SORT 36 (1) January-June 2012, 3-32

Analysing visual receptive fields through generalised additive models with interactions

María Xosé Rodríguez-Álvarez[1], Carmen Cadarso-Suárez[2]
and Francisco González[3,4]

Abstract

Visual receptive fields (RFs) are small areas of the visual field where a stimulus induces a responses of a particular neuron from the visual system. RFs can be mapped using reverse cross-correlation technique, which produces raw matrices containing both spatial and temporal information about the RF. Though this technique is frequently used in electrophysiological experiments, it does not allow formal comparisons between RFs obtained under different experimental conditions. In this paper we propose the use of Generalised Additive Models (GAM) including complex interactions, to obtain smoothed spatio-temporal versions of RFs. Moreover, the proposed methodology also allow for the statistical comparisons of the RFs obtained across various experimental conditions. Data analysed here derive from studies of neurons' activity in the visual cortex of behaving monkeys. Our results suggest that the GAM-based technique proposed in this paper can be a flexible and powerful tool for assessing receptive field properties.

MSC: 65D07, 65D10, 62P10

Keywords: Visual receptive fields, reverse cross-correlation, visual cortex, smoothing, B-splines, P-splines, tensor product splines.

1. Introduction

One of the techniques used in neurophysiology is electrophysiology, which records the electrical activity produced by neurons. Electrophysiology allows to study the

[1] Clinical Epidemiology and Biostatistics Unit, Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), Spain. maria.jose.rodriguez.alvarez2@sergas.es

[2] Unit of Biostatistics, Dept. of Statistics and OR. Universidade de Santiago de Compostela, Spain.

[3] Department of Phyisiology. School of Medicine. Universidade de Santiago de Compostela, Spain.

[4] Service of Ophthalmology, Complexo Hospitalario Universitario de Santiago de Compostela, Spain.
Received: December 2011

association between sensory stimuli and the neural response in any part of the brain, such as in the visual cortex. Neurons produce sudden changes in their membrane potential known as 'spikes', that can be recorded with microelectrodes. These spikes encode the information produced by the neurons. The analysis of the frequency of spike discharges gives us insights on how the neurons and the nervous system work.

Visual receptive fields (RF) are a basic feature of single cells in the visual system. They are small areas of the visual field that a particular visual neuron 'sees'. They have different sizes depending on the visual area we are considering. From the cell reponses (spikes) we can infer the spatio-temporal properties of the RFs. RFs can be mapped using different techniques. Initially, receptive field maps were manually mapped using simple dots, lines and edges. Despite its simplicity, this early mapping showed that the spatial structure of visual receptive fields as divided into 'on' and 'off' subregions (Barlow, 1953; Hartline and Ratliff, 1958; Hubel and Wiesel, 1962; Kuffler, 1953). Each of these regions responded to the onset of a bright ('on') or dark ('off') spot respectively and was the basis of the circular center-surround organization of visual receptive fields of retinal cells. In the primary visual cortex, Hubel and Wiesel (1962) classified cortical neurons into two groups, simple and complex. Simple cells had receptive fields divided with separate 'on' and 'off' regions, whereas complex cells did not have such division.

Visual cells from area V1 – which is a primary visual cortical area – show ocular dominance. This effect occurs because these cells receive stronger inputs from one eye than from the other. Some cells receive inputs with equal strength from both eyes. Ocular dominance makes that the RF of a given cell can be different depending on what eye is stimulated.

Further quantitative techniques of receptive field mapping have contributed significantly to refine our knowledge about receptive field properties and organization. Post-stimulus responses to flashing or moving bars or dots were used to map receptive fields in the visual system of both cats (Bishop et al., 1973; Henry and Bishop, 1972; Pei et al., 1994) and monkeys (Tsao et al., 2003).

Receptive field mapping techniques that use a reverse correlation analysis have been effective in providing detailed receptive field maps of neurons in early stages of the visual pathway (DeAngelis et al., 1993; Jones and Palmer, 1987; Krause et al., 1987; Reid et al., 1997). Binocular receptive fields have also been mapped using reverse correlation techniques and with stimuli with various binocular disparities (Gonzalez et al., 2001).

Briefly, reverse cross-correlation is a technique that can be used for studying how visual neurons process signals from different positions in their receptive field, and can provide both spatial and temporal information about their RF. A schematic illustration on how the reverse cross-correlation technique was used in our experiments is shown in Figure 1. The animal was viewing two monitors (Figure 1A) with a fixation target. Within a square area over the cell receptive field a bright or dark spot was flashed in different positions in a pseudorandom manner. Cell spikes were recorded while the stimulus was delivered (Figure 1B). When a spike was produced (t_0), the stimulus

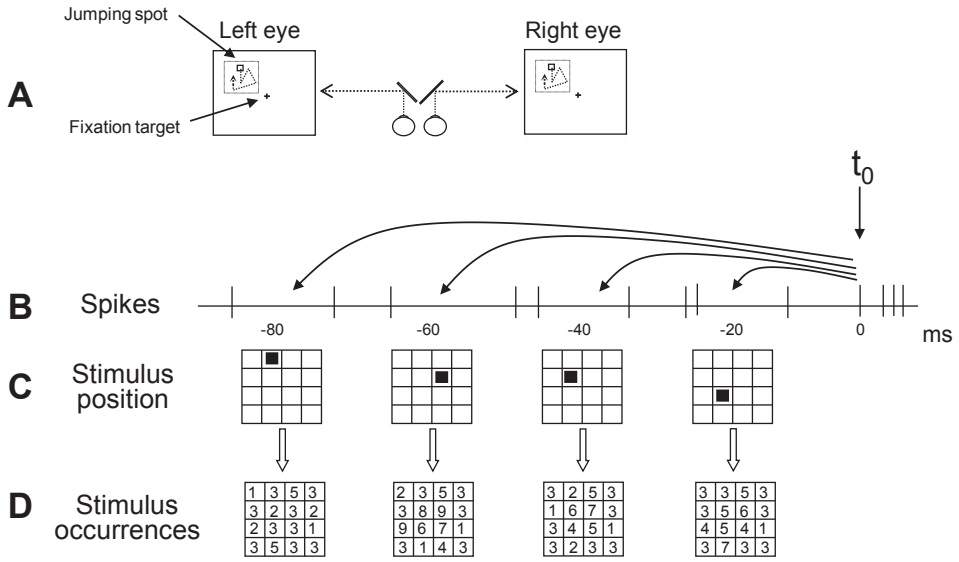


Figure 1: *Reverse cross-correlation technique. The animal was viewing two monitors (A) with a fixation target. Within a square area over the cell receptive field a bright or dark spot was flashed in different positions in a pseudorandom manner. Cell spikes were recorded while the stimulus was delivered (B). When a spike was produced (t_0), the stimulus position at several pre-spike times (-20, -40, …, -320 ms) was read (C) and the corresponding position was increased by one. The result was a numerical matrix containing the number of stimulus occurrences at each position (D).*

position at several pre-spike times $(-20, -40, \ldots, -320$ ms) was read (Figure 1C), and the corresponding position was increased by one. The result was a numerical matrix containing the number of stimulus occurrences at each position (Figure 1D). The graphical representation of this matrix is what we call receptive field map (RFmap).

Although this technique provides raw receptive field maps, it does not allow further aspects of receptive field analysis such as formal comparisons between left and right receptive field maps, 'on' and 'off' maps, or between monocular and binocular receptive field maps.

The main objective of this paper is to use flexible regression models including complex interactions for modelling the visual receptive fields over time, which in turn may vary across various experimental conditions. Specifically, we suggest the use of Poisson Generalised Additive Model (GAM, Hastie and Tibshirani, 1990) which expresses the cell response (i.e., number of spikes) as a smooth function of both space and time, including high-order interactions. Advantages of using this regression model-based approach include the following:

- construction of smoothed versions of RF maps, by including spatial effects in the model.;
- explanation of differences in the course of the cell activity in a unified way.

So far, several ways to smooth cell activity have been studied in the literature. An overview of the application of smoothing techniques in neuronal data can be found in Kass et al. (2003). Cadarso et al. (2006) and Roca-Padiñas et al. (2006, 2011), for example, employ a flexible modeling technique based on the logistic GAM with local linear kernel smoothers. Faes et al. (2008) apply a flexible method based on natural cubic splines to model synchrony in neuronal firing. Other recent techniques in this context include the Bayesian adaptive regression splines (Behseta and Kass, 2005; Behseta et al., 2005; DiMatteo et al., 2001). Flexible regression-based techniques come out favourably since they enjoy the flexibility of capturing the spatio-temporal evolution without the restriction of parametric modeling as well as the possibility to include covariate or factor information. We revisit this aspect in Section 3 where the models discussed happen to share similar properties.

Though GAMs were successfully applied in electrophysiological experiments, no attempt was made so far to use spatio-temporal GAM models when modelling neural data. To this aim, in this paper we follow the GAM approach suggested by Wood (2006a), as a flexible way to model the temporal evolution of visual RFs, across different conditions. The estimation algorithm used for fitting GAMs is based on penalised regression splines (Eilers and Marx, 1996), in combination with B-splines basis functions and the representation of the GAM as a mixed model. This representation is very appealing, since it allows to select the amount of smoothing automatically via restricted (or residual) maximum likelihood (REML). This statistical approach provides a computationally efficient way of estimating the model and making inference when dealing with neural data.

The paper is organized as follows. In Section 2, the electrophysiological experiment is discussed. GAMs including interactions are introduced in Section 3. In Section 4, we present the main results that were obtained for our study. Finally, we point out some conclusions in Section 5.

2. The electrophysiological experiment

2.1. Animal preparation

The experimental setup and physiological recording were reported in detail elsewhere (Gonzalez et al., 1993; Gonzalez and Krause, 1994; Gonzalez et al., 2001). Two monkeys (*Macaca mulatta*) were trained to perform with their head fixed a task that required a steady visual fixation on a small target (0.3×0.2 deg). The behavourial task consisted of a series of trials from 1 to 2 s separated by an intertrial interval of 1 s. Single-cell activity was recorded by means of metal microelectrodes (5 Mohm, AMSystems Inc., Washington, USA) inserted in the brain. For this, a stainless steel cylindrical chamber was attached to the skull covering area V1. To allow access to the visual cortex, small craniotomies of 5 mm diameter were made in the exposed skull within the cham-

ber. Microelectrode penetrations were made by using an electrohydraulic microdriver (Narishige, Japan) mounted on the chamber. Neural voltage signals were amplified and filtered using conventional electronic equipment. A time-amplitude window discriminator (Bak Electronics, Rockville, Maryland, USA) was used to convert the amplified, filtered neural signal to TTL pulses, which were collected by a conventional computer. At the end of the experiments the animal was deeply anaesthetized with nembutal, the brain perfused with 10% formalin and the region where the penetrations were made was blocked and sliced. Sections were stained with toluidine blue and the electrode tracks were reconstructed. Ocular movements were monitored by means of a video camera under infrared illumination. A frame grabber (Imagenation PXC200, Oregon, USA) attached to a conventional personal computer was used to detect the corneal reflex on the left eye and abort the trials when eye movements exceeded a fixation window of 1×1 deg. All surgical procedures were made under general anesthesia (ketamine i.m. 10 mg/kg, followed by sodium pentobarbital i.v. 27 mg/kg). Supplementary pentobarbital was given whenever necessary during the surgical procedure and analgesics and antibiotics (noramidopirine i.m. 150 mg/kg, penicillin i.m. 50000 IU/kg) were given at the end of the surgery. Cleaning and asepsis of the implant was made periodically. All animal procedures were performed in accordance with the guidelines of the Bioethic Committee of our institution.

2.2. Visual stimulation

The presentation of stimuli and the collection, storage and on-line display of analysed data were controlled by five conventional personal computers running software developed in our own laboratory (Gonzalez and Krause, 1994). The animal was placed in front of a two-mirror system allowing simultaneous and separate viewing of two monitors (Model CPD- 520GST, Sony) placed laterally 57.7 cm away from the monkey's eyes subtending 44.8×28 of visual field and set for a resolution of 320×200 pixels (1 pixel = 0.14) and 70 frames per second. Once the assessment of the basic properties of the RF of the cell under study was made, the stimulation procedure to obtain the RF maps (see below) was started.

2.3. Data acquisition

To obtain data to produce the RFmaps we first used a reverse cross-correlation technique (DeBoer and Kuyper, 1968; Jones and Palmer, 1987; Krause et al., 1987; Pérez et al., 2005). For this, we flashed a small bright square (jumping spot), 14.0 cd/m^2) on a grid with 16×16 spatial locations (2.2×2.2 deg). The spot size was adjusted to cover $1/16$th of the grid side and flashed for a duration of $1/70$s (one frame of the monitor) at different grid locations (Pérez et al., 2005). Figure 1 schematically shows this technique. The

background of the monitor screens had a constant luminance of 2 cd/m^2. To obtain 'on' responses, a bright (14.0 cd/m^2) jumping spot was used. To obtain 'off' responses, a dark (0.28 cd/m^2) jumping spot was used. For each successive presentation the location of the spot on the grid was chosen randomly. Each time a neural spike was detected we correlated it with the stimulus position at various prespike times and constructed series of matrices for different pre-spike times. To produce reliable maps, each sequence of stimulation typically required at least 20 presentations of the stimulus on each location of the stimulus grid. For each cell, a series of matrices with 256 grid positions each, covering a prespike time from 20 to 320 ms (at 20 ms interval), were obtained, under different experimental conditions. A local specific coupling between the stimulus and response at a particular prespike time and disparity will produce an emerging set of data with high values on this particular location in the matrix.

3. Statistical methodology

Regression analysis plays a fundamental role in statistics. The objective of this statistical methodology is to evaluate the influence of some explanatory variables (called covariates), $\boldsymbol{x} = (x_1, \ldots, x_p)$, on the mean of a measure of interest y. This relation is given by

$$E[y|x_1, \ldots, x_p] = E[y|\boldsymbol{x}] = m(x_1, \ldots, x_p), \tag{1}$$

where $m(\cdot)$ is a multivariate function, usually denoted as the mean regression function.

The most usual way to model the dependence between the response variable and the covariates is through the multiple linear regression model. In this model, the response y given covariates $\boldsymbol{x}$ is assumed to be normally distributed $\left(y|\boldsymbol{x} \sim N\left(m(\boldsymbol{x}), \sigma^2\right)\right)$, and the covariates are assumed to have a linear effect on the response. Specifically, the following model is assumed

$$E[y|\boldsymbol{x}] = m(x_1, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\mathsf{T}}$ is a vector of unknown regression coefficients.

However, in some circumstances, the assumption of linearity in the effects of the continuous covariates is very restrictive and is not supported by the data at hand. In this setting non-parametric regression techniques are involved in modelling the dependence between y and $\boldsymbol{x}$, but without specifying, in advance, the function $m(\cdot)$ in (1) that relates the covariates and the response. However, if no restrictions are imposed to the funcion $m(\cdot)$, some problems arise. First, fully non-parametric data analysis can be afflicted by the so-called *curse of dimensionality* (Bellman, 1961): as the dimension of the number of covariates increases, it becomes exponentially more difficult to estimate the function $m(\cdot)$. Second, it is difficult to visualize a regression surface $m(\cdot)$ for more than two

covariates. To overcome these difficulties, additive regression models (AMs, Hastie and Tibshirani, 1990) are an alternative to unconstrained non-parametric regression with several covariates. The additive regression model is defined as

$$E[y|\boldsymbol{x}] = m(x_1,\ldots,x_p) = f_1(x_1) + \cdots + f_p(x_p), \tag{2}$$

where f_k $(k = 1,\ldots,p)$ are smooth and unknown functions.

In many applications, the response y can be discrete (e.g. count or binary data). These situations require a different model for the conditional expectation of y, since a direct connection between $E[y|\boldsymbol{x}]$ and the additive predictor $f_1(x_1) + \cdots + f_p(x_p)$ in (2) is no longer possible unless some constraints are imposed. In such cases, the generalised additive models (GAMs) extends the AM by allowing for different distributions of the response variable y. Specifically, GAMs assume that the distribution of the response variable y given covariates $\boldsymbol{x}$ belongs to the exponential family (McCullagh and Nelder, 1989). In these models, the relationship between $E[y|\boldsymbol{x}]$ and the covariates is specified through the following model

$$E[y|\boldsymbol{x}] = g\left(f_1(x_1) + \cdots + f_p(x_p)\right), \tag{3}$$

where $g(\cdot)$ is a monotonic known function (the inverse of link function).

A weakness of the AM and GAM given in equations (2) and (3) is that these models completely ignore the fact that the functional form of a covariate effect often varies according to the values taken by one or more of the remaining covariates. It is not unusual to find situations in which more complex models are needed. This is the case of our real data example, in which we could expect that the response of interest varies smoothly across the visual receptive field (defined in terms of x- and y- coordinates). In recent years, a number of papers have appeared which address the problem of estimating AMs and GAMs with interaction terms. Hastie and Tibshirani (1990) discussed various approaches using smoothing splines. Wahba (1990), among others, proposed the use of smoothing spline ANOVA methods. Roca-Padiñas et al. (2006, 2008) also proposed alternative methods based on kernel-type smoothers. In the context of penalised regression splines, Brezger and Lang (2006), Currie et al. (2006), Eilers and Marx (2003), Lee and Durbán (2011), Wood (2006b), among others, are several references related to multidimensional smoothing. This paper is focused on penalised regression splines, and more specifically on those approaches implemented in the `mgcv` package [Wood (2006a)] of the `R` (R Development Core Team, 2011) statistical software. We begin this section by presenting the idea of penalised spline smoothing in the univariate case and then we address the multivariate and multidimensional regression problem.

3.1. Penalised spline smoothing

For simplicity, we introduce the idea of penalised spline smoothing in the univariate Gaussian case. Specifically, let

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad i = 1, \ldots, n, \tag{4}$$

where $f(\cdot)$ is a smooth and unknown function of covariate x that needs to be estimated from the data points (x_i, y_i), $i = 1, \ldots, n$.

To estimate the function $f(\cdot)$, it is assumed that this function can be represented as a linear combination of d known basis functions B_j, i.e.

$$f(x) = \sum_{j=1}^{d} \theta_j B_j(x), \tag{5}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^{\mathsf{T}}$ is a vector of unknown regression coefficients. Under this representation model (4) is purely parametric, and therefore it can be easily estimated using ordinary least squares

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{B}^{\mathsf{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathsf{T}}\boldsymbol{y},$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathsf{T}}$ and $\mathbf{B}$ is the design matrix

$$\mathbf{B} = \begin{pmatrix} B_1(x_1) & B_2(x_1) & \cdots & B_d(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ B_1(x_n) & B_2(x_n) & \cdots & B_d(x_n) \end{pmatrix}.$$

There are several alternatives for the choice of the basis functions B_j $(j = 1, \ldots, d)$ in (5). The simplest basis functions are the polynomials, where $B_j(x) = x^{j-1}$. However, these basis functions have the disadvantage of not being very flexible, and, tend to produce an artificial behaviour at the boundaries. Some alternatives are the so-called *spline* basis, as truncated polynomial, B-splines (de Boor, 2001), or thin plate regression splines (Wood, 2003). For the sake of illustration, this paper is focused on B-splines. In this case, the function $f(\cdot)$ in (5) is specified with respect to a given set of knots

$$x_{min} = k_1 < k_2 < \cdots < k_k = x_{max}$$

placed at equidistant or noequidistant points over the domain of x. The knots divide the domain of x, such that each interval will be covered by $p+1$ B-splines of degree p, and the number of B-splines in (5) is $d = k + p - 1$. Figure 2(a) shows an example

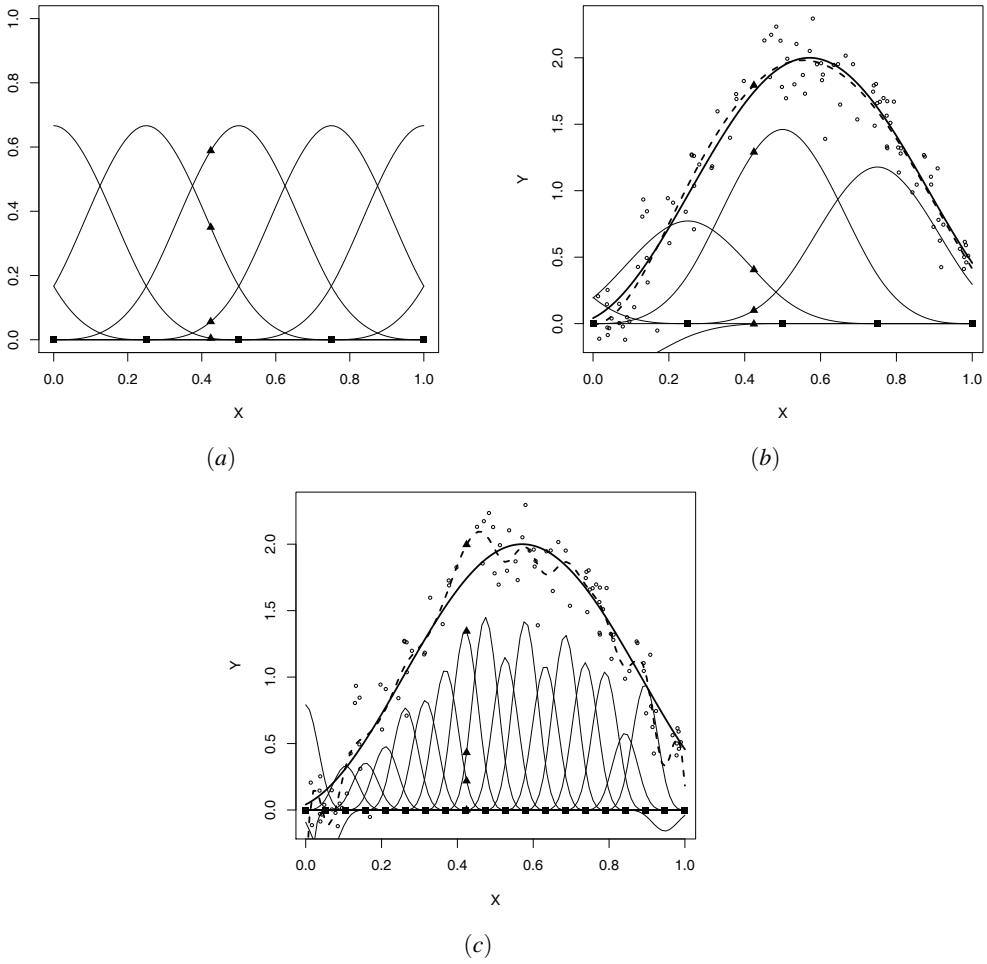


Figure 2: *(a) B-splines basis functions of degree $p = 3$ based on $k = 5$ knots (black squares). The solid triangles indicate the non-zero B-splines at $x = 0.42$. Shown in (b) and (c) are the "weighted" (accordingly to the coefficient estimate $\hat{\boldsymbol{\theta}}$) B-Spline basis function (solid lines), the true function $f(x) = 1 + \sin(5(x+1))$ (bold solid line), and the fitted curve (bold dotted lined) based on $k = 5$ and $k = 20$ knots respectively.*

of B-spline basis functions with $k = 5$ knots and degree $p = 3$. In order to estimate the function $f(\cdot)$ in (5) (or more precisely, the vector $\boldsymbol{\theta}$) we have therefore to specify the number and location of knots and the degree of the B-spline. As an illustration of the impact of the number of knots, we simulated $n = 100$ data from the function $f(x_i) = 1 + \sin(5(x_i+1))$, with $x_i \sim U[0,1]$ and independent $\varepsilon_i \sim N(0, 0.2)$. Figures 2(b) and 2(c) show the estimated curves (via ordinary least squares) with $k = 5$ and $k = 20$ equidistant knots respectively. As can be observed, as the number of knots increases, the estimated curve (in comparison with the true curve) becomes too wiggly, meaning that the data are overfitted. Thus, a wrong choice in the number (and also the location) of knots can lead to estimates, and therefore conclusions, that could be erroneous.

