"`). The Gaussian kernel density of the residuals in Figure 2(a), obtained by using the `density()` function of the **stats** package, is bell-shaped around zero. Furthermore, residuals *versus* age in Figure 2(b), and residuals *versus* fits in Figure 2(c), move around the horizontal zero line and do not show any particular systematic feature. To investigate the dependence structure of the residuals, the autocorrelogram and the autodependogram in Figure 3 are obtained with



(a) density

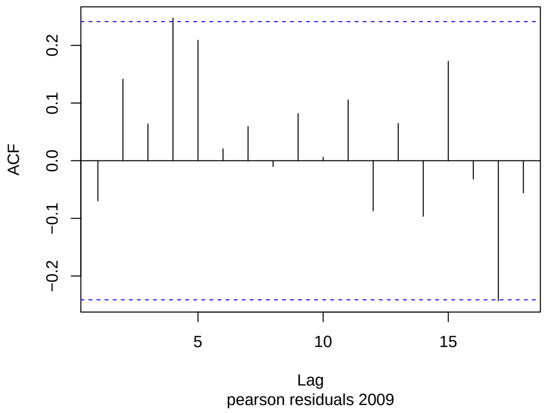


(b) residuals *versus* age

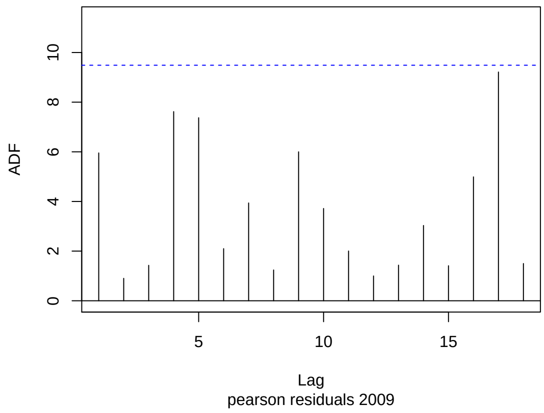


(c) residuals *versus* fits

Figure 2: Plot of residuals from the fitted model (`ItalyM` dataset, year 2009).



(a) autocorrelogram



(b) autodependogram

Figure 3: Serial dependence diagrams on the residuals from the fitted model (`ItalyM` dataset, year 2009).

```
R> plot(res1, plottype = "checksd", restype = "pearson")
```

Both plots show that the linear/nonlinear dependence structure of the data is captured by the fitted model. Note that, for further analyses, residuals can be extracted with

```
R> residuals(res1, type = "pearson")
```

```
           2009
5   -0.008221649
6   -0.491638996
7    1.418005014
.    .
.    .
.    .
68   2.432295179
69   1.430594665
70   0.785573483
```

Bivariate analysis

In the second example, graduation is performed over ages and years, via the command

```
R> res2 <- dbkGrad(obsq = obsq, limx = c(6, 46), limy = c(60, 104),
+    exposure = population, cvres = "res", transformation = "logit",
+    bwtypex = "VC", bwtypey = "EX", cvhx = TRUE, cvhy = TRUE,
+    cvsx = TRUE, cvsy = TRUE, adaptx = "ab", adapty = "b")
```

```
It.  0,  RSS = 17.0044,  Par. =  0.002       0.002       0.2       0.2
It.  1,  RSS = 16.5043,  Par. =  0.00344181  0.00281997  0.428614  0
It.  2,  RSS = 16.435,   Par. =  0.0033923   0.00304048  0.378265  0
  .   .    .    . .        .    .  .           .           .         .
  .   .    .    . .        .    .  .           .           .         .
  .   .    .    . .        .    .  .           .           .         .
It. 25,  RSS = 16.2223,  Par. =  0.00301489  0.00404372  0.346014  0
It. 26,  RSS = 16.2223,  Par. =  0.00301489  0.00404372  0.346014  0
It. 27,  RSS = 16.2223,  Par. =  0.00301489  0.00404372  0.346014  0
```