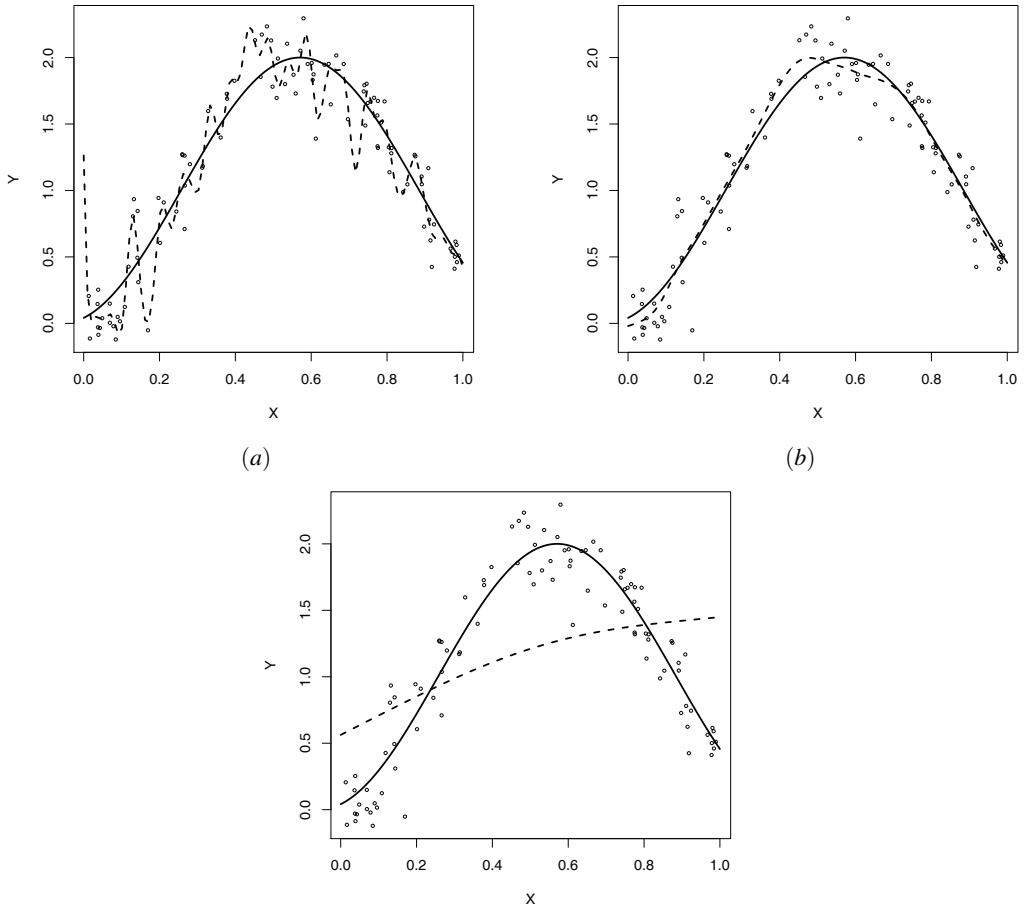


Figure 3: *True function* $f(x) = 1 + \sin(5(x+1))$ *(bold solid line), and the fitted curve (bold dotted lined) based on* $p = 3$ *degree,* $k = 40$ *knots; and* $q = 2$*-nd order derivative: (a)* $\lambda = 10^{-6}$*; (b)* $\lambda = 7$*; and (c)* $\lambda = 10^4$.

Although several approaches have been proposed to select an optimal set of knots (see e.g. Fried and Silverman, 1989; Lee, 2002), all of them have the disadvantage of being computationally intensive. To overcome this problem O'Sullivan (1986), and later Eilers and Marx (1996), introduced the idea of penalised splines, where a smoothness penalty is added to the least squares criterion when estimating the regression coefficients $\boldsymbol{\theta}$ in (5). Although in Eilers and Marx (1996) a discrete penalty is proposed, for the sake of presentation we considered here a penalty based on the integrated derivatives of function f. Specifically, in the case of penalised splines a large amount of knots (e.g. $\min\{n/4, 40\}$) is chosen, and instead of fitting the model by minimising the sum of squares,

$$||\boldsymbol{y} - \mathbf{B}\boldsymbol{\theta}||^2,$$

it is fitted by minimising a *penalised* sum of squares

$$||\boldsymbol{y} - \mathbf{B}\boldsymbol{\theta}||^2 + \lambda \int_{x_{min}}^{x_{max}} [f^q(x)]^2 dx, \tag{6}$$

where the integral of the square of the q-th order derivative, f^q, penalises models that are too wiggly. The smoothing parameter λ controls the trade-off between the bias and the variance of the resulting estimates. The result of a using large smoothing parameter ($\lambda \to \infty$) is an oversmoothed curve, leading in the limit to the least squares $(q-1)$-order polynomial through the data. On the other hand, a small smoothing parameter ($\lambda \to 0$) tends reproduce the data. This can be observed in Figure 3, where the data presented in Figure 2 were re-analysed but incorporating the penalisation. Thus, the optimal amount of smoothing λ has to be chosen by compromising goodness of fit with complexity of the estimated function. This issue is the subject of the next subsection.

It sould be noted that since the function $f(\cdot)$ is linear in the coefficients θ_j (see (5)), the penalty can also be writen as (see, for instance, Marra and Radice, 2010)

$$\int_{x_{min}}^{x_{max}} [f^q(x)]^2 dx = \boldsymbol{\theta}^\top \mathbf{K} \boldsymbol{\theta}, \tag{7}$$

where $\boldsymbol{K}$ is a known $d \times d$ matrix, whose elements depend on the chosen spline basis. Accordingly, for a given λ, the minimiser of (6) is then

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{K}\right)^{-1} \mathbf{B}^\top \boldsymbol{y}.$$

In the case of B-splines, the integrated square of the q-th order derivative of function $f(\cdot)$ can be well approximated by differences of the sequence of regression coefficients $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. Accordingly, the smoothness penalty presented in (7) is equivalent to impose a penalty on q-th order differences of adjacent B-Spline coefficients. For a more detailed review of this topic, see Eilers and Marx (1996).

3.1.1. Smoothing parameter selection

As pointed out before, a crucial point in penalised spline smoothing is the choice of the smoothing parameter λ in (6). If a large smoothing parameter is chosen, the resulting curve estimate is very smooth, but if a small smoothing parameter is chosen the resulting estimate becomes too wiggly. Therefore, it is important to have procedures for helping in the selection of the optimal smoothing parameter. In this section, we review several methods to choose the 'optimal'value of λ.

(Generalised) Cross Validation When using cross validation (CV) to select the the optimal smoothing parameter, the objective is to obtain the λ value which minimises the

cross validation expression

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}^{-i}(x_i)\right)^2,$$

where $\hat{f}^{-i}$ denotes the fit obtained by leaving out the ith data point.

As can be observed in the expression above, CV implies to fit n different models (as many as the number of observations), which is computationally very intensive. Fortunately, a more efficient equivalent expression can be obtained Hastie and Tibshirani (1990):

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}(x_i)}{1 - h_{ii}}\right)^2,$$

where h_{ii} are the diagonal elements of the *hat matrix*

$$\mathbf{H} = \mathbf{B}\left(\mathbf{B}^{\mathsf{T}}\mathbf{B} + \lambda\mathbf{K}\right)^{-1}\mathbf{B}^{\mathsf{T}},$$

that is the matrix such that $\hat{\boldsymbol{y}} = \hat{f}(\boldsymbol{x}) = \mathbf{H}\boldsymbol{y}$. It should be noted that the trace of the hat matrix $\mathbf{H}$ define the *effective degrees of freedom* (edf) (i.e. the 'effective' number of parameters) of a smoother (see Hastie and Tibshirani, 1990; Wood, 2006a).

However, the CV criterion suffers from several drawbacks (see Wahba, 1990; Wood, 2006a). For instance: (a) in the case of more than one smooth function, it becomes computationally expensive; and (b) it presents a lack of invariance. A modified version of this criterion, the generalised cross validation (GCV), was suggested by Craven and Wahba (1979), and presents some advantages over CV (see Craven and Wahba, 1979; Wahba, 1990). The GCV is defined as:

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}(x_i)}{1 - \sum_{j=1}^{n} h_{jj}/n}\right)^2.$$

An efficient implementation – in the multivariate additive case – of the GCV criterion (Wood, 2000) was the origin of the `mgcv` package, and for this reason the package deserves its name (multivariate generalised cross validation). Later, Wood (2004) proposed an alternative method to the Wood (2000) which overcame some of its limitations. Primarily the new method is particularly robust numerically, and can deal with rank deficiency in the model. This is, by now, the default method (for the additive case) to choose the smoothing parameters in the `mgcv` package.

Akaike Information Criterion A common approach for the λ selection is to optimise criteria such as the Akaike's information criterion (AIC). In this case, the optimal

smoothing parameter λ is that which minimises

$$AIC(\lambda) = \text{dev}(\boldsymbol{y}; \boldsymbol{\theta}; \lambda) + 2\text{trace}(\mathbf{H}),$$

where dev denotes the deviance of the model and $\mathbf{H}$ is the hat matrix.

In the `mgcv` package, the AIC, or more specifically a rescaled AIC – the unbiased risk estimator (UBRE) (Craven and Wahba, 1979) – is the default method when the scale parameter of the exponential family to which the response y pertains is known (for instance, in the binomial and Poisson cases) (see Wood, 2004, 2008).

Mixed Model Representation A different approach to choose the optimal smoothing parameter λ comes from the fact that a P-Spline regression model can be formulated as a linear mixed model (Brumback et al., 1999; Currie and Durbán, 2002). Although a detailed presentation of this approach is beyond the scope of this paper, we briefly describe here the main ideas behind it.

In this approach, the design matrix $\mathbf{B}$ and the vector of regression coefficients $\boldsymbol{\theta}$ are reformulated is such a way that

$$\boldsymbol{y} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \mathbf{U}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}, \quad \text{with} \quad \boldsymbol{u} \sim N(0, \mathbf{G}) \quad \text{and} \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}_n)$$

where $\mathbf{I}_n$ is the identity matrix, $\mathbf{U}$ and $\mathbf{Z}$ are the model matrices, and $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are the fixed and random effects coefficients of the linear mixed model respectively. The random effects have covariance matrix $\mathbf{G}$, which depends on the variance σ_u^2, $\mathbf{G} = \sigma_u^2 \mathbf{I}_K$. Under this new configuration, it can be shown that λ is given by the ratio of the variance components, i.e., $\lambda = \frac{\sigma^2}{\sigma_u^2}$. Accordingly, the estimation of the P-Spline model can be obtained using standard procedures for the estimation of a linear mixed model, and the choice of the smoothing parameter becomes the estimation of the variance components (either via *maximum likelihood* (ML) or *restricted (or residual) maximum likelihood* (REML)).

Recently, Wood (2011) presented a computationally efficient way of estimating the smoothing parameters of P-splines models in which a Laplace approximation is used to obtain an approximate REML or ML. These methods are already implemented in the `gam()` function of the `mgcv` package.

Simulation results presented in Wood (2011) and Strasak et al. (2011) suggest that REML and ML methods offer some improvement in terms of the mean-square error and the stability of the estimator relative to GCV or AIC in most cases. Moreover, Reiss and Odgen (2009) also show that at finite sample sizes GCV or AIC are prone to undersmoothing and are more likely to develop multiple minima than REML. For all these reasons, the use of the REML method is recommended.

3.1.2. Additive models and identifiability constraints

In some circumstances, the objective of a real data analysis could be to jointly evaluate the effect of two (or more) covariates, x_1 and x_2, on the response of interest y. In this case, an appropiate model could be the addivive model

$$y_i = f_1(x_{i1}) + f_2(x_{i2}) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad i = 1, \ldots, n, \tag{8}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are smooth and unknown functions.

However, this model presents an identifiability problem due to the fact that it incorporates more than one covariate. We could subtract a constant c of any smooth function ($f_1(x_1) - c$), and add it to another one ($f_2(x_2) + c$), and the same regression model would be obtained. To avoid these free constants, it is necessary to impose some restrictions (that is, among all equivalent models, we have to choose one). In this case, the usual way to proceed to guarantee the identification of the model is to incorporate a constant α, and to "centre" each of the smooth functions in some way, for example by assuming:

$$\sum_{i=1}^{n} f_1(x_{1i}) = 0 \quad \text{and} \quad \sum_{i=1}^{n} f_2(x_{2i}) = 0,$$

yieding the model

$$y_i = \alpha + f_1(x_{i1}) + f_2(x_{i2}) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad i = 1, \ldots, n. \tag{9}$$

Once the identifiability problem has been solved, each of the smooth functions in (9) can be represented using regression splines, and the model can be estimated using penalised least squares (or the mixed model approach) as in the univariate case, with the smoothing parameters of each of the smooth functions being selected via any of the criteria previously presented.

3.1.3. Generalised additive models

The P-Spline methodology presented in the previous Sections can be easily extended to deal with a non-gaussian response y, given that the distribution of y conditional on the covariates $(x_1, \ldots, x_p)$ belongs to the exponential family. In this case, model (3) can be estimated on the basis of the penalised log-likelihood by means of penalised iterative re-weighted least squares (P-IRLS) (see Eilers and Marx, 1996; Wood, 2006a) or by its representation as a generalised linear mixed model (GLMM) (Breslow and Clayton, 1993).

3.2. Factor-by-curve interaction

In this section we are concerned with situations where the effect of a continuous covariate x on the response varies across groups defined by the levels of a categorical variable z (for example, a factor with M levels $\{1,\ldots,M\}$).

To face such situations, *varying coefficient* terms (Hastie and Tibshirani, 1993) can be used. Here, the effect of x is assumed to vary over the range of z, and the regression model then becomes

$$y_i = \alpha + \sum_{l=2}^{M} \alpha_l \mathbf{1}_{\{z_i=l\}} + \sum_{l=1}^{M} \mathbf{1}_{\{z_i=l\}} f_l(x_i) + \varepsilon_i, \tag{10}$$

where $\mathbf{1}_A$ denotes the indicator function of event A (used to construct the *dummy* variables).

As can be observed in equation (10), the model assumes a different effect of covariate x for each of the levels of the categorical covariate z. The inclusion of the "main" effect of z $\left(\sum_{l=2}^{M} \alpha_l \mathbf{1}_{\{z_i=l\}}\right)$ is needed due to identifiability contraints (given that each of the smooth functions $f_l(\cdot)$ is centred). Again, each function $f_l(\cdot)$ in (10) can be represented using regression splines, and the model can then be fitted using penalised least squares or the mixed model approach (or by P-IRLS and GLMM in the generalised case).

3.3. Continuous bivariate interactions

In many applications, the additive structure of the model presented in (8) is not appropriate for the data at hand, and more complex models are needed. In this section we are concerned with the case where the response y is expressed on the basis of a bivariate surface defined by covariates x_1 and x_2:

$$y_i = f_{12}(x_{i1}, x_{i2}) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad i = 1,\ldots,n, \tag{11}$$

To date, several P-spline approaches to estimate model (11) have been suggested in the statistical literature. For instance, Wood (2003) proposed to model the surface $f_{12}(x_1, x_2)$ by means of two dimensional thin-plate regression splines. However, this approach presents the drawback that only one smoothing parameter λ is incorporated into the smoothness penalty (i.e., an isotropic penalty), meaning that the same smoothness is assumed for both covariates x_1 and x_2. Whereas this isotropy could be justified when modelling, for instance, a smooth function of latitude and longitude, this is not always the case when the covariates x_1 and x_2 are measured in different units. As an alternative, the tensor product of one-dimensional spline basis functions has been suggested (see Eilers and Marx, 2003; Wood, 2006b). In this case, univariate functions f_1 and f_2 are associated with x_1 and x_2 respectively, each of which is represented by

a regression spline:

$$f_1(x_1) = \sum_{j=1}^{d_1} \theta_j^1 B_j^1(x_1) \quad \text{and} \quad f_2(x_2) = \sum_{k=1}^{d_2} \theta_k^2 B_k^2(x_2).$$

Accordingly, the bivariate surface is then defined as:

$$f_{12}(x_1,x_2) = \sum_{j=1}^{d_1}\sum_{k=1}^{d_2} \theta_{jk} B_j^1(x_1) B_k^2(x_2) = \sum_{j=1}^{d_1}\sum_{k=1}^{d_2} \theta_{jk} B_{jk}(x_1,x_2),$$

where the two-dimensional basis functions are given by the tensor product (see also Figure 4)

$$B_{jk}(x_1,x_2) = B_j^1(x_1) B_k^2(x_2).$$

Thus, model (11) can be expressed in matrix notation as:

$$\boldsymbol{y} = \mathbf{B}_{12}\boldsymbol{\theta}_{12} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\theta}_{12} = (\theta_{11}, \ldots, \theta_{d_1 1}, \ldots, \theta_{d_1 d2})^\mathsf{T}$ and

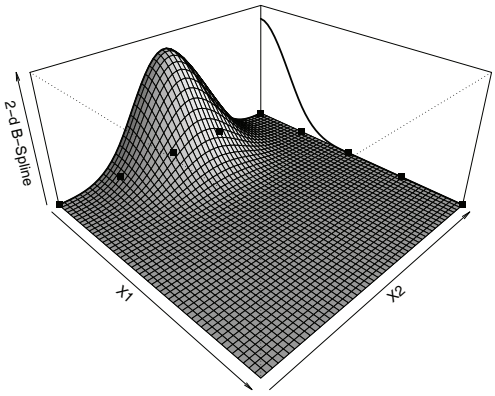
$$\mathbf{B}_{12} = \begin{pmatrix} B_{11}(x_{11},x_{12}) & \cdots & B_{d_1 1}(x_{11},x_{12}) & \cdots & B_{d_1 d_2}(x_{11},x_{12}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{11}(x_{n1},x_{n2}) & \cdots & B_{d_1 1}(x_{n1},x_{n2}) & \cdots & B_{d_1 d_2}(x_{n1},x_{n2}) \end{pmatrix}.$$

In this situation, how could be defined the penalty for the tensor product? First, it can be useful to view the vector of regression coefficients $\boldsymbol{\theta}_{12}$ as a two dimensional array:

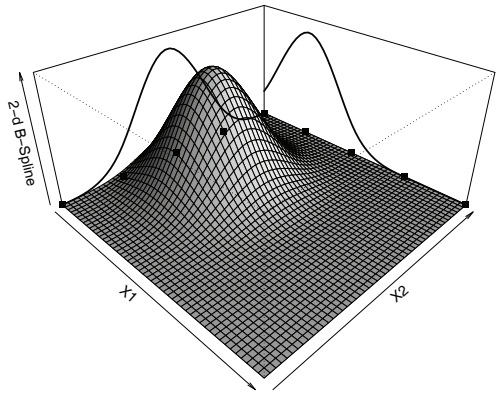
$$\boldsymbol{\Theta}_{12} = \begin{pmatrix} \theta_{11} & \cdots & \theta_{d_1 1} \\ \vdots & \ddots & \vdots \\ \theta_{1 d_2} & \cdots & \theta_{d_1 d_2} \end{pmatrix}.$$