Graduation is limited to ages from 5 to 45 (`limx = c(6, 46)`) and to years from 1965 to 2009 (`limy = c(60, 104)`); a preliminary logit transformation of the data (`transformation = "logit"`) is applied. An adaptive bandwidth is used for both variables; the local factors in (6) are used for age (`bwtypex = "VC"`) and those in (5) for years (`bwtypey = "EX"`). The type of adaptation is according to formulation (ab) of Section 3 (`adaptx = "ab"`) for age, while formulation (b) is used (`adapty = "b"`) for year. Cross-validation with residuals in (8) is used (`cvres = "res"`) for selecting global bandwidths (`cvhx = TRUE` and `cvhy = TRUE`) and sensitivity parameters (`cvsx = TRUE` and `cvsy = TRUE`). From the cross-validation iterations printed at video, we can see that the sensitivity parameter for variable year is zero; this means that in this case the exposures-based local factors are not useful.

In the bivariate case, plotting the `dbkGrad` object produces by default an image plot (Figure 4) of both observed and graduated mortality rates

```
R> plot(res2, plottype = "obsfit")
```

The plot shows how graduated data retain the important aspects coming from the changes of

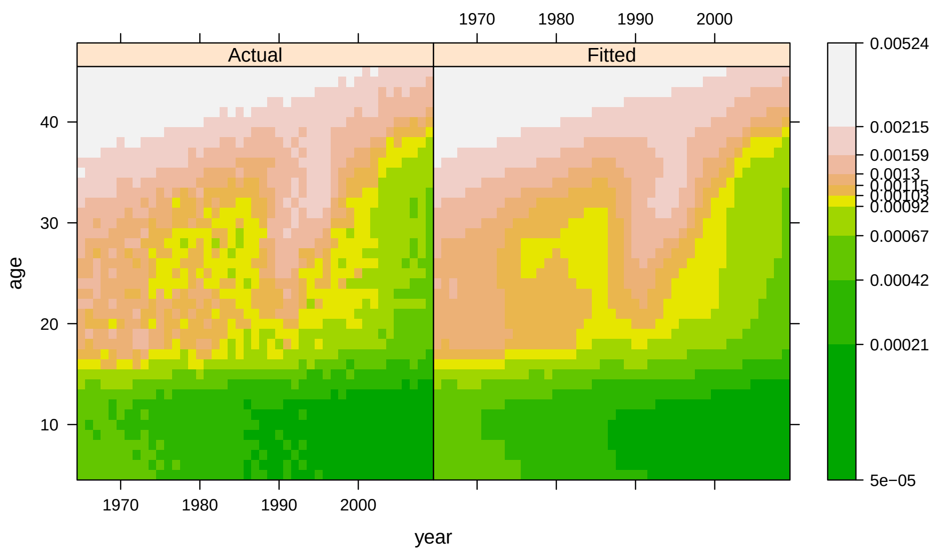
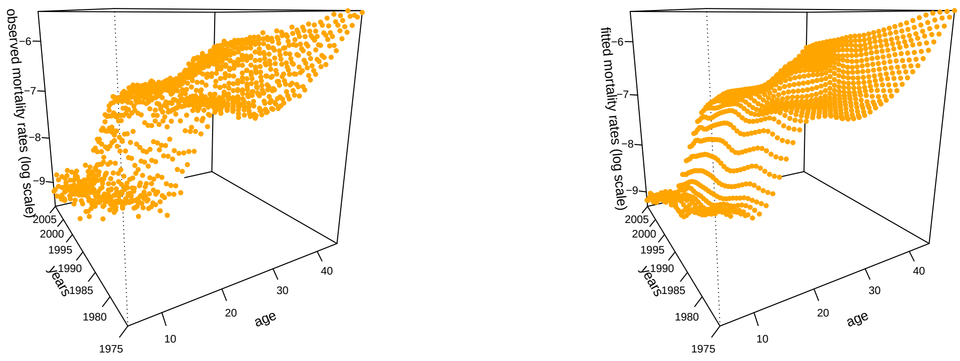


Figure 4: Observed and graduated bivariate mortality rates, in logarithmic scale (`ItalyM` dataset, ages 5–45, years 1965–2009). Global bandwidths and sensitivity parameters are selected, after a preliminary logit transformation, by cross-validation with residuals in (8).