Therefore, whereas the rows of $\boldsymbol{\Theta}_{12}$ correspond to the regression coefficients in the x_1 direction (see also Figure 4), the columns correspond to the x_2 direction. It then seems reasonable to consider the penalty for the bivariate surface separately for the rows and columns of $\boldsymbol{\Theta}_{12}$ (i.e., by considering separate penalties in the x_1 and x_2 directions). This leads to the penalty term in two dimensions (see Currie et al., 2006; Wood, 2006a, among others):

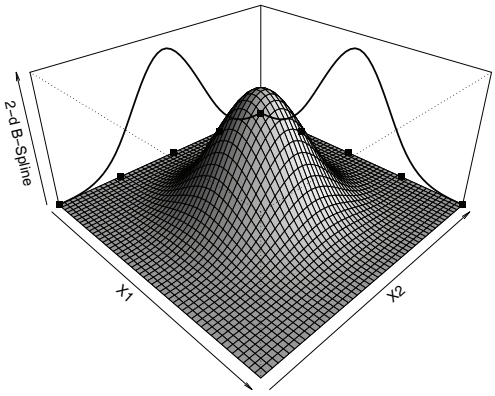
$$\lambda_1 \boldsymbol{\theta}_{12}^\mathsf{T} (\mathbf{I}_{d2} \otimes \mathbf{K}_1) \boldsymbol{\theta}_{12} + \lambda_2 \boldsymbol{\theta}_{12}^\mathsf{T} (\mathbf{K}_2 \otimes \mathbf{I}_{d1}) \boldsymbol{\theta}_{12},$$



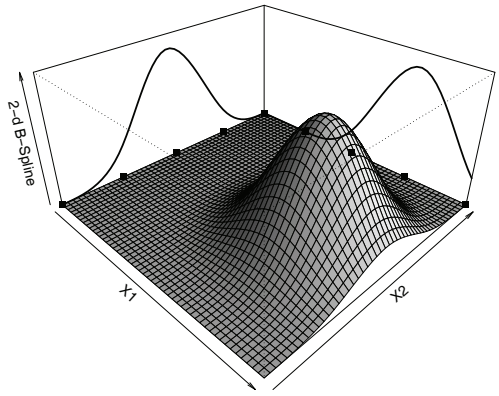
(a) $B_{24}(x_1,x_2)$



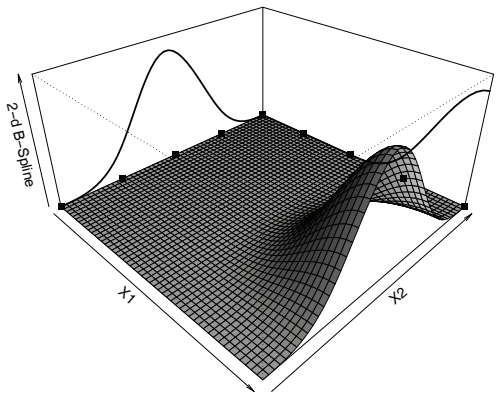
(b) $B_{34}(x_1,x_2)$



(c) $B_{44}(x_1,x_2)$



(d) $B_{54}(x_1,x_2)$



(e) $B_{64}(x_1,x_2)$

Figure 4: *Tensor product of univariate B-Spline basis functions of degree $p=3$ and $k=5$ knots. The figures show the tensor products obtained by varying the univariate B-Spline basis function along the x_1 direction $(B_j^1(x_1), j=2,\ldots,6)$ while maintaining fixed the B-Spline basis function in the x_2 direction $(B_4^2(x_2))$, i.e. $B_{j4}(x_1,x_2)=B_j^1(x_1)B_4^2(x_2), j=2,\ldots,6$.*

where $\otimes$ denotes the kronecker product, and $\mathbf{K}_1$ and $\mathbf{K}_2$ are univariate penalty matrices of dimension $d_1 \times d_1$ and $d_2 \times d_2$ respectively.

Parameters can now be estimated from the penalized least squares

$$||\boldsymbol{y} - \mathbf{B}_{12}\boldsymbol{\Theta}_{12}||^2 + \lambda_1 \boldsymbol{\theta}_{12}^{\mathsf{T}} (\mathbf{I}_{d2} \otimes \mathbf{K}_1)\, \boldsymbol{\theta}_{12} + \lambda_2 \boldsymbol{\theta}_{12}^{\mathsf{T}} (\mathbf{K}_2 \otimes \mathbf{I}_{d1})\, \boldsymbol{\theta}_{12},$$

or based on the mixed model representation.

3.4. Extension to the inclusion of higher order interactions

The methodology presented in Subsections 3.2 and 3.3 can also be extended to the inclusion of higher order interactions. For instance, in the analyses of visual receptive fields, we are interested on incorporating the time dimension to study the behaviour of the stimulus occurrences along pre-spike times. Thus, we are concerned on modelling a function of the form:

$$f_{12t}(x_1, x_2, t)\,.$$

Moreover, in order to evaluate the impact of different experimental conditions on cell response, we might need to incorporate this information into the regression model, and to allow for the previous three-dimensional function to vary across these experimental conditions ($l = 1, \ldots, M$):

$$\alpha + \sum_{l=2}^{M} \alpha_l \mathbf{1}_{\{z_i = l\}} + \sum_{l=1}^{M} \mathbf{1}_{\{z_i = l\}} f_{12t}^{l}(x_{i1}, x_{i2}, t_i)\,.$$

Based on the results presented in Subsection 3.3, $f_{12t}(x_1, x_2, t)$ (or, equivalently, $f_{12t}^{l}(x_{i1}, x_{i2}, t_i)$) can be represented by the tensor product of three univariate basis functions

$$f_{12t}(x_1, x_1, t) = \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} \sum_{m=1}^{d_t} \theta_{jkm} B_j(x_1) B_k(x_2) B_m(t) = \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} \sum_{m=1}^{d_t} \theta_{jkm} B_{jkm}(x_1, x_2, t)\,,$$

and the penalty term – organising the coefficients in some appropriate order – is now defined as

$$\lambda_1 \boldsymbol{\theta}_{12t}^{\mathsf{T}} (\mathbf{I}_{d_2} \otimes \mathbf{K_1} \otimes \mathbf{I}_{d_t})\, \boldsymbol{\theta}_{12t} + \lambda_2 \boldsymbol{\theta}_{12t}^{\mathsf{T}} (\mathbf{K_2} \otimes \mathbf{I}_{d_1} \otimes \mathbf{I}_{d_t})\, \boldsymbol{\theta}_{12t} + \lambda_t \boldsymbol{\theta}_{12t}^{\mathsf{T}} (\mathbf{I}_{c_2} \otimes \mathbf{I}_{c_1} \otimes \mathbf{K}_t)\, \boldsymbol{\theta}_{12t}.$$

4. Application to visual receptive fields

The dataset consists of a series of 16 matrices – with 256 grid positions each – of the form presented in Figure 1D, containing the number of stimulus occurrences at each position. Each matrix corresponds to the different pre-spike times considered (between -20 to -320 ms). The aim of the study was to obtain smooth RFmaps for a given pre-spike time, and to compare the obtained RFmaps between different experimental conditions. For illustration purposes, in the analyses presented in this paper we have considered the response to the onset of a bright spot ('ON') of two different cells (denoted by FAP0 and FBH4) of one monkey, and compared the RFmaps for the right and left eye.

4.1. Data analysis

We adopted a Poisson model with mean $n_{ijk}\lambda_{ijkt}$, where i indicates the row of the matrix, j the column ($i, j = 1, \ldots, 16$), k the eye (either left -0- or right -1-), and t the pre-spike time ($t = -20, \ldots, -320$). n_{ijk} denotes the number of stimulus presentations on each particular grid position. We considered different models for the intensity parameter (or firing rate) λ_{ijkt}, from the simplest model (and additive model) to the most complex (including the interaction between the row, column, time and eye). Specifically the following models were considered:

Model I

$$log\left(\lambda_{ijkt}\right) = \alpha_0 + \alpha_1 \mathbf{1}_{\{k=1\}} + f_{row}\left(i\right) + f_{col}\left(j\right) + f_{time}\left(t\right),$$

Model II

$$log\left(\lambda_{ijkt}\right) = \alpha_0 + \alpha_1 \mathbf{1}_{\{k=1\}} + f_{row,col}\left(i,j\right) + f_{time}\left(t\right),$$

Model III

$$log\left(\lambda_{ijkt}\right) = \alpha_0 + \alpha_1 \mathbf{1}_{\{k=1\}} + f_{row,col,time}\left(i,j,t\right),$$

Model IV

$$log\left(\lambda_{ijkt}\right) = \alpha_0 + \alpha_1 \mathbf{1}_{\{k=1\}} + \sum_{l=0}^{1} \mathbf{1}_{\{k=l\}} f^{l}_{row,col,time}\left(i,j,t\right),$$

The models were estimated using the package `mgcv`, version 1.7-9, in the `R` environment, version 2.14.0. In all cases, B-splines were used as spline basis functions, with degree $p = 3$ and penalty in the $q = 2$-nd order derivative, and tensor product smoothers

were used to model the multidimensional functions. For the univariate smooth functions, $k = 8$ knots placed at equidistant points were selected, whereas for the multidimensional case, $k = 3$ knots were chosen for each of the marginal basis functions. Different number of knots were also informally checked, but the results showed that the previous choices were enough. The smoothing parameters were automatically selected based on the REML criterion, and the models were compared in terms of the (conditional) AIC (cAIC).

4.2. Results

Tables 1 and 2 show the results for the selection of models of cells FBH4 and FAP0 respectively. As can be observed, the cAIC indicates that for cell FBH4 the best model is that including the spatio-temporal interaction and the categorical covariate eye (Model III). However, in the case of cell FAP0 the best model is the most complex model, i.e., that including the spatio-temporal interaction and the interaction with eye (Model IV). Table 3 presents a detailed description of these fitted models. In both cases, the firing rate depends on the eye, with left eye presenting more activity (p-value $<$ 0.001 in both cases). This indicates that these cells show dominance of the left eye.

Figure 5(a) shows the change over time of the standard deviation of the estimated firing rates of an ON-RFmap of a cortical visual cell (FBH4). The abscissa represents pre-spike time in ms. As it can be seen there are important changes. The RFmap begins

Table 1: *Results for the selection of models of cell FBH4.*

Model	Formula in `mgcv`	cAIC	edf	Deviance explained (%)
I	eye + s(row, bs = 'ps') + s(col, bs = 'ps') + s(time, bs = 'ps')	31981.67	13.13	3.12
II	eye + te(row, col, bs = 'ps') + s(time, bs = 'ps')	32002.94	20.00	3.24
III	eye + te(row, col, time, bs = 'ps')	31502.67	98.04	7.15
IV	eye + te(row, col, time, by =eye, bs = 'ps')	31518.41	164.76	7.88

Deviance explained (%): Percentage of the null deviance (null.dev - deviance of the model with just one constant term) explained by the fitted model, i.e. (null.dev - dev)/null.dev, where dev is the deviance of the fitted model.

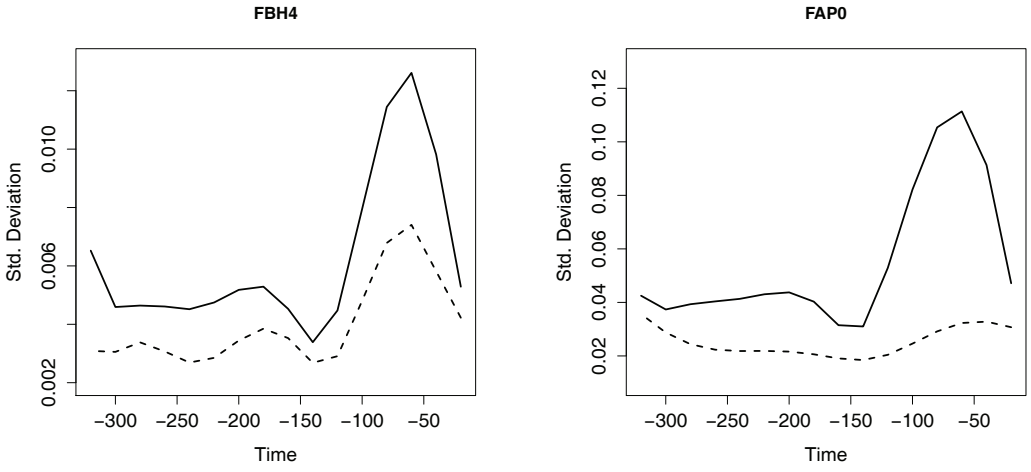
Table 2: *Results for the selection of models of cell FAP0.*

Model	Formula in `mgcv`	cAIC	edf	Deviance explained (%)
I	eye + s(row, bs = 'ps') + s(col, bs = 'ps') + s(time, bs = 'ps')	43248.77	12.04	16.1
II	eye + te(row, col, bs = 'ps') + s(time, bs = 'ps')	43207.59	21.61	16.6
III	eye + te(row, col, time, bs = 'ps')	42854.40	91.17	20.5
IV	eye + te(row, col, time, by =eye, bs = 'ps')	42831.73	138.94	21.5

s: smooth function
te: tensor product

Table 3: *Results of the fitted Models III and IV for cells FBH4 and FAP0 respectively.*

Term	Coefficient	edf	p-value
	Cell FBH4 - Model III		
Intercept	-3.448	—	< 0.001
eye_{Right}	-0.317	1	< 0.001
te(row, col, time)	—	96.04	< 0.001
	Cell FAP0 - Model IV		
Intercept	-0.485	—	< 0.001
eye_{Right}	-0.318	1	< 0.001
$\text{te(row, col, time)}_{Right}$	—	87.55	< 0.001
$\text{te(row, col, time)}_{Left}$	—	49.38	< 0.001



(*a*) (*b*)

Figure 5: *Change over time of the standard deviation of the estimated firing rates of the ON-RFmap of cells FBH4 and FAP0 for left (solid line) and right (dotted line) eye.*

to be structured about 140 ms pre-spike in both eyes and peaks approximately at 60 ms pre-spike, also in both eyes. Therefore the timing of the RFmap is similar for both eyes but the strength is different, indicating ocular dominance is constant over time. Figures 6 and 7 show a time series of ON-RFmaps obtained from a cortical visual cell (FBH4). As it can be seen, at about 200 ms pre-spike there is a central area with values below mean, and about 100 ms pre-spike a clear central area of values higher than the mean appears. This indicates that this area is the RF of the cell and that the optimal stimulus is a dot jumping from outside of the RF into the RF area.

Figure 5(b) is similar to Figure 5(a), however in this case the two cells studied show a different time-course of their RFmaps. The left eye stimulation produces a strong

RFmap with an onset at about 160 ms pre-spike and peaks at about 60 ms pre-spike. There is a clear dominance of the left eye, as it can be seen in Figures 8 and 9. Figure 8 shows the RFmaps of the non dominant eye (right eye). There is a very weak central area with almost no visible structure. Figure 9 shows the left ON-RFmap of the same cell shown in Figure 8. A clear central structure appears at 120 ms pre-spike and lasts until 20 ms prespike. The dominance of left eye induces stronger cell responses wich in turn produce stronger RFmap structures. As the cell of Figure 6, in this case there is a central area in the RFmaps from 220 to 180 ms prespike with low values, indicating that the cell response is better elicited when the spot jumps from outside the RF into the RF of the cell.

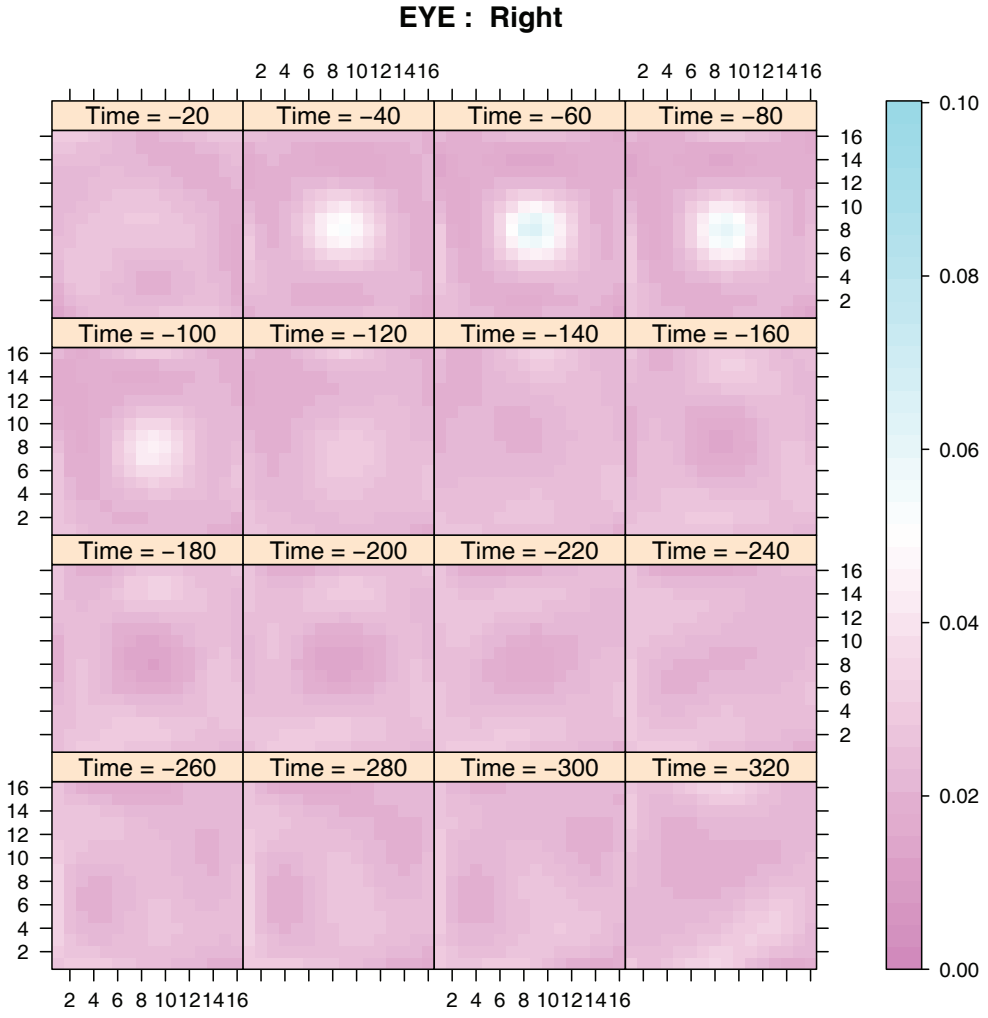


Figure 6: *Level plot of the estimated firing rates of the ON-RFmap based on Model III for the right eye of cell FBH4.*

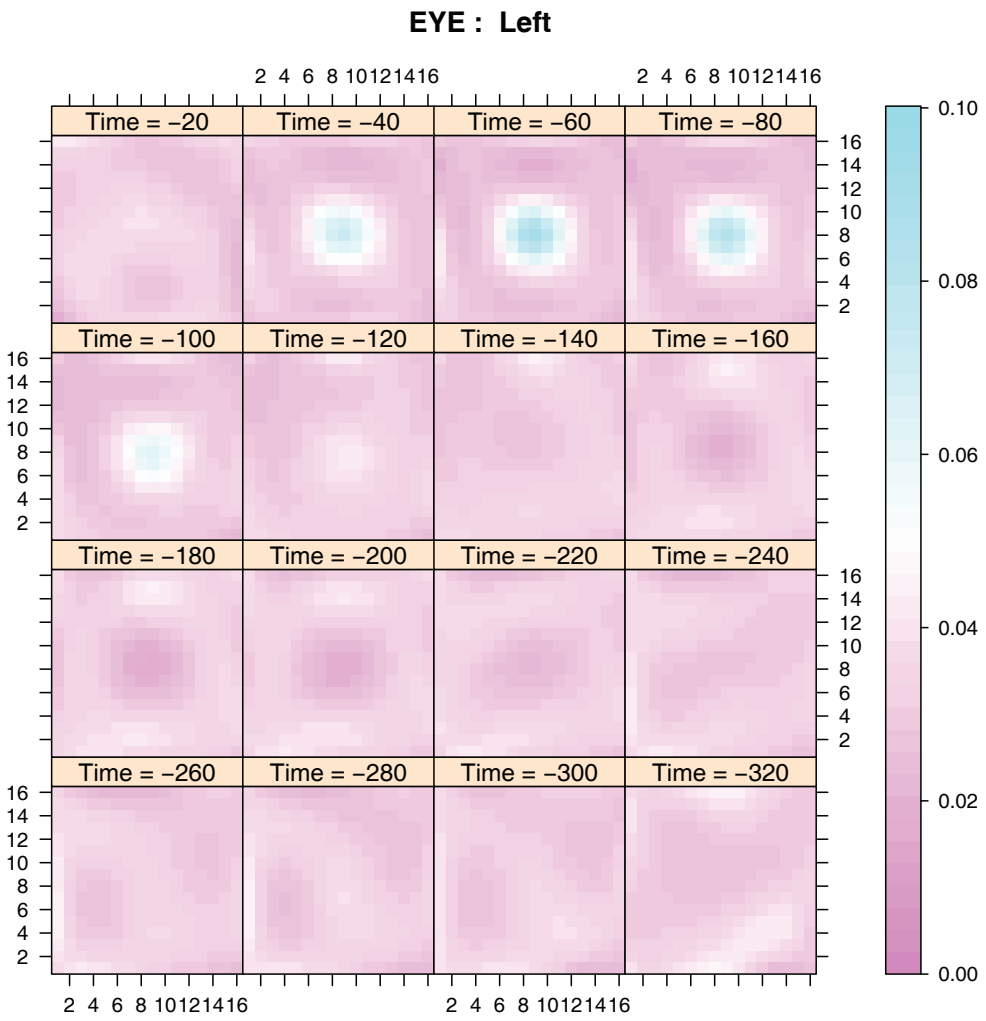


Figure 7: *Level plot of the estimated firing rates of the ON-RFmap based on Model III for the left eye of cell FBH4.*

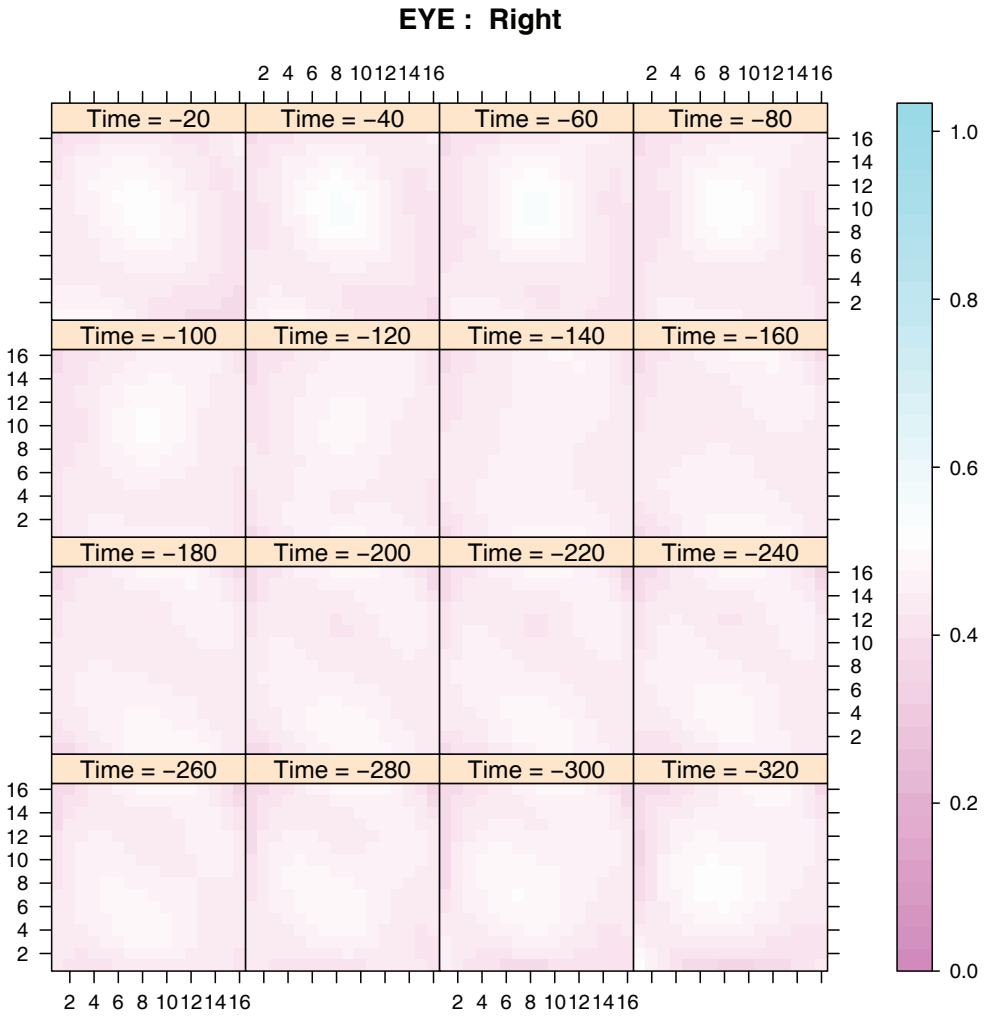


Figure 8: *Level plot of the estimated firing rates of the ON-RFmap based on Model IV for the right eye of cell FAP0.*

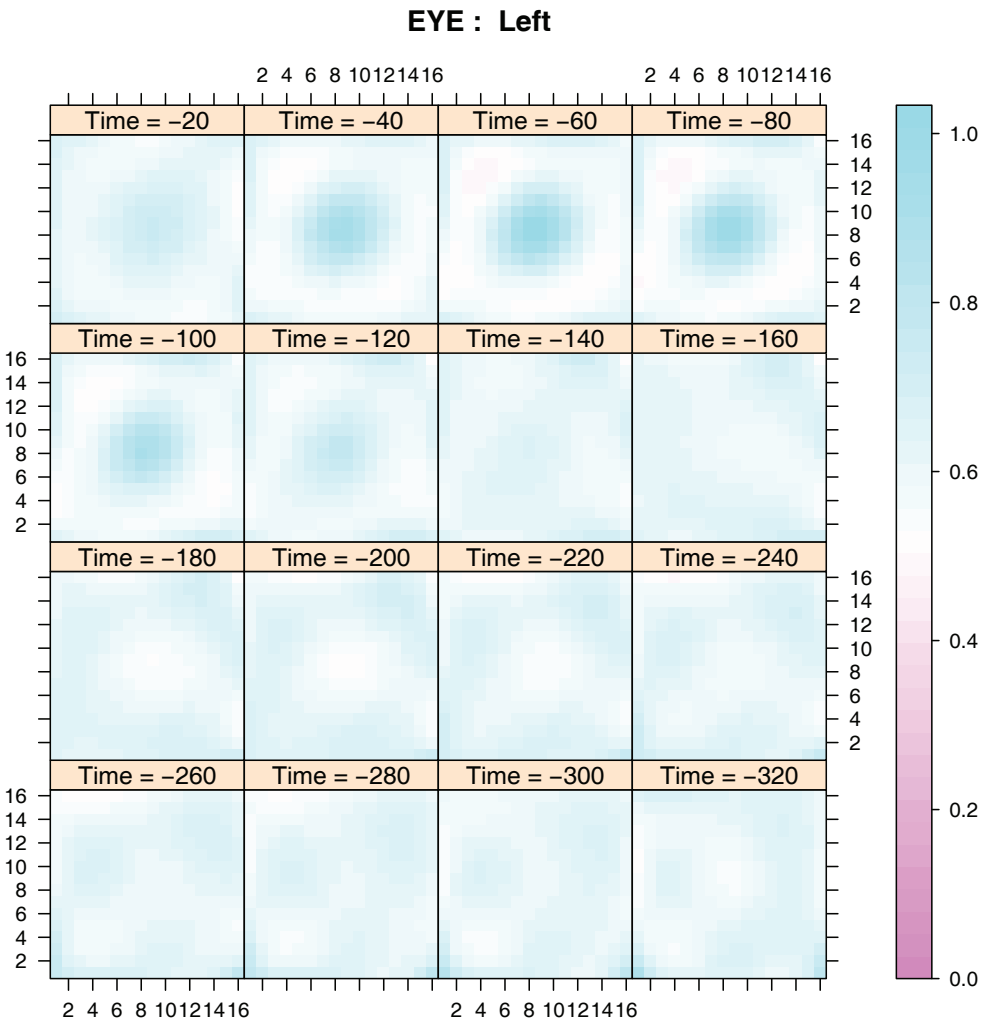


Figure 9: *Level plot of the estimated firing rates of the ON-RFmap based on Model IV for the left eye of cell FAP0.*

5. Conclusion

Receptive field mapping of cortical visual neurons is a critical procedure in neurophysiological studies of the visual system. In electrophysiological laboratories the reverse cross-correlation technique is frequently used to produce raw RFmaps – in the form of numerical matrices – which allow to give insights into the electrical activity of the visual neurons. However, the resulting RFmaps obtained from applying such technique, cannot be compared formally among different conditions.

In this paper, a Poisson GAM including interactions was proposed as a flexible statistical tool for (a) smoothing raw RFmaps of a single neuron, by including spatial effects in the model; and (b) estimating the temporal evolution of these maps, which in turn may vary across different experimental conditions.

The estimation algorithm used for fitting GAMs was based on P-splines and tensor product splines and the amount of smoothing was selected automatically via restricted (or residual) maximum likelihood (REML).

The proposed methodology was applied to study the activity of visual cells from area V1 (a primary visual cortical area). For the sake of illustration, in all the analyses perfomed in this paper, only visual neurons with response to the onset of a bright spot ('ON') were considered, and the temporal evolution of their RFmaps was compared for the right and left eye. However, it should be noted that our GAM methodology is very flexible and extensions to more complex interaction models are straightforward. This is important since physiologists are also interested in examining possible variations on RFmaps, when several levels of other covariates are considered. For example, comparisons between both monocular RFs and between monocular and binocular RFs are often required to assess the question on how the visual system handles binocular information. Therefore, GAM regression models designed to perform statistical comparisons between RFmaps would be very useful in electrophysiological experiments conducted in the visual system.

When calculating the time span of any RFmap, the onset and offset of the RFmap structures must be determined. In this paper, a relatively simple approach to solve this problem was proposed, and it consists of using the change of standard deviation of the matrix values over time. The authors are concisious, however, that a more refined analysis is needed, to extract accurate information from such RFmaps structures. Although it is beyond the scope of the current paper, it is an important topic for further research.

An `R` script implementing the nonparametric model estimation can be obtained by contacting the first author at maria.jose.rodriguez.alvarez2@sergas.es.

Acknowledgements

The authors would like to express their gratitude for the support received in the form of the MICINN (Spanish Ministry of Science and Innovation) grants MTM2008-01603 and DE2009-0030. Work of María Xosé Rodríguez-Álvarez was supported by grant CA09/0053 from the Instituto de Salud Carlos-III (Spanish Ministry of Science and Technology). Work of Francisco Gonzalez was supported by grant BFU-2010-14968 from the MCINN, Spain, FEDER, and 2010/PX142 from the Xunta de Galicia, Spain.

References

Barlow, H.B. (1953). Summation and inhibition in the frog's retina. *The Journal of Physiology*, 119, 69–88.

Behseta, S. and Kass, R. E. (2005). Testing equality of two functions using BARS. *Statistics in Medicine*, 24, 3523–3534.

Behseta, S., Wallstrom, G. L. and Kass, R. E. (2005). Hierachical models for assessing variability among functions. *Biometrika*, 92, 419–434.

Bellman, R. E. (1961). *Addaptive Control Processes*. Princenton University Press.

Bishop, P. O., Coombs, J. S. and Henry, G. H. (1973). Receptive fields of simple cells in the cat striate cortex. *The Journal of Physiology*, 231, 31–60.

de Boor, C. A. (2001). *A Practical Guide to Splines.* Revised Edition. Springer-Verlag, New York.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed model. *Journal of the American Statistical Association*, 88, 9–25. 116.

Brezger, A. and Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, 50, 967–991.

Brumback, B. A., Ruppert, D. and Wand, M. P. (1999). Variable selection and function estimation in additive nonparametric regression using a data-set prior: Comment. *Journal of the American Statistical Association*, 94, 794–797.

Cadarso-Suárez, C., Roca-Pardiñas, J., Molenberghs, G., Faes, C., Nácher, V., Ojeda, S. and Acuña, C. (2006). Flexible modelling of neuron firing rates across different experimental conditions. An application to neural activity in the prefrontal cortex during a discrimination task. *Applied Statistics*, 55, 431–447.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31, 377–403.

Currie, I. D. and Durbán, M. (2002). Flexible smoothing with P-splines: a unified approach. *Statistical Modelling*, 2, 333–349.

Currie, I. D., Durbán, M. and Eilers, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, Series B*, 68, 1–22.

DeAngelis, G. C., Ohzawa, I. and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69, 1091–1117.

DeBoer, E. and Kuyper, P. (1968). Triggered correlation. *IEEE Transactions on Biomedical Engineering*, 15, 169–179.

DiMatteo, I., Genovese, C. R. and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, 88, 1055–1071.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121.

Eilers, P. H. C. and Marx, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems*, 66, 159–174.

Faes, C., Geys, H., Molenberghs, G., Aerts, M., Cadarso-Suárez, C., Acuña, C. and Cano, M. (2008). A Flexible Method to Measure Synchrony in Neuronal Firing. *Journal of the American Statistical Association*, 103, 149–161.

Fried, J. and Silverman, B. (1989). Flexible parsimonious smoothing and additive modelling. *Technometrics*, 31, 3–21.

Gonzalez, F., Krause, F., Perez, R., Alonso, J. M. and Acuña, C. (1993). Binocular matching in monkey visual cortex: single cells responses to correlated and uncorrelated dynamic random dot stereograms. *Neuroscience*, 52, 933–939.

Gonzalez, F. and Krause, F. (1994). Generation of Dynamic Random element stereograms in real time with a system based on a personal computer. *Medical and Biological Engineering and Computing*, 32, 373–376.

Gonzalez, F., Perez, R., Justo, M. S. and Bermudez, M. A. (2001). Receptive field organization of disparity-sensitive cells in macaque medial superior temporal cortex. *European Journal of Neuroscience*, 14, 167–173.

Hartline, H. K. and Ratliff, F. (1958). Spatial summation of inhibitory of inhibitory influences in the eye of Limulus, and the mutual interaction of receptor units. *The Journal of General Physiology*, 41, 1049–1066.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability. Chapman & Hall.

Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, 55, 757–796.

Henry, G. H. and Bishop, P. O. (1972). Striate neurons: receptive field organization. *Investigative Ophthalmology*, 11, 357–368.

Hubel, D. H. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160, 106–154.

Jones, J. P. and Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58, 1187–1211.

Kass, R. E., Ventura, V. and Cai, C. (2003). Statistical smoothing of neuronal data. *Network: Computation in Neural Systems*, 14, 5–15.

Krause, F., Gonzalez, F., Nelson, J. I. and Eckhorn, R. (1987). A fast method for predicting coding properties of visual cortical simple cells. *Perception*, 16, 2–269 (A45).

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16, 37–68.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13, 183–212.

Lee, T. C. M. (2002). On algorithms for ordinary least squares regression spline fitting: a comparative study. *Journal of Statistical Computation and Simulation*, 72, 647–663.

Lee, D.-J. and Durbán, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, 11, 49–69.

Marra, G. and Radice, R. (2010). Penalised regression splines: theory and application to medical research. *Statistical Methods in Medical Research*, 19, 107–125.

Marx, B. D. and Eilers, P. C. H. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis*, 28, 193–209.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Second Edition. Chapman & Hall/CRC.

Nishimoto, S., Ishida, T. and Ohzawa, I. (2006). Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation. *The Journal of Neuroscience*, 26, 3269–3280.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1, 502–518.

Pei, X., Vidyasagar, T. R., Volgushev, M. and Creutzfeldt, O. D. (1994). Receptive field analysis and orientation selectivity of postsynamptic potentials of simple cells in cat visual cortex. *The Journal of Neuroscience*, 14, 7130–7140.

Pérez, R., Castro, A. F., Justo, M. S., Bermúdez, M. A. and Gonzalez, F. (2005). Retinal correspondence of monocular receptive fields in disparity-sensitive complex cells from area V1 in the awake monkey. *Investigative Ophthalmology & Visual Science*, 46, 1533–1539.

R Development Core Team. R: A language and environment for statistical computing (2011). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reid, R. C., Victor, J. D. and Shapley, R. M. (1997). The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Visual Neuroscience*, 14, 1015–1027.

Reiss, P. T. and Ogden, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society, Series B*, 71, 505–524.

Roca-Pardiñas, J., Cadarso-Suárez, C., Nácher, V. and Acuña, C. (2006). Bootstrap-based methods for testing factor-by-curve interactions in generalized additive models: assessing prefrontal cortex neural activity related to decision-making. *Statistics in Medicine*, 25, 2483–2501.

Roca-Pardiñas, J., Cadarso-Suárez, C., Tahoces, P. G. and Lado, M. J. (2008). Assessing continuous bivariate effects among different groups through nonparametric regression models: An application to breast cancer detection. *Computational Statistics and Data Analysis*, 52, 1958–1970.

Roca-Pardiñas, J., Cadarso-Suárez, C., Pardo-Vazquez, J. L., Leboran, V., Molenberghs, G., Faes, C. and Acuña, C. (2011). Assessing neural activity related to decision-making through flexible odds ratio curves and their derivatives. *Statistics in Medicine*, 14, 1695–1711.

Strasak, A. M., Umlauf, M., Pfeiffer R. M. and Lang, S. (2011). Comparing penalized splines and fractional polynomials for flexible modelling of the effects of continuous predictor variables. *Computational Statistics and Data Analysis*, 55, 1540–1551.

Tsao, D. Y., Conway, B. R. and Livingstone, M. S. (2003). Receptive fields of disparity-tuned simple cells in macaque V1. *Neuron*, 38, 103–114.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society, Series B*, 62, 413–428.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society, Series B*, 65, 95–114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter for generalized additive models. *Journal of the American Statistical Association*, 99, 673–686.

Wood, S. N. (2006a). *Generalized Additive Models. An Introduction with R*. Chapman & Hall/CRC.

Wood, S. N. (2006b). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62, 1025–1036.

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society, Series B*, 70, 495–518.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society, Series B*, 73, 3–36.

Discussion of "Analysing visual receptive fields through generalised additive models with interactions" by María Xosé Rodríguez-Álvarez, Carmen Cadarso-Suárez and Francisco Gonzalez

María L. Durbán
Universidad Carlos III de Madrid

It is a pleasure for me to comment on the paper by Rodriguez-Alvarez, Cadarso-Suarez and Gonzalez. I would like to thank the editors for their invitation. This paper proposes a Poisson generalized additive model (GAM) for smoothing visual receptive fields (RFs) over time, and uses multidimensional interactions to compare data under several experimental conditions. As the authors point out, the use of high-dimensional smoothing for the analysis of spatial and spatio-temporal data has been a subject of great interest in recent years (specially in the context of P-splines). Also, the possibility of fitting a factor by curve interaction has made these type of models even more flexible, and capable of capturing very complex structures present in the data (Durbán et al. (2005)). The authors go one step further and use multidimensional interactions, and use a unified approach for: i) spatio-temporal modelling, and ii) comparison across different experimental conditions. The models proposed and the results obtained in the paper unavoidably give rise to several related and interesting issues, which may help to improve and extend the scope of the models and their application.

One of the main problems encountered when multidimensional smoothing is used is the computational time that the fitting of the model might need. P-splines, and in general, low-rank smoothers, are the best approach to take. However, even in this case, great care needs to be taken when choosing the size of the bases used in the model. The authors have, obviously, come across this problem. I am concerned with their choice of size for the bases. First, the fact that the size of the bases is different in the univariate case ($k = 8$), and in the multivariate case ($k = 3$) result in models that are not nested (then, care must be taken when comparing them). To ensure that they are, bases for multidimensional smooth terms (models II, III and IV) should be constructed using marginal bases identical to those used for the main effects (model I). But, sometimes, using bases of the same size results in models that can be computationally very demanding, or even, impossible to fit. One possible solution is to reduce the number of parameters for the interaction term, by reducing the number of knots for the marginal bases for the tensor product, but preserving the nested nature of the models. This idea can be explained by analogy to classical ANOVA models, where, in general, the main effects are more significant than interactions. Lee and Durbán (2012) propose the use of nested B-spline bases for the interaction term: a B-spline basis such that the space spanned by the marginal basis in the interaction, is a subset of the space spanned by the basis in the unidimensional case and the hierarchical nature of the models is preserved. Another point related to the bases size is the fact that, using three knots for marginal bases might

not be enough to capture the structure of the firing rates. I agree that larger bases are difficult to handle by the package `mgcv` (specially when the data have an array structure). However, as a general rule (Ruppert, 2002), the minimum number of knots used to construct a basis is at least four. The results for generalized linear array models (GLAMs), in Eilers et al. (2006), can speed up the computation when the size of the bases increases.

A major contribution of the paper is the development of a model to compare the receptive fields properties by introducing interaction between factor and multidimensional smooth terms, and check, for example, if the firing rate depends on eye. Models III and IV test if the spatio-temporal effects are similar in both eyes or not, but do not separate the spatial and temporal component. It might be interesting to know if the change of firing rates over time is the same for both eyes, or to estimate a common spatial pattern in time and check if a space-time interaction is necessary. Maybe an ANOVA-type model (Lee and Durbán (2011)) could shed light on some interesting issues. Thus, the spatio-temporal term in model III and IV could be substituted by:

$$f_{\text{row, col, time}}(i,j,t) \Rightarrow f_{\text{time}} + f_{\text{row, col}}(i,j) + f_{\text{row, col, time}}(i,j,t),$$

and appropriate penalties and identifiability constraints can be easily imposed.

Let me finish by congratulating again the authors for an interesting and relevant paper for the statistical and medical scientific community.

References

Durbán, M., Harezlak, J., Wand, M. P. and Carroll, R. J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, 24, 1153–1167.

Eilers, P. H. C., Currie, I. D. and Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, 50, 61–76.

Lee, D-J. and Durbán, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, 11, 49–69.

Lee, D-J. and Durbán, M. (2012). Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases (submitted).

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11, 735–757.