(a) observed (b) graduated

Figure 5: Bivariate mortality rates, in logarithmic scale (`ItalyM` dataset, ages 5–45, years 1975–2005). Global bandwidths and sensitivity parameters are selected, after a preliminary logit transformation, by cross-validation with residuals in (8).

the mortality rates and, at the same time, leave out random noise. A different "perspective" of the same results can be obtained via the command

```
R> plot(res2, plottype = "obsfit", plotstyle = "persp", columns = 11:41,
 +       col = "orange")
```

which produces the 3D plots in Figure 5. In order to improve the legibility of the plot, we have focused the attention on the years from 1975 to 2005 via the specification `columns = 11:41`.

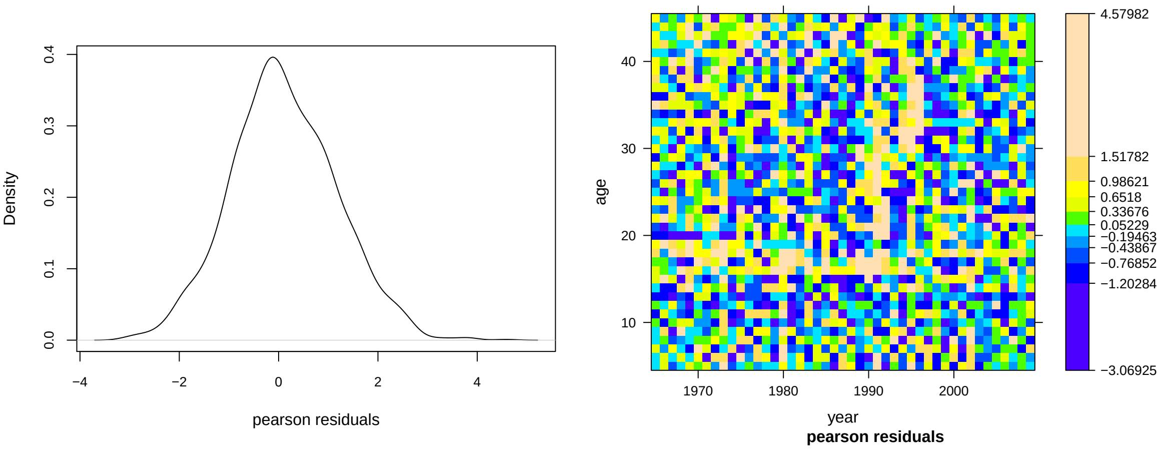
To investigate residuals, the command

```
R> plot(res2, plottype = "residuals", restype = "pearson", palette = "topo.colors")
```

produces the two plots in Figure 6 using the color palette `topo.colors`. For other color palettes see `?topo.colors`. Note that, the Gaussian kernel density of the Pearson residuals, in Figure 6(a), is obtained by considering all the available residuals simultaneously. From this density we can see that, as desired, the residuals are bell-shaped around zero. With regard to the image plot in Figure 6(b), apart from a ridge running from the point of coordinates (1970,20) to (1990,15) and then up to (1995,40), no further particular systematic features are observed.

7. Conclusions

In this paper we present **DBKGrad**, an R package for graduating mortality rates using discrete beta kernels. It may be used for smoothing probabilities defined on a univariate/bivariate finite domain, but the emphasis here is on probabilities of dying. The package — thanks to



(a) density

(b) image plot

Figure 6: Pearson residuals from bivariate graduation (`ItalyM` dataset, ages 5–45, years 1965–2009). Bandwidths and sensitivity parameters are selected, after a preliminary logit transformation, by cross-validation with residuals in (8).

the use of a discrete approach — is meant to be a user-friendly tool for demographers and actuaries who deal with (discrete) life tables; flexibility is achieved by providing the user with many options. For example, he/she may choose among fixed and adaptive bandwidths, these latter being based on three different formulations and each allowing two different ways of incorporating the exposed to the risk of dying. The global bandwidth and/or a dampening factor may be indicated by the user or chosen by cross-validation; the cross-validation score being minimized may be based on the traditional sum of squared residuals or on an alternative formulation used in the graduation literature, that is the sum of squared proportional residuals. Several preliminary data transformations, different diagnostic checks of residuals, and both pointwise confidence intervals and simultaneous confidence bands, are provided. Given the importance of graphical analysis in the nonparametric context, several plots and diagrams can be easily produced, with different styles for univariate and bivariate datasets. We believe that the **DBKGrad** package may prove useful either as a modeling tool or, if parametric models are to be used, to carry out a diagnosis of parametric models or simply to examine data.

References

Bagnato L, De Capitani L, Punzo A (2013a). "Improving the Autodependogram Using the Kulback-Leibler Divergence." *Technical Report 1306.5006*, arXiv.org e-print. URL `http://arxiv.org/abs/1306.5006`.

Bagnato L, De Capitani L, Punzo A (2013b). "Testing Serial Independence via Density-Based Measures of Divergence." *Methodology and Computing in Applied Probability*. `doi:10.1007/s11009-013-9320-4`.

Bagnato L, De Capitani L, Punzo A (2014). "Detecting Serial Dependencies with the Reproducibility Probability Autodependogram." *Advances in Statistical Analysis*, **98**(1), 35–61.

Bagnato L, Punzo A (2010). "On the Use of χ^2-Test to Check Serial Independence." *Statistica & Applicazioni*, **VIII**(1), 57–74.

Bagnato L, Punzo A (2012a). "Checking Serial Independence of Residuals from a Nonlinear Model." In W Gaul, A Geyer-Shulz, L Schmidt-Thieme, J Kunze (eds.), *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, Studies in Classification, Data Analysis and Knowledge Organization, pp. 203–211. Springer-Verlag, Berlin.

Bagnato L, Punzo A (2012b). "Finite Mixtures of Unimodal Beta and Gamma Densities and the k-Bumps Algorithm." *Computational Statistics*, **28**(4), 1571–1597.

Bagnato L, Punzo A (2013). "Using the Autodependogram in Model Diagnostic Checking." In N Torelli, F Pesarin, A Bar-Hen (eds.), *Advances in Theoretical and Applied Statistics*, Studies in Theoretical and Applied Statistics, pp. 129–139. Springer-Verlag, Berlin.

Bagnato L, Punzo A, Nicolis O (2012). "The Autodependogram: A Graphical Device to Investigate Serial Dependences." *Journal of Time Series Analysis*, **33**(2), 233–254.

Bloomfield DSF, Haberman S (1987). "Graduation: Some Experiments With Kernel Methods." *Journal of the Institute of Actuaries*, **114**(2), 339–369.

Camarda CG (2012). "**MortalitySmooth**: An R Package for Smoothing Poisson Counts with P-Splines." *Journal of Statistical Software*, **50**(1), 1–24. URL http://www.jstatsoft.org/v50/i01/.

Chen SX (2000). "Beta Kernel Smoothers for Regression Curves." *Statistica Sinica*, **10**(1), 73–91.

Copas JB, Haberman S (1983). "Non-Parametric Graduation Using Kernel Methods." *Journal of the Institute of Actuaries*, **110**(1), 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), 215–233.

Elzhov TV, Mullen KM, Spiess AN, Bolker B (2013). ***minpack.lm**: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK*. R package version 1.1-8. URL http://CRAN.R-project.org/package=minpack.lm.

Gavin JB, Haberman S, Verrall RJ (1993). "Moving Weighted Average Graduation Using Kernel Estimation." *Insurance: Mathematics and Economics*, **12**(2), 113–126.

Gavin JB, Haberman S, Verrall RJ (1994). "On the Choice of Bandwidth for Kernel Graduation." *Journal of the Institute of Actuaries*, **121**(1), 119–134.