Acknowledgements

The work of Maria Durbán is supported by the the Spanish Ministry of Science and Education (projects MTM 2008-02901 and MTM2011-28285-C02-02).

Thomas Kneib
Georg-August-Universität Göttingen

Thank you very much for the possibility to discuss this paper that introduces and discusses additive model including different types of interactions along the analysis of visual receptive. The authors do a remarkable job in reviewing the current state of the literature and in demonstrating the applicability of the methods in a complex setting that requires careful inclusion of interaction effects in different ways. Penalized spline smoothing forms the basis for the models considered and mixed model based inference for the smoothing parameters yields a data-driven amount of smoothness for all functions. My comments mainly focus on the use of mixed model methodology (and potential alternatives) and some specific modelling choices.

- As one advantage of the mixed model approach, the authors discuss the possibility of a formal comparison across models of different complexity. However, inference in mixed models has proved to be notoriously difficult due to deviations from regularity conditions underlying standard likelihood based inferential approaches. For example, likelihood ratio tests for the inclusion / exclusion of specific effects approached via testing smoothing variances for deviations from zero have been shown to lead to rather complex distributions for the test statistic that deviate considerably from the standard result that would be a χ^2 distribution with one degree of freedom (Crainiceanu and Ruppert 2004, Crainiceanu, Ruppert, Claeskens and Wand 2005, Scheipl, Greven and Küchenhoff 2008). In a similar vein, Greven and Kneib (2010) show that both marginal AIC and the conditional AIC (in their standard forms) are problematic when comparing models that differ by the inclusion / exclusion of specific effects obtained by setting the smoothing variance to zero. While this is not exactly the type of model comparison considered here, I wonder how generally applicable formal comparisons in the considered model class can be and how the authors suggest to deal with such problems.
- A related question concerns the performance of model choice and the p-values presented in Table 3. Given the difficulties discussed in the previous comment, more guidelines on model choice and details on how the p-values in Table 3 have been computed would be very helpful from the perspective of practitioners.
- When using varying coefficient terms for modelling interactions, the coding of the (categorical) interaction variable will strongly influence the results. For example, the model

$$y_i = \sum_{l=1}^{M} \mathbb{1}_{z_i=l} f_l(x_i) + \varepsilon_i$$

models separate curves per each of the M levels of the categorical covariate z while

$$y_i = f_1(x_i) + \sum_{l=2}^{M} \mathbb{1}_{z_i=l} g_l(x_i) + \varepsilon_i$$

models the effect in category 1 as $f_1(x)$ and deviations from this effect in remaining groups $l = 2,\ldots,M$ as $g_l(x)$. While both models would be equivalent when restricting $f_l(x)$ and $g_l(x)$ to be linear, this does no longer hold in the penalised spline smoothing context due to the estimation of the smoothing parameter that is not invariant under such transformations. In addition, when using effect coding instead of dummy coding for the categorical covariates, another set of results would be obtained. How would the authors deal with this difficulty and what would be their general recommendations for choosing a specific parameterisation?

- I wonder whether (spatial) smoothing is really desired in this application. I would presume that there may be sharp edges around the receptive fields while the given approach assumes a smooth transition from active to inactive cells. As a consequence, edge-preserving or adaptive procedures may be more appropriate at least for the effect representing the receptive field. In a Bayesian formulation, adaptiveness could for example be achieved by making the smoothing variance depending on the spatial location (see Lang and Brezger (2004) for such an approach).
- For the spatial effect, radial basis functions may also be a useful alternative to the bivariate P-splines proposed here since P-splines may induce some artificial structure in the estimates due to their inherent non-radiality. Although the amount of non-radiality is decreasing with the spline degree and may therefore be negligible for the cubic splines considered here, it would be good to have some general advice on the selection of either radial bases or bivariate penalized splines.

References

Crainiceanu, C. and Ruppert, D. (2004). Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis*, 91, 35–52.

Crainiceanu, C., Ruppert, D., Claeskens, G. and Wand, M. P. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika*, 92, 91–103.

Greven, S. and Kneib, T. (2010). On the behavior of marginal and conditional Akaike information criteria in linear mixed models. *Biometrika*, 97, 773–789.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13, 183–212.

Scheipl, F., Greven, S. and Küchenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis*, 52, 3283–3299.

Rejoinder

First of all, we would like to thank the invited discussants for the time spent discussing our work and for all their valuable comments and sugggestions made on our paper.

1. Comments from Prof. Thomas Kneib

Prof. Thomas Kneib primarily centres his comments on various problems arising in the context of the inference in Generalised Additive Models. Although Prof. Kneib mainly focuses his comments on the mixed model methodology, similar difficulties can be encountered as well using different approaches and smoothing parameters selectors. We agree with the discussant that this is an area of enormous interest and active research, not yet completely solved. We are conscious that our paper lacks an explicit discussion of this point, and the use of the word 'formal' probably overstates the main objective of the paper. We would like to thank Prof. Kneib for his valuable comments on this issue and for giving us the opportunity of shedding more light on this challenging point. However, we believe that this issue probably requires a new paper.

With respect to model comparisons and model choice (in the frequentist framework), in the context of GAMs (and GLMs), two different approaches are usually applied, depending on if the models are nested or not (see also the comments from Prof. María Durbán). If the models are not nested, the comparison between models can be based on some information criteria such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). If the models are nested, the generalised likelihood ratio test or the F-ratio test can be applied. However, in GAMs the distribution of these test statistics under the null hypothesis is only approximate (due to the fact that the smoothing parameters are treated as if they are known), and the obtained p-values should be analised with caution (specially when they are close to the significance level) (Wood, 2006a). As pointed out by Prof. Simon Wood (in the help file of the functions `mgcv::summary.gam` and `mgcv::anova.gam`), various simulations studies suggested that the p-values obtained under ML and REML smoothness selection, have the best behaviour. This cautionary note can also be extended to the inference about each smooth term in a model, and the way in how p-values are computed. In this case, two different approaches can be used, based on the frequentist or the Bayesian covariance matrix of the coefficient estimates (Wood, 2006a). If the objective is testing for smooth terms in a GAM for equality to zero, the frequentist approach can be used (see Marra and Radice, 2010; Wood, 2006a). However, even in this case the performance of the

obtained p-values is not good, in terms of type I error. The Bayesian approach is then recommended in this case (as for the construction of the 'confidence' intervals; see the help of the `mgcv` package). Therefore, the p-values reported in this paper were based on the bayesian covariance matrix. Again, the (Bayesian) p-values obtained under ML and REML smoothness selection present the best behaviour. This is in accordance with the results of a small simulation study performed by us, in which type I errors (using REML) proved to be relatively close to nominal errors.

Under the representation of an additive model as a linear mixed model, some other alternatives (and difficulties) for model choice appear. In this framework, the random effects u parameterise the deviation of the smooth function from a given polynomial (depending on the degree p of the spline basis). Therefore, testing for a polynomial (or constant) function versus a smooth function is equivalent to test if the corresponding smoothing variance is zero ($\sigma_u^2 = 0$). As pointed out by Prof. Kneib, this field has received a considerably amount of research attention in the last years. For instance, in Scheipl et al. (2008) a comparative study of different restricted likelihood ratio tests (RLRT) and F-type tests for a zero variance (also in the context of penalised splines) is presented. The authors conclude that the RLRT statistic proposed by Crainiceanu and Ruppert (2004) – and extended by Greven et al. (2008) to test for more than one variance component – presents the best behaviour in terms of the computational time, with type I error rates and power almost equivalent to the bootstrap-based tests. Moreover, this test is implemented in the `R` package `RLRSim`. However, for the model comparisons performed in this paper, the use of the RLRT statistics present in this package is not possible unless we consider nested models and isotropic interactions (see Lee 2010), and, moreover, these tests may not be applicable to GAM. As regards the use of the AIC for comparison purposes, under the mixed model representation, two different alternative definitions can be used: the marginal and the conditional AIC. The conditional AIC (in its standard form) is the AIC supplied by the `mgcv` package, and its use is 'recommended' in the context of penalised splines. However, Greven and Kneib (2010) showed that the standard form of the conditional AIC has an undesirable performance, in the sense that it always chooses the inclusion of the random effect u, unless u is predicted to be exacly zero ($\hat{\sigma}_u^2 = 0$). Although the comparisons made in Greven and Kneib (2010) – where the authors compare the presence of a linear effect versus a smooth effect – are not the type of model comparison considered in our paper, it would be worthwhile evaluating the performance of the conditional AIC in this setting. Moreover, the extension of the corrected version of the conditional AIC Liang et al. 2008, Greven and Kneib 2010) to generalised linear mixed models and GAMs is is an open area of research.

The comment from Prof. Kneib related to the use of varying cofficient terms is connected to this previous issue. From an applied point of view, the decision about the coding of the categorical variable that should be used depends on both the problem at hand and the research question to be answered. On the one hand the use of the dummy coding allows to separately evaluate the effect of the continuous covariate in

each of the levels of the categorical covariate. However, this coding makes it difficult to 'visually' judge the presence of interaction from several curves that are estimated separately. On the other hand, when the interest is focused on finding out whether there actually is a difference among several groups, it may be advantageous to use the effect parameterisation ($y_i = f(x_i) + \sum_{l=1}^{M} \mathbf{1}_{\{z_i=l\}} g_l(x_i) + \varepsilon_i$, given that $\sum_{l=1}^{M} g_l(x_i) = 0$). In this case, the deviations from the main effect, $f(x_i)$, in each of the levels of the categorical covariate, $g_l(\cdot)$ ($l = 1, \ldots, M$), exactly tells us the amount of difference. Moreover, this approach also allows to evaluate the presence of interaction based on the inference about $g_l(\cdot)$, ($l = 1, \ldots, M$). If these functions are equal to zero, it implies that no interaction is present. The dummy coding is the way in which the `mgcv` package treats, by default, the interaction between continuous and categorical covariates, and this was the coding used in the paper. However, the use of the effect coding can also be implemented in the `mgcv` package, by first creating the effect coding variable. For illustration purposes, we have analysed the visual receptive field data (Model IV) using this approach:

```
R > fbh4$eff_eye <- as.numeric((fbh4$eye == 'Right') - (fbh4$eye == 'Left'))
R > fit.fbh4 <- gam(spikes ~ eye + te(row, col, time, by = eff_eye, bs = 'ps')
+   + offset(log(trial)), data = fbh4, family = poisson, method = 'REML')
```

and

```
R > fap0$eff_eye <- as.numeric((fap0$eye == 'Right') - (fap0$eye == 'Left'))
R > fit.fap0 <- gam(spikes ~ eye + te(row, col, time, by = eff_eye, bs = 'ps')
+   + offset(log(trial)), data = fap0, family = poisson, method = 'REML')
```

Table 1: *Results of the fitted models for cells FBH4 and FAP0.*

Term	Coefficient	edf	p-value
Cell FBH4 - Model IV			
Intercept	-3.466	—	< 0.001
eye_{Right}	-0.283	1	< 0.001
f(row, col, time)	—	96.00	< 0.001
g_{Right}(row, col, time)	—	20.13	0.107
Cell FAP0 - Model IV			
Intercept	-0.480	—	< 0.001
eye_{Right}	-0.330	1	< 0.001
f(row, col, time)	—	88.83	< 0.001
g_{Right}(row, col, time)	—	24.86	< 0.001

Table 1 presents a detailed description of these fitted models. It should be noted that in this case, due to the identifiability constraints imposed on the model, $g_{Right}(\cdot) = -g_{Left}(\cdot)$. As can be observed, whereas for cell FBH4 there is no evidence to suggest the presence of interaction with eye (p-value = 0.107), for cell FAP0 the results suggest that the RFmap is different in the left and right eye (p-value < 0.001). This is in

concordance with the results obtained using the dummy coding and the cAIC for comparison purposes.

As regards the use of radial basis for the spatial effect, we agree with Prof. Kneib that this approach could have been also used for the visual receptive field analysis in our paper. In contrast to tensor product smoothers, where the knots are placed separately in each direction (row and column in our example), radial bases (as thin plate splines, for example) consider bivariate knots located in the surface. Therefore, one of the main advantages of radial bases is that they allow to adapt the placement of knots to the data (i.e., placing more knots in areas with dense data), and therefore to take into account the possible correlation between the covariates (see Fahrmeir and Kneib 2011 for a detailed comparison between tensor product and radial basis). However, in contrast to tensor product smoothers, radial bases assume the same amount of smoothing in each direction. As pointed out in the paper, this isotropy could be justified when the covariates are measured in the same units (and, of course, the same amount of smoothing is expected in each direction). This is the case, for instance, of spatial effects, but not of the spatio-temporal interaction, due to the different scaling of time and space. For this reason, the use of radial bases is not recomended in this context. However, the `mgcv` package allows to combine in a trivariate smoother, the tensor product of a bivariate radial basis and a one dimensional smoother. We would like to point out that the visual receptive field data was also analised using this approach. Since the results provided by this model and the model using the tensor product of univariate smoothers were pretty similar, we decided to present in the paper the results of the trivariate tensor product.

Finally, as pointed out by Prof. Kneib the use of adaptive penalized splines could be more appropiate for this application, as can be observed from the figures. So far, the `mgcv` package allows the use of locally adaptive procedures but in the context of radial bases, and cannot be applied for tensor product smoothers. Although, as pointed out before, this `R` package allows to combine the tensor product of a bivariate radial basis and a one dimensional smoother, it is not possible to use adaptive smoothers for the bivariate radial basis. Moreover, as far as we know, no statistical software implements the use of locally adaptive procedures for trivariate smoothers. This is an interesting field for further research.

2. Comments from Prof. María Durbán

Prof. María Durbán comments on the importance of the selection of the basis sizes, due to its impact on: (a) the final estimates (if the basis dimension is not large enough), (b) the computational time; and (c) the inference procedures.

As pointed out by María Durbán, from a computational point of view, as the number of knots increases (specially in the multidimensional case), so it does the computational time. For instance, using an Intel(R) Core (TM) i5 CPU 2.40 GHz and 4.00 GB RAM computer, the estimation of model IV using $k = 3$ or $k = 4$ was around 100 sec. and 403

sec. respectively. For this reason, we chose $k = 3$ knots for each of the marginal basis functions in the multidimensional case. However, in order to evaluate the adequacy of the basis dimensions, we performed a sensitivity analysis by increasing the number of knots, and comparing the results obtained by the different fitted models. Since in this case the observed differences were almost negligible, we finally selected $k = 3$ knots. Moreover, we also evaluated the adequacy of the basis dimensions based on an analysis of the deviance residuals of the fitted model (see the help associated with the function `choose.k` in the `mgcv` package). For each smooth term in our models (I- IV), we fitted an equivalent, single, smooth to the residuals, using a larger number of knots to see if there were still a pattern in the residuals that could potentially be explained by increasing the basis dimension in the original models. Again, these analyses suggested that $k = 8$ and $k = 3$ were enough for the univariate and the multidimensional cases respectively. However, we are conscious that this could not be the always the case, and in some circumstances, the need of larger bases could make prohibitive the fit of the model using the package `mgcv`. As pointed out by María Durbán, when the data are in an array structure, as in the case of the visual receptive field data, the use of GLAMs can make the fit of the model feasible in a reasonable computing time. As far as we know, neither the `mgcv` package nor other `R` packages and statistical software implement this approach. This is, indeed, an interesting field for further work.

As Prof. Durbán indicates, the use of different sizes of the bases for the univariate and the multivariates case, results in models that are not nested. Accordingly, it is not possible to apply formal testing procedures, as the generalised likelihood ratio test or the F-ratio test, to make the comparison between alternative models. Although in this paper we have used the conditional AIC for comparison purposes, the use of formal formal testing procedures could be also an alternative approach, ensuring that the models are nested. Moreover, the ANOVA-type models proposed by Lee and Durbán (2011) is an interesting approach to study in future. We completely agree with Prof. Durbán that the use ANOVA-type models could help in a better understanding of the Visual receptive field behaviour. It should be noted, however, that in the visual receptive field data, the experiment design assigns each spike to the stimulus position in all pre-spike times. It implies that each spatial matrix contains exactly the same number of spike ocurrences. Therefore, the spatio-temporal interaction is probably needed in this context.

References

Crainiceanu, C. and Ruppert, D. (2004). Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis*, 91, 35–52.

Fahrmeir, L. and Kneib, T. (2011). *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxforf University Press.

Greven, S., Crainiceanu, C. M. and Küchenhoff, H. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics*, 17, 870–891.

Greven, S. and Kneib, T. (2010). On the Behavior of Marginal and Conditional Akaike Information Criteria in Linear Mixed Models. *Biometrika*, 97, 773–789.

Lee, D.-J. (2010). Smothing mixed model for spatial and spatio-temporal data. PhD thesis, Department of Statistics, Universidad Carlos III de Madrid, Spain.

Lee, D.-J. and Durbán, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, 11, 49–69.

Liang, H., Wu, H. and Zou, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika*, 95, 773–778.

Marra, G. and Radice, R. (2010). Penalised regression splines: theory and application to medical research. *Statistical Methods in Medical Research*, 19, 107–125.

Scheipl, F., Greven, S. and Küchenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis*, 52, 3283–3299.

Wood, S. N. (2006). *Generalized Additive Models. An Introduction with R*. Chapman & Hall/CRC.

SORT 36 (1) January-June 2012, 45-68

A comparison of some confidence intervals for estimating the population coefficient of variation: a simulation study

Monika Gulhar[1], B. M. Golam Kibria[2], Ahmed N. Albatineh[3] and Nasar U. Ahmed[4]

Abstract

This paper considers several confidence intervals for estimating the population coefficient of variation based on parametric, nonparametric and modified methods. A simulation study has been conducted to compare the performance of the existing and newly proposed interval estimators. Many intervals were modified in our study by estimating the variance with the median instead of the mean and these modifications were also successful. Data were generated from normal, chi-square, and gamma distributions for CV = 0.1, 0.3, and 0.5. We reported coverage probability and interval length for each estimator. The results were applied to two public health data: child birth weight and cigarette smoking prevalence. Overall, good intervals included an interval for chi-square distributions by McKay (1932), an interval estimator for normal distributions by Miller (1991), and our proposed interval.

MSC: Primary 62F10; Secondary 62F35

Keywords: Average width, coefficient of variation, inverted coefficient of variation, confidence interval, coverage probability, simulation study, skewed distributions.

1. Introduction

The population coefficient of variation (CV) is a dimensionless (unit-free) measure of the dispersion of a probability distribution. More specifically, it is a measure of variabil-

[1] Department of Biostatistics. Florida International University. Miami. FL 33199. USA. Mgulhar@gmail.com.

[2] Department of Mathematics and Statistics. Florida International University. Miami. FL 33199. USA. kibriag@fiu.edu.

[3] Department of Biostatistics. Florida International University. Miami. FL 33199. USA. aalbatin@fiu.edu.

[4] Department of Epidemiology. Florida International University. Miami. FL 33199. USA. ahmedn@fiu.edu.

Received: April 2011
Accepted: December 2011

ity relative to the mean. This measure can be used to make comparisons across several populations that have different units of measurement. The population CV is defined as a ratio of the population standard deviation (σ) to the population mean ($\mu \neq 0$)

$$\mathrm{CV} = \frac{\sigma}{\mu} \tag{1}$$

In real life instances the population parameters σ and μ are estimated by the sample estimators s and $\bar{x}$, respectively. One obvious disadvantage arises when the mean of a variable is zero. In this case, the CV cannot be calculated. For small values of the mean, the CV will approach infinity and hence becomes sensitive to small changes in the mean. Even if the mean of a variable is not zero, but the variable contains both positive and negative values and the mean is close to zero, then the CV can be misleading.

The CV is widely used in health sciences in descriptive and inferential manners (Kelley, 2007). Some previous applications include measuring the variation in the mean synaptic response of the central nervous system (Faber and Korn, 1991) and measuring the variability in socioeconomic status and prevalence of smoking among tobacco control environments (Bernatm et al., 2009). The CV has also been used to study the impact of socioeconomic status on hospital use in New York City (Billings et al., 1993). These studies generated the CV for ambulatory care sensitive admissions for nine age cohorts and discussed the variance relative to the average number of admissions.

Perhaps, the most important use of the CV is in descriptive studies (Panichchkikosol-kul, 2009). Because the CV is a unit-less measure, the variation of two or more different measurement methods can be compared to each other. The CV is used often in public health. For instance, when assessing the overall health of an individual, the CV may be useful in a comparison of variability in blood pressure measurement (mmHg) and cholesterol measurement (mg/dL). If variance were used, rather than the CV, then these two measures would not be comparable as their units of measurement differ.

Another important application of the CV is its use in the field of quality control. The inverse of the CV or ICV is equal to the signal to noise ratio (SNR), which measures how much signal has been corrupted by noise (McGibney and Smith, 1993):

$$\mathrm{SNR} = \mathrm{ICV} = \frac{\mu}{\sigma} \tag{2}$$

The CV can be used as a measure of effect or as a point estimator (Kelley, 2007). To test the significance of the CV, a hypothesis test can be conducted and a confidence interval can be generated to reject or accept the null hypothesis. Confidence intervals associated with point estimates provide more specific knowledge about the population characteristics than the p-values in the test of hypothesis (Visintainer and Tejani, 1998). A p-value only allows one to determine if results are significant or non-significant, but a confidence interval allows for the examination of an additional factor, precision. The

precision of a confidence interval can be seen through the width and coverage probability of the interval. Given constant coverage, as the width of the $(1-\alpha)$ 100% confidence interval decreases, the accuracy of the estimate increases (Kelley, 2007). The coverage level is the probability that the estimated interval will capture the true CV value (Banik and Kibria, 2010).

There are various methods available for estimating the confidence interval for a population CV. For more information on the confidence interval for the CV, we refer to Koopmans et al. (1964), Miller (1991), Sharma and Krishna (1994), McKay (1932), Vangel (1996), Curto and Pinto (2009) and recently Banik and Kibria (2011), among others. The necessary sample size for estimating a population parameter is important. Therefore, determining the sample size to estimate the population CV is also important. Tables of necessary sample sizes to have a sufficiently narrow confidence intervals under different scenarios are provided by Kelly (2007).

Many researchers considered several confidence intervals for estimating the population CV. Since the studies were conducted under different simulation conditions they are not comparable as a whole. The objective of this paper is to compare several confidence interval estimators based on parametric, nonparametric and modified methods, which are developed by several researchers at several times under the same simulation conditions. Six confidence intervals that already exist in literature are considered. Additionally, we have made median and bootstrap modifications to several existing intervals in an attempt to improve the interval behaviour. Also, we have proposed our own confidence interval estimator. A simulation study is conducted to compare the performance of the interval estimators. Most of the previous simulations and comparisons found in literature were conducted under the normality assumption. Since in real life the data could be skewed we also conducted simulations under skewed distributions (chi-square and gamma). Finally, based on the simulation results, the intervals with high coverage probability and small width were recommended for practitioners.

The organization of this paper is as follows: In Section 2 we present the proposed confidence intervals. In Section 3, the simulation technique and results are provided and discussed. Two real life data are analyzed in Section 4. Finally, some concluding remarks are presented in Section 5. Due to space limitations, only graphs representing simulations for the normal distribution are presented but the full simulation results are presented in the Appendix.

2. Statistical methodology

Let $x_1, x_2, x_3, \ldots, x_n$ be an independently and identically distributed (iid) random sample of size n from a distribution with finite mean, μ, and finite variance, σ^2. Let $\bar{x}$ be the sample mean and s be the sample standard deviation. Then $\widehat{\mathrm{CV}} = s/\bar{x}$ would be the estimated value of the population CV. We want to find the $(1-\alpha)100\%$ confidence intervals for the population CV. In this section we will review some existing interval

estimators and propose some new interval estimators for CV. A total of 15 intervals will be considered.

2.1. The existing confidence interval estimators

Six existing (parametric and nonparametric) confidence intervals for the CV are reviewed in this section.

2.1.1. Parametric confidence intervals

1. Miller's (1991) confidence interval for normal distribution (referred to as Mill). Miller showed that $\frac{s}{\overline{x}}$ approximates an asymptotic normal distribution with mean $\frac{\sigma}{\mu}$ and variance $m^{-1}(\frac{\sigma}{\mu})^2\left(0.5+(\frac{\sigma}{\mu})^2\right)$. Then, the $(1-\alpha)100\%$ approximate confidence interval for the population CV $=\frac{\sigma}{\mu}$ is given by

$$\frac{s}{\overline{x}} \pm Z_{\frac{\alpha}{2}}\sqrt{m^{-1}\left(\frac{s}{\overline{x}}\right)^2\left(0.5+\left(\frac{s}{\overline{x}}\right)^2\right)} \tag{3}$$

which can be expressed as

$$\frac{s}{\overline{x}} - Z_{\frac{\alpha}{2}}\sqrt{m^{-1}\left(\frac{s}{\overline{x}}\right)^2\left(0.5+\left(\frac{s}{\overline{x}}\right)^2\right)} < \text{CV} < \frac{s}{\overline{x}} + Z_{\frac{\alpha}{2}}\sqrt{m^{-1}\left(\frac{s}{\overline{x}}\right)^2\left(0.5+\left(\frac{s}{\overline{x}}\right)^2\right)} \tag{4}$$

where $m=n-1, \overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ is the sample mean, $s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2}$ is the sample standard deviation and Z_α is the upper $(1-\alpha)100^{th}$ percentile of the standard normal distribution.

2. Sharma and Krishna's (1994) confidence interval for inverted CV (S&K): This confidence interval is given by

$$\frac{\overline{x}}{s}+\frac{\phi^{-1}_{(\alpha/2)}}{\sqrt{n}} < \frac{1}{\text{CV}} < \frac{\overline{x}}{s}-\frac{\phi^{-1}_{(\alpha/2)}}{\sqrt{n}} \tag{5}$$

where ϕ cumulative standard normal distribution. Therefore, the $(1-\alpha)100\%$ confidence interval for the population CV is given by

$$\left(\frac{\overline{x}}{s}-\frac{\phi^{-1}_{(\alpha/2)}}{\sqrt{n}}\right)^{-1} < \text{CV} < \left(\frac{\overline{x}}{s}+\frac{\phi^{-1}_{(\alpha/2)}}{\sqrt{n}}\right)^{-1} \tag{6}$$

3. Curto and Pinto's (2009) iid assumption (C&P):

$$\frac{s}{\bar{x}} - Z_{(\alpha/2)}\sqrt{\frac{1}{n}\left(\left(\frac{s}{\bar{x}}\right)^4 + \frac{1}{2}\left(\frac{s}{\bar{x}}\right)^2\right)} < \text{CV} < \frac{s}{\bar{x}} + Z_{(\alpha/2)}\sqrt{\frac{1}{n}\left(\left(\frac{s}{\bar{x}}\right)^4 + \frac{1}{2}\left(\frac{s}{\bar{x}}\right)^2\right)} \tag{7}$$

2.1.2. Confidence intervals based on chi-square distribution

1. McKay's (1932) confidence interval for chi-square distribution (McK):

$$\left(\frac{s}{\bar{x}}\right)\sqrt{\left(\frac{\chi^2_{\nu,1-\alpha/2}}{\nu+1} - 1\right)\left(\frac{s}{\bar{x}}\right)^2 + \frac{\chi^2_{\nu,1-\alpha/2}}{\nu}} < \text{CV} < \left(\frac{s}{\bar{x}}\right)\sqrt{\left(\frac{\chi^2_{\nu,\alpha/2}}{\nu+1} - 1\right)\left(\frac{s}{\bar{x}}\right)^2 + \frac{\chi^2_{\nu,\alpha/2}}{\nu}} \tag{8}$$

   where $\nu = n-1$ and $\chi^2_{\nu,\alpha}$ is the 100α-th percentile of a chi-square distribution with ν degrees of freedom.

2. Modified McKay (1996) confidence interval (MMcK): Vangel (1996) modified McKay's original (1932) interval:

$$\left(\frac{s}{\bar{x}}\right)\sqrt{\left(\frac{\chi^2_{(\nu,1-\alpha/2)}+2}{\nu+1} - 1\right)\left(\frac{s}{\bar{x}}\right)^2 + \frac{\chi^2_{\nu,1-\alpha/2}}{\nu}} < \text{CV} < \left(\frac{s}{\bar{x}}\right)\sqrt{\left(\frac{\chi^2_{(\nu,\alpha/2)}+2}{\nu+1} - 1\right)\left(\frac{s}{\bar{x}}\right)^2 + \frac{\chi^2_{\nu,\alpha/2}}{\nu}} \tag{9}$$

3. Panichkitkosolkul's (2009) confidence interval (Panich): Panichkitkosokul further modified the Modified McKay interval by replacing the sample CV with the maximum likelihood estimator for a normal distribution, $\tilde{k}$

$$\tilde{k}\sqrt{\left(\frac{\chi^2_{(\nu,1-\alpha/2)}+2}{\nu+1} - 1\right)\left(\tilde{k}\right)^2 + \frac{\chi^2_{\nu,1-\alpha/2}}{\nu}} < \text{CV} < \tilde{k}\sqrt{\left(\frac{\chi^2_{(\nu,\alpha/2)}+2}{\nu+1} - 1\right)\left(\tilde{k}\right)^2 + \frac{\chi^2_{\nu,\alpha/2}}{\nu}} \tag{10}$$

   where $\tilde{k} = \frac{\sqrt{\sum_{i=1}^n (x_i-\bar{x})^2}}{\sqrt{n}\bar{x}}$

2.2. Median modifications of existing intervals

Kibria (2006) and Shi and Kibria (2007) claimed that for a skewed distribution, the median describes the centre of the distribution better than the mean. Thus, for skewed data it makes more sense to measure sample variability in terms of the median rather than the mean. Following Shi and Kibria (2007), the $(1-\alpha)100\%$ CI for the CV are obtained for four of the existing estimators and provided below. These median modifications are made in attempt to improve the performance of the original intervals. The intervals selected for modification represent parametric and non-parametric estimators.

1. Median Modified Miller Estimator (Med Mill):

$$\frac{\tilde{s}}{\bar{x}} - Z_{\alpha/2}\sqrt{m^{-1}\left(\frac{\tilde{s}}{\bar{x}}\right)^2\left(0.5+\left(\frac{\tilde{s}}{\bar{x}}\right)^2\right)} < \text{CV} < \frac{\tilde{s}}{\bar{x}} + Z_{\alpha/2}\sqrt{m^{-1}\left(\frac{\tilde{s}}{\bar{x}}\right)^2\left(0.5+\left(\frac{\tilde{s}}{\bar{x}}\right)^2\right)} \tag{11}$$

where $\tilde{s} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\tilde{x})^2}$ and $\tilde{x}$ is the sample median.

2. Median Modification of McKay (Med McK):

$$\left(\frac{\tilde{s}}{\bar{x}}\right)\sqrt{\left(\frac{\chi^2_{\nu,1-\alpha/2}}{\nu+1}-1\right)\left(\frac{\tilde{s}}{\bar{x}}\right)^2+\frac{\chi^2_{\nu,1-\alpha/2}}{\nu}} < \text{CV} < \left(\frac{\tilde{s}}{\bar{x}}\right)\sqrt{\left(\frac{\chi^2_{\nu,\alpha/2}}{\nu+1}-1\right)\left(\frac{\tilde{s}}{\bar{x}}\right)^2+\frac{\chi^2_{\nu,\alpha/2}}{\nu}} \tag{12}$$

3. Median Modification of Modified McKay (Med MMcK):

$$\left(\frac{\tilde{s}}{\bar{x}}\right)\sqrt{\left(\frac{\chi^2_{\nu,1-\alpha/2}+2}{\nu+1}-1\right)\left(\frac{\tilde{s}}{\bar{x}}\right)^2+\frac{\chi^2_{\nu,1-\alpha/2}}{\nu}} < \text{CV} < \left(\frac{\tilde{s}}{\bar{x}}\right)\sqrt{\left(\frac{\chi^2_{\nu,\alpha/2}+2}{\nu+1}-1\right)\left(\frac{\tilde{s}}{\bar{x}}\right)^2+\frac{\chi^2_{\nu,\alpha/2}}{\nu}} \tag{13}$$

4. Median Modified Curto and Pinto(2009) iid assumption (Med C&P):

$$\frac{\tilde{s}}{\bar{x}} - Z_{(\alpha/2)}\sqrt{\frac{\left(\frac{\tilde{s}}{\bar{x}}\right)^4+0.5\left(\frac{\tilde{s}}{\bar{x}}\right)^2}{n}} < \text{CV} < \frac{\tilde{s}}{\bar{x}} + Z_{(\alpha/2)}\sqrt{\frac{\left(\frac{\tilde{s}}{\bar{x}}\right)^4+0.5\left(\frac{\tilde{s}}{\bar{x}}\right)^2}{n}} \tag{14}$$

2.3. Bootstrap confidence intervals

Bootstrapping is a commonly used computer-intensive, nonparametric tool (introduced by Efron, 1979), which is used to make inference about a population parameter when there are no assumptions regarding the underlying population available. It will be especially useful because unlike other methods, this technique does not require any assumptions to be made about the underlying population of interest (Banik and Kibria, 2009). Therefore, bootstrapping can be applied to all situations. This method is implemented by simulating an original data set then randomly selecting variables several times with replacement to estimate the CV. The accuracy of the bootstrap CI depends on the number of bootstrap samples. If the number of bootstrap samples is large enough, the CI may be very accurate. We will consider four different bootstrap methods which are summarized in this section.

Let $X^{(*)} = X_1^{(*)}, X_2^{(*)}, X_3^{(*)}, \ldots, X_n^{(*)}$ where the i^{th} sample is denoted $X^{(i)}$ for $i = 1, 2, \ldots, B$ and B is the number of bootstrap samples. Efron (1979) suggests using a minimum value of $B = 1000$. Following Banik and Kibria (2009), we will consider four different bootstrap methods for estimating the population CV which are summarized in this section.

2.3.1. Non-parametric bootstrap CI (NP BS)

The CV for all bootstrap samples is computed and then ordered as follows:

$$\mathrm{CV}^*_{(1)} \leq \mathrm{CV}^*_{(2)} \leq \mathrm{CV}^*_{(1)} \cdots \mathrm{CV}^*_{(B)}$$

The lower confidence level (LCL) and upper confidence level (UCL) for the population CV is then

$$\mathrm{LCL} = \mathrm{CV}^*_{[(\alpha/2)B]} \quad \text{and} \quad \mathrm{UCL} = \mathrm{CV}^*_{[(1-\alpha/2)B]}$$

Therefore, in a case where $B = 1000$, LCL = the 25^{th} bootstrap sample and UCL = the 975^{th} bootstrap sample.

2.3.2. Bootstrap t-approach (BS)

Following Banik and Kibria (2009), we will propose a version called the bootstrap-t, defined as

$$\mathrm{LCL} = \widehat{\mathrm{CV}} + T^*_{(n-1),\alpha/2} S_{\widehat{\mathrm{CV}}} \quad \text{and} \quad \mathrm{UCL} = \widehat{\mathrm{CV}} + T^*_{(n-1),1-\alpha/2} S_{\widehat{\mathrm{CV}}}$$

where $\widehat{\mathrm{CV}}$ is the sample CV, $T^*_{\alpha/2}, T^*_{1-\alpha/2}$ are the $(\alpha/2)^{th}$ and $(1-\alpha/2)^{th}$ sample quantiles of

$$T_i^* = \frac{\mathrm{CV}_i^* - \overline{\overline{\mathrm{CV}}}}{\widehat{\sigma}_{\mathrm{CV}}}$$

where

$$\widehat{\sigma}_{\mathrm{CV}} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}\left(\mathrm{CV}_i^* - \overline{\overline{\mathrm{CV}}}\right)^2} \quad \text{and} \quad \overline{\overline{\mathrm{CV}}} = \frac{1}{B}\sum_{i=1}^{B}\mathrm{CV}_i^*$$

is a bootstrap CV.

2.3.3. Modified median Miller based on critical value from bootstrap samples (BSMill)

$$\frac{\tilde{s}}{\overline{x}}+T^*_{\alpha/2}\sqrt{m^{-1}\left(\frac{\tilde{s}}{\overline{x}}\right)^2\left(0.5+\left(\frac{\tilde{s}}{\overline{x}}\right)^2\right)}<\mathrm{CV}<\frac{\tilde{s}}{\overline{x}}+T^*_{1-(\alpha/2)}\sqrt{m^{-1}\left(\frac{\tilde{s}}{\overline{x}}\right)^2\left(0.5+\left(\frac{\tilde{s}}{\overline{x}}\right)^2\right)} \tag{15}$$

2.3.4. Modified median Curto and Pinto (2009) based on BS sample (BS C&P)

$$\frac{\tilde{s}}{\overline{x}}-T^*_{(\alpha/2)}\sqrt{\frac{\left(\frac{\tilde{s}}{\overline{x}}\right)^4+0.5\left(\frac{\tilde{s}}{\overline{x}}\right)^2}{n}}<\mathrm{CV}<\frac{\tilde{s}}{\overline{x}}+T^*_{1-(\alpha/2)}\sqrt{\frac{\left(\frac{\tilde{s}}{\overline{x}}\right)^4+0.5\left(\frac{\tilde{s}}{\overline{x}}\right)^2}{n}} \tag{16}$$

where $T^*_{\alpha/2}$ and $T^*_{1-\alpha/2}$ are the $(\alpha/2)^{th}$ and $(1-\alpha/2)^{th}$ sample quantiles of $T^*_i=\frac{(\mathrm{CV}^*_i-\overline{\overline{\mathrm{CV}}})}{\widehat{\sigma}_{\mathrm{CV}}}$

2.4. Proposed confidence interval based on estimator of σ^2

Another parametric confidence interval that was compared in the simulation was one based on the known formula for calculating the confidence interval for σ^2. Let $x_1,x_2,x_3,\ldots,x_n$ be an iid random sample which is normally distributed with finite mean μ and variance σ^2. Then the $(1-\alpha)100\%$ CI for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\nu,1-\alpha/2}}<\sigma^2<\frac{(n-1)s^2}{\chi^2_{\nu,\alpha/2}} \tag{17}$$

Assuming that $\mu\neq 0$, dividing this interval by μ^2 results in

$$\frac{(n-1)s^2}{\left(\chi^2_{\nu,1-\alpha/2}\right)\mu^2}<\left(\frac{\sigma}{\mu}\right)^2<\frac{(n-1)s^2}{\left(\chi^2_{\nu,\alpha/2}\right)\mu^2} \tag{18}$$

Since μ is not known, we can replace it by the unbiased estimator of μ which is $\overline{x}$ resulting in

$$\frac{(n-1)s^2}{\left(\chi^2_{\nu,1-\alpha/2}\right)\overline{x}^2}<\mathrm{CV}^2<\frac{(n-1)s^2}{\left(\chi^2_{\nu,\alpha/2}\right)\overline{x}^2} \tag{19}$$

Taking the square root results in the final proposed interval estimator given by

$$\frac{\sqrt{n-1}\widehat{\mathrm{CV}}}{\sqrt{\chi^2_{\nu,1-\alpha/2}}}<\mathrm{CV}<\frac{\sqrt{n-1}\widehat{\mathrm{CV}}}{\sqrt{\chi^2_{\nu,\alpha/2}}} \tag{20}$$

3. Simulation study

Many researchers have considered several confidence intervals for estimating the population CV. However, these studies were all conducted under different simulation conditions and therefore they are not comparable as a whole. In this study we considered 15 useful confidence intervals (six existing intervals, eight modified intervals, and one proposed interval) for estimating the population CV and compared them under the same simulation conditions. A Monte-Carlo simulation was conducted using the R statistical software (2010) version 2.10.1 to compare the performance of the interval estimators. The performance of the estimators was considered for various CV values, sample sizes, and distributions.

Table 1: *The CV and skewness of data from normal, chi-square, and gamma distributions.*

Distribution	CV	Skewness
N(μ, σ)	$\frac{\sigma}{\mu}$	0
chi-square (ν)	$\sqrt{\frac{2}{\nu}}$	$2\sqrt{\frac{2}{\nu}}$
gamma (α,2)	$\frac{1}{\sqrt{\alpha}}$	$\frac{2}{\sqrt{\alpha}}$

Table 2: *Estimated lower and upper confidence limits and corresponding widths for all proposed methods for Example 1 ($n = 189$).*

Method	Lower Confidence Limit	Upper Confidence Limit	Width
Mill	0.223719	0.277478	0.053759
S&K	0.223659	0.277365	0.053706
McK	0.223073	0.276620	0.053546
MMcK	0.221061	0.274088	0.053027
C&P	0.239134	0.256633	0.017499
Panich	0.230751	0.286434	0.055683
Proposed	0.224879	0.275406	0.050527
Med Mill	0.223940	0.277752	0.053812
MedMcK	0.223878	0.277643	0.053765
MedMMcK	0.221277	0.274362	0.053086
Med C&P	0.221347	0.274292	0.052945
NP BS	0.220954	0.274950	0.053996
P BS	0.219723	0.274653	0.054930
BS Mill	0.220414	0.273982	0.053568
BS C&P	0.220486	0.273912	0.053426

3.1. Simulation technique

To study the behaviour of small and large sample sizes, we used $n = 15, 25, 50, 100$. To examine the behaviour of the intervals when the sample size is increased even further, a second simulation was conducted using $n = 500$. A random sample was generated with specific parameters from a normal distribution, chi-square distribution and gamma distribution. The probability density functions of the distributions are given below.

- Normal distribution, with $\mu = 10$ and $\sigma = 1, 3, 5$ where

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-10}{\sigma}\right)^2}, \quad -\infty < x < \infty, \quad \sigma > 0$$

- Chi-square distribution, with degrees of freedom (df) $\nu = 200, 22, 8$, where

$$f(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\nu/2)} e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}, \quad x \geq 0$$

- Gamma distribution with $\alpha = 100, 11.11, 4$ and $\beta = 2$, where

$$f(x) = \frac{1}{\sqrt{\alpha}} \frac{e^{-\frac{x}{2}} x^{\alpha-1}}{2^{\alpha}}, \quad x > 0, \alpha > 0$$

The CV was calculated for each distribution type by utilizing the equations in Table 1. Thus, based on the CV formula presented in Table 3.1, all three distribution types consider CV $= 0.10, 0.30, 0.50$. The number of simulation replications (M) was 2000 for each case. For intervals that utilize the bootstrapping technique, 1000 bootstrap samples (B) are used for each n.