Gavin JB, Haberman S, Verrall RJ (1995). "Graduation by Kernel and Adaptive Kernel Methods with a Boundary Correction." *Transactions of Society of Actuaries*, **47**, 173–209.

Goldstein JR (2011). "A Secular Trend toward Earlier Male Sexual Maturity: Evidence from Shifting Ages of Male Young Adult Mortality." *PLoS ONE*, **6**(8), e14826.

Härdle W (1992). *Applied Nonparametric Regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Hayfield T, Racine JS (2008). "Nonparametric Econometrics: The **np** Package." *Journal of Statistical Software*, **27**(5), 1–32. URL http://www.jstatsoft.org/v27/i05/.

Heligman L, Pollard JH (1980). "The Age Pattern of Mortality." *Journal of the Institute of Actuaries*, **107**(1), 49–80.

Human Mortality Database (2013). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at http://www.mortality.org/ or http://www.humanmortality.de/ (data downloaded in May 2013).

Hyndman RJ, Booth H, Tickle L, Maindonald J (2014). ***demography**: Forecasting Mortality, Fertility, Migration and Population Data.* R package version 1.17. URL http://CRAN.R-project.org/package=demography.

London D (1985). *Graduation: The Revision of Estimates.* Actex Publications Abington, Connecticut.

Mazza A, Punzo A (2011). "Discrete Beta Kernel Graduation of Age-Specific Demographic Indicators." In S Ingrassia, R Rocci, M Vichi (eds.), *New Perspectives in Statistical Modeling and Data Analysis*, Studies in Classification, Data Analysis and Knowledge Organization, pp. 127–134. Springer-Verlag, Berlin.

Mazza A, Punzo A (2013a). "Graduation by Adaptive Discrete Beta Kernels." In A Giusti, G Ritter, M Vichi (eds.), *Classification and Data Mining*, Studies in Classification, Data Analysis and Knowledge Organization, pp. 243–250. Springer-Verlag, Berlin.

Mazza A, Punzo A (2013b). "Using the Variation Coefficient for Adaptive Discrete Beta Kernel Graduation." In P Giudici, S Ingrassia, M Vichi (eds.), *Statistical Models for Data Analysis*, Studies in Classification, Data Analysis and Knowledge Organization, pp. 225–232. Springer International Publishing, Switzerland.

Moré J (1978). "The Levenberg-Marquardt Algorithm: Implementation and Theory." In GA Watson (ed.), *Numerical Analysis*, volume 630 of *Lecture Notes in Mathematics*, pp. 104–116. Springer-Verlag, Berlin.

Parkes AS (1976). *Patterns of Sexuality and Reproduction.* Oxford University Press, New York.

Peristera P, Kostaki A (2005). "An Evaluation of the Performance of Kernel Estimators for Graduating Mortality Data." *Journal of Population Research*, **22**(2), 185–197.

Preston SH (1976). *Mortality Patterns in National Populations: With Special Reference to Recorded Causes of Death.* Academic Press, New York.

Punzo A (2010). "Discrete Beta-Type Models." In H Locarek-Junge, C Weihs (eds.), *Classification as a Tool for Research*, Studies in Classification, Data Analysis and Knowledge Organization, pp. 253–261. Springer-Verlag, Berlin.

Punzo A, Zini A (2012). "Discrete Approximations of Continuous and Mixed Measures on a Compact Interval." *Statistical Papers*, **53**(3), 563–575.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Šidák Z (1967). "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions." *Journal of the American Statistical Association*, **62**(318), 626–633.

Affiliation:

Angelo Mazza, Antonio Punzo
Department of Economics and Business
University of Catania
Corso Italia, 55, 95129 Catania, Italy
E-mail: a.mazza@unict.it, antonio.punzo@unict.it
URL: http://www.economia.unict.it/punzo/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 57, Code Snippet 2 *Submitted:* 2012-09-24
April 2014 *Accepted:* 2013-11-04