Coverage probability and width of the interval were measured for each case. The most common 95% confidence interval ($\alpha = 0.05$) is used for measuring the confidence level. The coverage probability is calculated by counting the number of times the true CV is captured between the upper and lower limits. Generally, since the sampling distribution of the sample CV is approximately normally distributed, see Miller (1991), as the sample size (number of sample CV's) increases the distribution becomes more symmetric and it is expected for the coverage probability to approach $(1-\alpha)$. When $\alpha = 0.05$, an interval that has perfect performance in terms of coverage probability will capture the true CV between the upper and lower limits 95% of the time. The coverage probability is an excellent method for evaluating the success of a particular interval in capturing the true parameter. An interval width is calculated by subtracting a lower limit from an upper limit. A smaller width is better because it means that the true CV is captured within a smaller span and the results are more precise. In cases where the coverage probability is comparable, the interval length is especially important because the smaller width will give a more accurate and precise result. The simulated coverage probabilities and interval widths for the normal, chi-square and

Table 3: *Estimated lower and upper confidence limits and corresponding widths for all proposed methods for Example 2 ($n = 24$).*

Method	Lower Confidence Limit	Upper Confidence Limit	Width
Mill	0.181510	0.340507	0.158997
S&K	0.181261	0.338874	0.157613
McK	0.177526	0.331398	0.153872
MMcK	0.163988	0.307669	0.143680
C&P	0.215497	0.260397	0.044900
Panich	0.186213	0.350015	0.163802
Proposed	0.165501	0.306156	0.140655
Med Mill	0.183289	0.330811	0.147522
MedMcK	0.187048	0.350897	0.163849
MedMMcK	0.186665	0.349749	0.163084
Med C&P	0.168763	0.317286	0.148523
NP BS	0.170327	0.315722	0.145396
P BS	0.188882	0.340905	0.152023
BS Mill	0.121866	0.305541	0.183675
BS C&P	0.136045	0.319795	0.183750

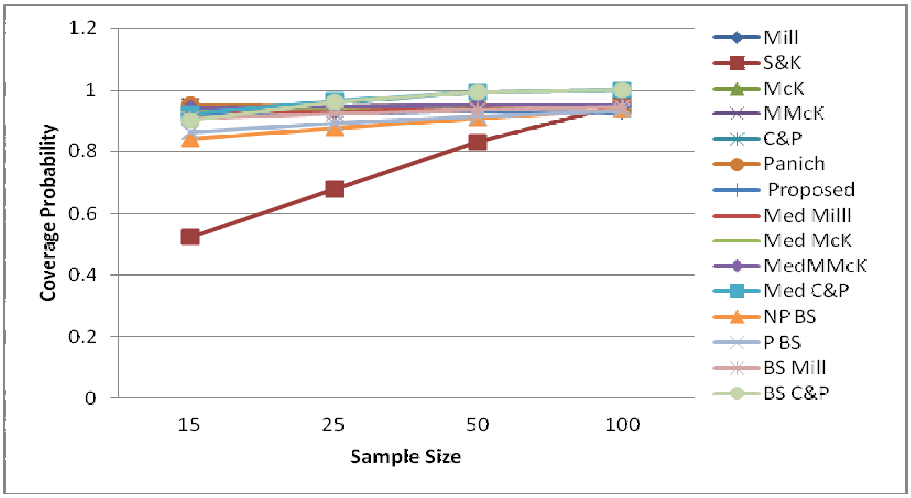


Figure 1: *Sample size vs. coverage probability for normal distribution when* CV *= 0.30.*

gamma distributions are presented in Tables 4, 5 and 6 of the appendix, respectively, and for $n = 500$ of all distributions are presented in Table 7 of the appendix. Each table gives results for the various sample sizes and CV values previously mentioned. Due to

space limitation simulation results of the normal distribution are graphically displayed in Figure 1 through Figure 4, but full simulation results are given in the Appendix.

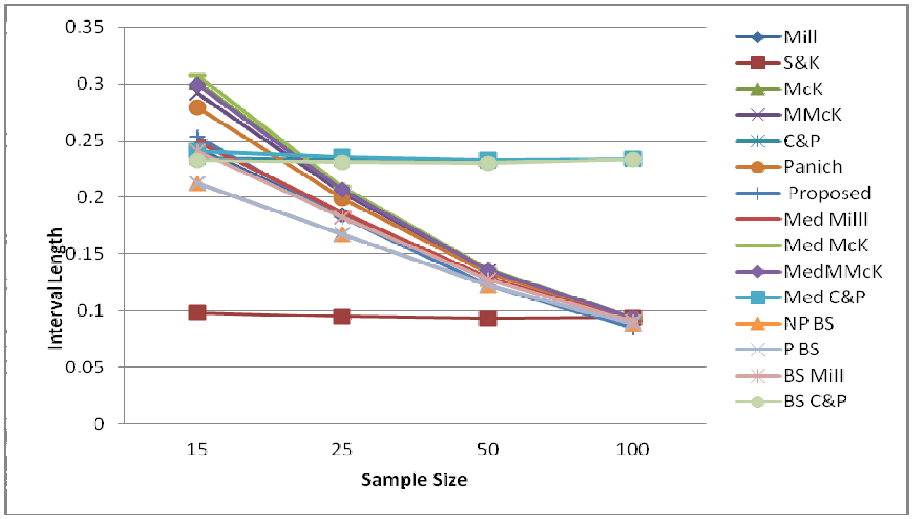


Figure 2: *Sample size vs. interval length for normal distribution when CV = 0.30.*

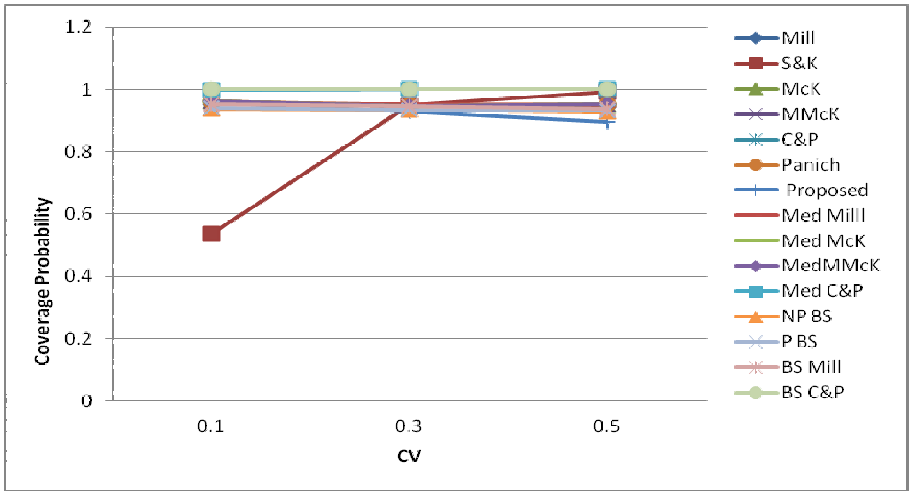


Figure 3: *Coefficient of variation vs. coverage probability for normal distribution when* $n = 100$.

3.2. Simulation results

A main trend that was noted throughout all distributions was the net increase in coverage probability from $n = 15$ to $n = 100$. The results will be discussed below by distribution type. To aid in visualization of results, multiple graphs have been referred to.

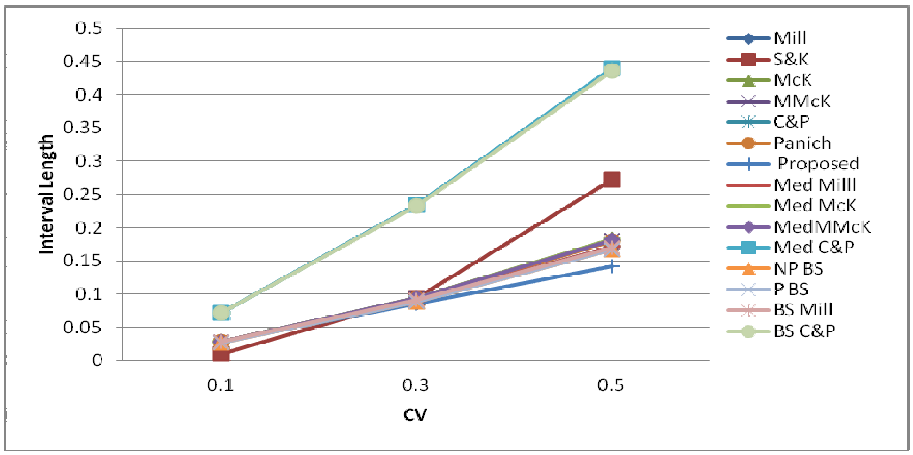


Figure 4: *Coefficient of variation vs. interval length for normal distribution when* $n = 100$.

3.2.1. Normal distribution

Table 4 and Figures 1–4 present simulated results for the normal distribution. Figure 1 displays sample size vs. coverage probability for all tested intervals when CV $= 0.30$. This figure clearly shows that S&K had the weakest overall performance with its lowest coverage probability at 21% (Table 4). However, S&K's performance improves as sample size increases. By $n = 100$, almost all intervals are performing at a similar level (Figure 1). All C&P intervals (C&P, Med C&P, and BS C&P) over exceeded the expected coverage probability of 95% and reached 100% and are clear outliers. Figure 2 displays the sample size vs. interval length for all intervals when CV $= 0.30$. All C&P intervals seem to remain stable and large throughout all sample sizes. The S&K interval also remains stable; however, the values are, as desired, low. All other intervals follow a similar trend: as sample size increases, interval length decreases. When $n = 100$, all intervals are performing similarly, however the C&P intervals are clear outliers. Figure 3 represents the CV vs. coverage probability for normal distribution when $n = 100$. All intervals seem to fall between a coverage range of .9 and 1 however, when CV $= 0.10$ the S&K interval performs very poorly.

The largest sample size presented in Tables 4, 5, and 6 is $n = 100$. For instances where one wants to consider an even larger sample size, Table 7 presents the behaviour of each interval for all distribution types when $n = 500$. Overall, all intervals seem to perform at a similar level when $n = 500$. One key difference that should be noted is that when sample size in increased to 500, there is a drastic improvement in the performance of S&K.

3.2.2. Skewed distributions

The simulated results for chi-square distribution are presented in Table 5 which indicates that many of the same trends as seen in the normal distribution. Similar to the normal distribution, the S&K interval had the weakest overall performance for the chi-square distribution. Table 5 shows that as sample size increases, coverage probability for all intervals has a general increasing trend. The S&K interval is notably lacking in performance relatively to other intervals. Table 5 shows that as sample size increases, the interval length decreases for most of the intervals. Similar to the results seen for the normal distribution, all C&P intervals seem to remain stable and large throughout all sample sizes and do not follow the decreasing trend. The S&K interval also did not follow the trend as it remained low throughout all sample sizes. Table 5 indicates that with the exception of S&K, all intervals had coverage probabilities between 0.9 and 1 for all CV values. S&K performed much lower than the other intervals when $CV = 0.10$. The same trend was seen for the normal distribution. Table 5 shows that as the CV value increases, all interval lengths become increasingly wider. The widths are narrowest when $CV = 0.10$ and widest when $CV = 0.50$. The C&P intervals have the greatest length for all CV values. When $CV = 0.50$, S&K also has a notably higher interval length than the other intervals and the narrowest width is observed in the BS and PBS intervals.

The results for gamma distribution are presented in Table 6. Although the gamma distribution has a greater skewness than the chi-square distribution, this difference did not have a major effect on the trends seen between the two distribution types. All trends in the gamma distribution are comparable to results from the chi-square distribution because the two distribution types are related. When sample size is increased to 500 for skewed distributions, all intervals perform at a similar level (Table 7). The weakest interval when $n = 500$ is C&P. C&P has its lowest coverage probability of 21% when $CV = 0.10$ in the gamma distribution.

4. Application

To illustrate the findings of the paper, some real life health data are analyzed in this section.

4.1. Example 1: Birth weight data

The first data set was obtained from Hosmer and Lemeshow (2000), which was collected from the Baystate Medical Center, in Springfield, Massachusetts (University of Massachusetts Amherst, 2000). Child birth weight data were collected from 189 women. A baby weighing less than 2500 grams was defined as a “low birth weight” child. Among them, 59 women had low birth weight babies and 130 women had normal birth weight babies. The average birth weight was 2944.66 grams, with a standard deviation

of 729.022 grams. The coefficient of variation for the low birth data is 0.248. The histogram of the data is presented in Figure 5, which showed a mound shaped distribution. The Shapiro-Wilk test (W = 0.9925, p-value = 0.4383) supported that the data follow a normal distribution.

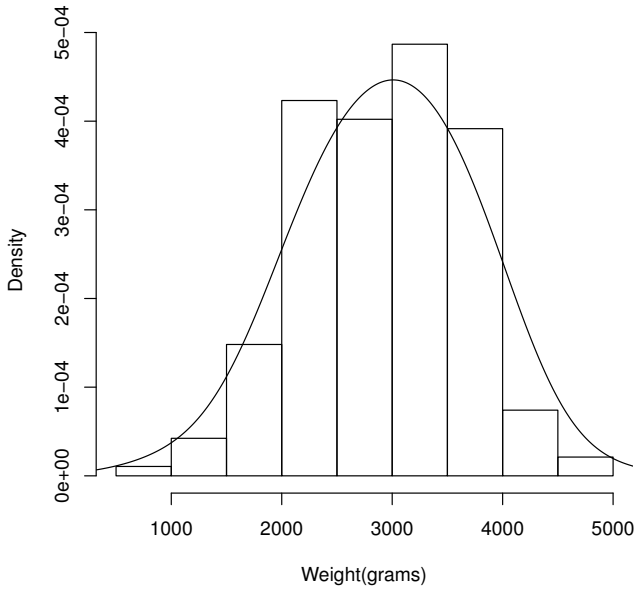


Figure 5: *Histogram of birth weight of babies from 189 women, with modelled normal distribution.*

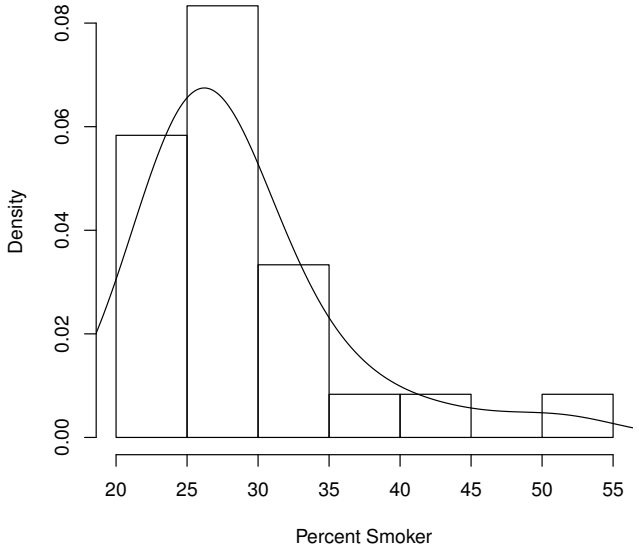


Figure 6: *Histogram of the percent of smokers in the United States for 24 selected years, with modelled gamma distribution.*

Using SPSS (2008), the Kolmogorov-Smirnov (ks) goodness of fit test result (ks = 0.043, p-value = 0.200[1]) for a hypothesized normal distribution with mean $\mu = 2945$ and standard deviation $\sigma = 755$, indicates that the data follow a normal distribution with mean of 2945 gm and standard deviation of 755 gm. The corresponding population coefficient of variation, CV $= 755/2945 = 0.2564$. The resulting 95% confidence intervals and corresponding widths for all proposed intervals are reported in Table 2. Results showed that all confidence intervals covered the hypothesize population CV 0.2564. We note that the C&P interval has the shortest width followed by our proposed interval. Panich has the largest width. Based on the overall simulation results presented and Figure 3, we may suggest S&K or the proposed one to be the interval of choice for this example. In fact the proposed interval performed better than the majority of the other interval estimators.

4.2. Example 2: Cigarette smoking among men, women, and high school students: United States, 1965-2007

The data set for this example were obtained from the Center for Disease Control and Prevention (CDC, 2009). The CDC used the National Health Interview Survey and the Youth Risk Behavior Survey to compile the data. Adults were classified as cigarette smokers if they smoked 100 cigarettes in their lifetime and if now smoke daily or occasionally.

A high school student was categorized as a smoker if he or she had even one cigarette in the past month. The data is a compilation of cigarette smoking among adult and high school population for selected years between 1965 and 2007. The total sample size for the data is $n = 24$ selected years between 1965 and 2007. The mean percent of smokers is 28.9%, with a standard deviation of 6.8%. The coefficient of variation for the data is 0.236. The histogram of the data is presented in Figure 6, which shows a right skewed distribution. Using SAS (2008), the Kolmogorov-Smirnov (ks) goodness of fit test result (ks = 0.188, p-value > 0.25) for a hypothesized gamma distribution with shape parameter, $\alpha = 17.6$ and scale parameter $\beta = 1.6$, indicates that the data follow a gamma distribution with $\alpha = 17.6$ and $\beta = 1.6$. For more information on goodness of fit for the gamma distribution, see D'Agostino and Stephens (1986). For this example, the corresponding population coefficient of variation, CV $= \frac{1}{\sqrt{17.6}} = 0.238$. The resulting 95% confidence intervals and corresponding widths for all proposed intervals are reported in Table 3. Results showed that all confidence intervals covered the hypothesize population CV of 0.238. Based on the simulation results presented, we would have expected the S&K interval to have the narrowest width,

1. Lillifors significance correction is applied. This is a lower bound of the true significance.

however, in this example the narrowest width was observed in the original C&P interval. S&K's interval had results that were very comparable to all other intervals, but it was not the best. This example indicates that the C&P bootstrap interval is the narrowest, a finding that is confirmed by our simulation results.

5. Concluding remarks

We have considered several confidence intervals for estimating the population CV. A simulation study has been conducted to compare the performance of the estimators. For simulation purposes, we have generated data from both symmetric (normal) and skewed distributions (chi-square and gamma) to see the performance of the interval estimators. After thorough examination of all individual intervals, and each overall simulation condition, it can be concluded that the intervals that performed well are: Mill, McK, MMcK, Panich, MedMill, MedMcK, MedMMcK, and our proposed interval. The S&K interval was not a good estimator for small sample sizes and had the weakest performance in terms of coverage probability. However, its performance is comparable to other intervals when the sample size is large. All C&P intervals did not perform well for large sample sizes as the interval length for these estimators was too high relative to other estimators. This wide interval length is not desired as it indicates that the results are imprecise. It is known that for symmetric distributions the mean equals the median, but as pointed out by a reviewer, the median modified intervals of the CV for skewed distributions works well for moderate sample sizes, but for a very large sample size, the point estimate of the CV using the median will be very close but doesn't converge to the true population CV. To see the large sample behaviour, we have generated data from normal, chi-square and gamma distribution for $n = 500$ and presented the coverage probability and average widths in Table 7. Clearly, Table 7 shows that the coverage probabilities for all estimators are close to the nominal size of 0.95. For large sample size any of the intervals except S&K could be used. We do not encourage the researchers to use bootstrap confidence intervals as their performances are not significantly better than others and they are very time consuming. However, if some researchers are very conservative about the assumptions of the distributions and willing to accept the extensive computation they might consider the bootstrap methods. Most importantly, in many instances, our newly proposed interval produced the best interval length, especially as CV increased from 0.1 to 0.5. Higher values of CV indicate more variability in the data, a characteristic that is often seen in health sciences where sample sizes are frequently small. Because our proposed interval performed well for the higher values of CV, it will be a good interval to use for many health science data. We also recommend using this interval over other intervals because it is a very user-friendly interval in calculation.

Acknowledgment

The authors would like to thank the editor and two anonymous reviewers for their valuable comments and suggestions that greatly improved the content and presentation of an earlier draft of the paper.

Appendix

The appendix contains Table 4 through Table 7.

Table 4: *Estimated coverage probabilities and widths of the intervals using normal distribution ($M = 2000$).*

	Mill[2]	S&K	McK	MMcK	C&P	Panich	Prop	Med Mill	Med McK	Med MMcK	MedC&P	NPBS	PBS	BSMill	BS C&P
N=15, CV=.1															
Coverage	0.9235	0.212	0.9535	0.9535	0.9155	0.951	0.952	0.9325	0.9445	0.9445	0.9265	0.8395	0.87	0.909	0.9025
Length	0.0746	0.0104	0.0857	0.0854	0.072066	0.0824	0.0842	0.0759	0.0871	0.0869	0.073321	0.0645	0.0645	0.0736	0.071132
N=15, CV=.3															
Coverage	0.9165	0.5235	0.9465	0.9505	0.912	0.9525	0.928	0.9295	0.939	0.9435	0.9245	0.84	0.863	0.9075	0.9015
Length	0.2437	0.0978	0.3015	0.2917	0.235397	0.2791	0.2534	0.2491	0.3072	0.2989	0.240624	0.2122	0.2125	0.2411	0.232897
N=15, CV=.5															
Coverage	0.918	0.7265	0.952	0.9545	0.9115	0.9535	0.8915	0.9295	0.949	0.954	0.922	0.865	0.886	0.9115	0.906
Length	0.4678	0.2948	0.8265	0.689	0.451932	0.6405	0.4244	0.4788	0.8401	0.7156	0.462575	0.424	0.4239	0.4631	0.447386
N=25, CV=.1															
Coverage	0.9305	0.2785	0.9495	0.95	0.964	0.952	0.949	0.9345	0.945	0.945	0.9675	0.877	0.895	0.925	0.9565
Length	0.0565	0.0101	0.0611	0.061	0.071406	0.0598	0.0603	0.0571	0.0618	0.0617	0.072236	0.0513	0.0512	0.0559	0.070708
N=25, CV=.3															
Coverage	0.929	0.6775	0.945	0.944	0.9605	0.945	0.9235	0.934	0.9435	0.9445	0.9645	0.8745	0.89	0.922	0.962
Length	0.1843	0.0948	0.207	0.2039	0.2331	0.1989	0.1822	0.1867	0.2094	0.2068	0.236185	0.1676	0.1675	0.1827	0.231161
N=25, CV=.5															
Coverage	0.918	0.824	0.946	0.9515	0.953	0.95	0.8795	0.93	0.936	0.943	0.9585	0.8825	0.892	0.9145	0.956
Length	0.3503	0.2812	0.4454	0.422	0.443089	0.407	0.3043	0.3557	0.4505	0.4304	0.449977	0.3234	0.3232	0.3476	0.439721
N=50, CV=.1															
Coverage	0.9365	0.374	0.9475	0.9475	0.995	0.9465	0.9455	0.9395	0.9475	0.9475	0.995	0.908	0.911	0.9365	0.9945
Length	0.0398	0.0102	0.0414	0.0413	0.071923	0.0409	0.0409	0.04	0.0416	0.0416	0.072323	0.0377	0.0377	0.0395	0.071469
N=50, CV=.3															
Coverage	0.94	0.829	0.9525	0.952	0.994	0.9495	0.932	0.9445	0.952	0.9515	0.9955	0.909	0.9145	0.9365	0.993
Length	0.1281	0.0929	0.1355	0.1347	0.231613	0.1331	0.1222	0.1289	0.1362	0.1355	0.233008	0.122	0.122	0.1273	0.230112
N=50, CV=.5															
Coverage	0.934	0.9535	0.9455	0.945	0.9925	0.9495	0.8855	0.9385	0.943	0.945	0.9935	0.908	0.921	0.938	0.992
Length	0.2459	0.2794	0.2755	0.2703	0.444506	0.2662	0.2065	0.2478	0.2771	0.2726	0.447815	0.2365	0.2368	0.245	0.442756
N=100, CV=.1															
Coverage	0.954	0.5355	0.961	0.961	0.9995	0.9615	0.96	0.956	0.963	0.963	0.9995	0.938	0.9405	0.953	1
Length	0.0281	0.0102	0.0286	0.0286	0.07217	0.0285	0.0283	0.0282	0.0287	0.0287	0.072372	0.0271	0.0271	0.028	0.071852
N=100, CV=.3															
Coverage	0.9495	0.951	0.9475	0.9475	1	0.9485	0.928	0.952	0.947	0.948	1	0.935	0.9345	0.946	1
Length	0.091	0.094	0.0937	0.0934	0.233686	0.0928	0.0851	0.0912	0.0939	0.0937	0.23442	0.0881	0.0883	0.0907	0.23298
N=100, CV=.5															
Coverage	0.943	0.992	0.951	0.9505	1	0.9525	0.8955	0.9485	0.952	0.9515	1	0.9275	0.9315	0.9385	1
Length	0.1709	0.2724	0.182	0.1805	0.438955	0.1792	0.1416	0.1715	0.1826	0.1813	0.440591	0.1667	0.1664	0.1699	0.43646

Table 5: *Estimated coverage probabilities and widths of the intervals using chi-square distribution* ($M = 2000$).

	Mill[3]	S&K	McK	MMcK	C&P	Panich	Prop	Med Mill	Med McK	Med MMcK	MedC&P	NPBS	PBS	BSMill	BS C&P
N=15, CV=.1															
Coverage	0.9195	0.208	0.958	0.9585	0.913	0.9585	0.9565	0.932	0.9505	0.9505	0.9245	0.827	0.8625	0.909	0.902
Length	0.0735	0.0101	0.0844	0.0842	0.071006	0.0812	0.083	0.0749	0.086	0.0858	0.072382	0.0626	0.0626	0.0725	0.070002
N=15, CV=.3															
Coverage	0.9175	0.564	0.9605	0.9615	0.912	0.966	0.946	0.9275	0.952	0.953	0.9225	0.8335	0.867	0.9055	0.901
Length	0.2406	0.0954	0.2969	0.2875	0.232453	0.2752	0.2509	0.2472	0.3038	0.2964	0.238825	0.1917	0.1917	0.2376	0.229511
N=15, CV=.5															
Coverage	0.9275	0.777	0.9805	0.98	0.921	0.9745	0.9405	0.9415	0.971	0.978	0.938	0.8415	0.8675	0.9235	0.9165
Length	0.4473	0.2713	0.7173	0.6183	0.43216	0.5775	0.4127	0.465	0.7387	0.6593	0.449276	0.32	0.3195	0.4401	0.425139
N=25, CV=.1															
Coverage	0.925	0.2825	0.9395	0.9395	0.957	0.944	0.9365	0.927	0.9375	0.9385	0.9605	0.868	0.887	0.9185	0.954
Length	0.0565	0.0101	0.0611	0.061	0.071407	0.0598	0.0603	0.0571	0.0618	0.0617	0.072231	0.0509	0.0508	0.0559	0.070711
N=25, CV=.3															
Coverage	0.9305	0.694	0.9645	0.9635	0.966	0.961	0.946	0.9365	0.957	0.958	0.971	0.865	0.8885	0.929	0.9635
Length	0.1833	0.0938	0.2057	0.2027	0.231837	0.1978	0.1814	0.1866	0.2089	0.2066	0.235998	0.153	0.1532	0.1818	0.229914
N=25, CV=.5															
Coverage	0.949	0.884	0.9705	0.9715	0.9725	0.97	0.93	0.9515	0.957	0.9615	0.9745	0.8755	0.8935	0.9485	0.9695
Length	0.3446	0.2724	0.4318	0.4112	0.435865	0.3978	0.3014	0.3556	0.442	0.4286	0.449855	0.26	0.2603	0.3409	0.431175
N=50, CV=.1															
Coverage	0.953	0.3895	0.961	0.961	0.997	0.9615	0.9585	0.956	0.9575	0.9575	0.9975	0.915	0.9295	0.9495	0.9965
Length	0.0398	0.0101	0.0413	0.0413	0.071908	0.0409	0.0409	0.04	0.0416	0.0416	0.072358	0.0373	0.0374	0.0395	0.071459
N=50, CV=.3															
Coverage	0.9475	0.8605	0.9585	0.9595	0.998	0.9615	0.9445	0.95	0.953	0.9535	0.999	0.9135	0.917	0.9455	0.9975
Length	0.1293	0.0943	0.1368	0.136	0.233722	0.1343	0.1232	0.1309	0.1382	0.1376	0.236501	0.113	0.1131	0.1284	0.232103
N=50, CV=.5															
Coverage	0.9645	0.9735	0.979	0.9785	0.998	0.978	0.929	0.9685	0.9645	0.97	0.998	0.9035	0.914	0.9645	0.9985
Length	0.2409	0.2696	0.2687	0.2638	0.435342	0.2599	0.2035	0.2466	0.2734	0.2709	0.445631	0.1916	0.1916	0.2389	0.431806
N=100, CV=.1															
Coverage	0.9475	0.5185	0.959	0.959	1	0.9555	0.9565	0.949	0.9575	0.9575	1	0.9295	0.932	0.948	1
Length	0.028	0.0101	0.0286	0.0286	0.072027	0.0284	0.0283	0.0281	0.0287	0.0287	0.072271	0.027	0.0269	0.0279	0.071638
N=100, CV=.3															
Coverage	0.958	0.9605	0.963	0.963	1	0.963	0.95	0.9565	0.9595	0.9595	1	0.922	0.932	0.9575	1
Length	0.091	0.094	0.0937	0.0934	0.233826	0.0929	0.0852	0.0918	0.0944	0.0943	0.235882	0.0821	0.0821	0.0906	0.232691
N=100, CV=.5															
Coverage	0.954	0.9965	0.969	0.968	1	0.964	0.917	0.9545	0.96	0.9635	1	0.9055	0.9095	0.9505	1
Length	0.1697	0.2692	0.1806	0.1791	0.435943	0.1778	0.141	0.1732	0.1833	0.1831	0.444846	0.142	0.142	0.1683	0.432382

Table 6: *Estimated coverage probabilities and widths of the intervals using gamma distribution* ($M = 2000$).

	Mill[4]	S&K	McK	MMcK	C&P	Panich	Prop	Med Mill	Med McK	Med MMcK	MedC&P	NPBS	PBS	BSMill	BS C&P
N=15, CV=.1															
Coverage	0.914	0.218	0.948	0.948	0.9075	0.9515	0.947	0.9265	0.9445	0.9455	0.9175	0.8335	0.8595	0.9	0.8955
Length	0.0735	0.0101	0.0844	0.0842	0.071036	0.0813	0.083	0.0749	0.086	0.0858	0.072374	0.0638	0.0638	0.0725	0.070013
N=15, CV=.3															
Coverage	0.913	0.5505	0.967	0.966	0.905	0.963	0.953	0.9255	0.9605	0.9615	0.9205	0.814	0.8485	0.8995	0.894
Length	0.2368	0.0927	0.2913	0.2825	0.228816	0.2704	0.2476	0.2429	0.2977	0.2906	0.234679	0.1887	0.1888	0.2336	0.225688
N=15, CV=.5															
Coverage	0.926	0.7725	0.985	0.983	0.918	0.981	0.951	0.9405	0.978	0.98	0.9325	0.8355	0.866	0.917	0.9135
Length	0.4398	0.2639	0.6845	0.6029	0.424926	0.5634	0.4077	0.4569	0.7043	0.6366	0.441444	0.3156	0.3161	0.4348	0.420057
N=25, CV=.1															
Coverage	0.928	0.269	0.949	0.949	0.965	0.946	0.948	0.9335	0.944	0.9445	0.9695	0.8715	0.8895	0.9215	0.9615
Length	0.0568	0.0102	0.0615	0.0614	0.071874	0.0602	0.0607	0.0575	0.0622	0.0622	0.072733	0.051	0.051	0.0563	0.071211
N=25, CV=.3															
Coverage	0.937	0.6925	0.968	0.968	0.9695	0.963	0.947	0.9415	0.96	0.96	0.9735	0.8755	0.8975	0.9305	0.965
Length	0.1818	0.0924	0.2039	0.201	0.229989	0.1961	0.1802	0.1851	0.2071	0.2049	0.234137	0.1525	0.1528	0.1803	0.228007
N=25, CV=.5															
Coverage	0.944	0.8695	0.9735	0.973	0.9715	0.97	0.9295	0.951	0.9605	0.9655	0.9765	0.863	0.8825	0.9385	0.969
Length	0.3443	0.2722	0.4314	0.4109	0.435489	0.3975	0.3012	0.3551	0.4414	0.4275	0.449127	0.2578	0.2579	0.3406	0.430841
N=50, CV=.1															
Coverage	0.9425	0.397	0.9565	0.9565	0.995	0.955	0.954	0.9465	0.9535	0.953	0.9955	0.919	0.9275	0.9425	0.994
Length	0.0398	0.0102	0.0414	0.0413	0.071965	0.0409	0.0409	0.0401	0.0416	0.0416	0.072405	0.0373	0.0373	0.0396	0.071531
N=50, CV=.3															
Coverage	0.95	0.8565	0.9605	0.961	0.9975	0.964	0.946	0.949	0.954	0.9555	0.9975	0.906	0.9175	0.9445	0.9975
Length	0.1282	0.0929	0.1356	0.1347	0.231671	0.1331	0.1223	0.1296	0.1368	0.1363	0.234208	0.1124	0.1124	0.1272	0.229973
N=50, CV=.5															
Coverage	0.9575	0.9685	0.9765	0.975	0.997	0.9755	0.933	0.9605	0.9625	0.966	0.997	0.898	0.9075	0.9555	0.997
Length	0.2413	0.2704	0.2692	0.2643	0.43611	0.2604	0.2038	0.247	0.2739	0.2715	0.446485	0.1937	0.1934	0.2386	0.431276
N=100, CV=.1															
Coverage	0.9515	0.533	0.953	0.953	1	0.9555	0.95	0.954	0.951	0.9515	1	0.9295	0.9345	0.9525	1
Length	0.0281	0.0102	0.0287	0.0287	0.072294	0.0285	0.0284	0.0282	0.0288	0.0288	0.072542	0.0271	0.0271	0.028	0.071857
N=100, CV=.3															
Coverage	0.954	0.9535	0.958	0.959	1	0.9575	0.939	0.9525	0.9515	0.9525	1	0.918	0.9265	0.949	1
Length	0.0905	0.093	0.0931	0.0929	0.23239	0.0923	0.0847	0.0912	0.0938	0.0937	0.234425	0.0826	0.0828	0.09	0.231312
N=100, CV=.5															
Coverage	0.969	0.996	0.973	0.973	1	0.9745	0.9285	0.9675	0.959	0.963	1	0.92	0.923	0.9645	1
Length	0.17	0.2698	0.1809	0.1794	0.436682	0.1782	0.1412	0.1736	0.1838	0.1836	0.446032	0.1417	0.1418	0.1689	0.433915

Table 7: *Estimated coverage probabilities and widths of the intervals using normal, chi-square and gamma distributions* ($n = 500, M = 2000$).

	Mill[5]	S&K	McK	MMcK	C&P	Panich	Prop	Med Mill	Med McK	Med MMcK	MedC&P	NPBS	PBS	BSMill	BS C&P
Normal Distribution															
CV=.1															
Coverage	0.9475	0.9475	0.946	0.9455	0.2167	0.9455	0.9465	0.949	0.949	0.9445	0.9445	0.9415	0.942	0.9435	0.9435
Length	0.0126	0.0126	0.0126	0.0125	0.0018	0.01252	0.0125	0.0126	0.0126	0.0125	0.01253	0.0124	0.0124	0.0125	0.01248
N=15, CV=.3															
Coverage	0.947	0.9465	0.9475	0.9455	0.5645	0.9455	0.925	0.9435	0.9435	0.945	0.9445	0.943	0.9405	0.942	0.9415
Length	0.0408	0.0408	0.0407	0.0405	0.0158	0.04041	0.0374	0.0408	0.0408	0.0405	0.04044	0.04	0.04	0.0403	0.04027
N=15, CV=.5															
Coverage	0.95	0.9505	0.952	0.953	0.671	0.9525	0.909	0.949	0.95	0.9535	0.953	0.944	0.9505	0.9485	0.9485
Length	0.058	0.0579	0.0579	0.0571	0.0281	0.05701	0.0498	0.058	0.058	0.0571	0.05705	0.0565	0.0565	0.0568	0.05678
Chi-Square Distribution															
N=25, CV=.1															
Coverage	0.9485	0.9485	0.9485	0.95	0.2245	0.95	0.9455	0.9475	0.9475	0.947	0.947	0.941	0.941	0.9435	0.9435
Length	0.0126	0.0126	0.0126	0.0125	0.0018	0.01252	0.0124	0.0126	0.0126	0.0125	0.01253	0.0123	0.0123	0.0125	0.01246
N=25, CV=.3															
Coverage	0.961	0.962	0.963	0.958	0.566	0.958	0.937	0.947	0.9475	0.951	0.9505	0.943	0.942	0.956	0.956
Length	0.041	0.041	0.0409	0.0407	0.0159	0.04061	0.0375	0.0412	0.0412	0.0409	0.04087	0.0383	0.0383	0.0405	0.04048
N=25, CV=.5															
Coverage	0.9655	0.9655	0.965	0.962	0.7075	0.962	0.935	0.949	0.95	0.953	0.953	0.9415	0.9435	0.962	0.962
Length	0.0578	0.0578	0.0577	0.0569	0.028	0.05688	0.0497	0.0584	0.0585	0.0576	0.05753	0.0523	0.0523	0.0567	0.05663
Gamma Distribution															
N=50, CV=.1															
Coverage	0.9455	0.9455	0.946	0.945	0.2095	0.9445	0.9415	0.944	0.944	0.9445	0.9445	0.9355	0.937	0.942	0.942
Length	0.0126	0.0126	0.0126	0.0125	0.0018	0.01253	0.0125	0.0126	0.0126	0.0126	0.01254	0.0123	0.0123	0.0125	0.01247
N=50, CV=.3															
Coverage	0.9515	0.951	0.95	0.95	0.579	0.9495	0.9325	0.946	0.946	0.9455	0.9455	0.938	0.9345	0.9455	0.9455
Length	0.0408	0.0408	0.0407	0.0404	0.0158	0.0404	0.0374	0.041	0.041	0.0407	0.04066	0.0381	0.0381	0.0403	0.04028
N=50, CV=.5															
Coverage	0.9735	0.974	0.975	0.968	0.7785	0.968	0.9255	0.948	0.9505	0.957	0.957	0.938	0.9415	0.9655	0.9645
Length	0.0781	0.078	0.0779	0.0759	0.0439	0.07585	0.0622	0.0792	0.0795	0.0773	0.07726	0.0673	0.0673	0.0756	0.0755

References

Banik, S. and Kibria, B. M. G. (2010). Comparison of some parametric and nonparametric type one sample confidence intervals for estimating the mean of a positively skewed distribution. *Communications in Statistics-Simulation and Computation*, 39, 361–389.

Banik, S. and Kibria, B. M. G. (2011). Estimating the population coefficient of variation by confidence intervals . *Communications in Statistics-Simulation and Computation*, 40, 1236–1261.

Bernatm, D. H., Lazovich, D., Forster, J. L., Oakes, J. M. and Chen, V. (2009). Area-level variation in adolescent smoking. *Preventing Chronic Disease*, 6(2), 1–8.

Billings, J., Zeitel, L., Lukomink, J., Carey, T.S., Blank, A. E. and Newman, L. (1993). Impact of socioeconomic status of hospital use in New York City. *Health Affairs*, 162–173.

Center for Disease Control and Prevention (2009). Health, United States, 2009. Retrieved from: http://www.cdc.gov/nchs/data/hus/hus09.pdf.

Curto, J. D. and Pinto, J. C. (2009). The coefficient of variation asymptotic distribution in the case of non-iid random variables. *Journal of Applied Statistics*, 36(1), 21–32.

D'Agostino, Ralph B. and Stephens, Michael A. (1986). Goodness-of-fit-techniques (Statistics: a Series of Textbooks and Monographs, vol. 68).

Efron, B. (1979). Bootstrap methods: another look at the jack knife. *Annals of Statistics*, 7, 1–26.

Efron, B. (1987). Better bootstrapping confidence intervals. *Journal of American Statistical Association*, 82(397), 171–185.

Faber, D. S. and Korn, H. (1991). Applicability of the coefficient of variation for analyzing synaptic plasticity. *Biophysical Society*, 60, 1288–1294.

Hosmer D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*: Second edition. New York: John Wiley & Sons.

Kelly, K. (2007). Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research Methods*, 39(4), 755–766.

Kibria, B. M. G. (2006). Modified confidence intervals for the mean of the Asymmetric distribution. *Pakistan Journal of Statistics*, 22(2), 109–120.

Koopmans, L. H., Owen, D. B. and Rosenblatt, J. I. (1964). Confidence intervals for the coefficient of variation for the normal and log normal distributions. *Biometrika Trust*, 51(1/2), 25–32.

McGibney, G. and Smith, M. R. (1993). An unbiased signal?to?noise ratio measure for magnetic resonance images. *Medical Physics*, 20(4), 1077–1079.

McKay, A. T. (1932). Distribution of the coefficient of variation and the extended t distribution. *Journal of Royal Statistical Society*, 95, 695–698.

Miller, E. G. (1991). Asymptotic test statistics for coefficient of variation. *Communication in Statistic-Theory & Methods*, 20, 3351–3363.

Panichkitkosolkul, W. (2009). Improved confidence intervals for a coefficient of variation of a normal distribution. *Thailand Statistician*, 7(2), 193–199.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org.

SAS Institute Inc. (2008). SAS/STAT User's Guide, Version 9.2, Cary, NC: SAS Institute Inc.

Sharma, K. K. and Krishna, H. (1994). Asymptotic sampling distribution of inverse coefficient-of-variation and its applications. *IEEE Transactions on Reliability*, 43(4), 630–633.

Shi, W. and Kibria, B. M. G. (2007). On some confidence intervals for estimating the mean of a skewed population. *International Journal of Mathematical Education in Science and Technology*, 38(3), 412–421.

SPSS for Windows (2008). SPSS Inc., Chicago, IL.

Vangel, M. G. (1996). Confidence intervals for a normal coefficient of variation. *The American Statistician*, 15(1), 21–26.

Visintainer P. F. and Tejani, N. (1998). Understanding and using confidence intervals in clinical research. *The Journal of Maternal-Fetal Medicine*, 7, 201–2.

SORT 36 (1) January-June 2012, 69-80

Comparing and calibrating discrepancy measures for Bayesian model selection

Julián de la Horra[1] and María Teresa Rodríguez-Bernal[2]

Abstract

Different approaches have been considered in the literature for the problem of Bayesian model selection. Recently, a new method was introduced and analysed in De la Horra (2008) by minimizing the posterior expected discrepancy between the set of data and the Bayesian model, where the chi-square discrepancy was used. In this article, several discrepancy measures are considered and compared by simulation, and it is obtained that the chi-square discrepancy is reasonable to use. Then, an easy method for calibrating discrepancies is proposed, and the behaviour of this approach is studied on simulated data. Finally, a set of real data is analysed.

MSC: 62F15

Keywords: Bayesian model selection, discrepancy measure, calibration, posterior expected discrepancy.

1. Introduction

In recent years, many articles have been written about Bayesian model selection. Many of these articles have relied on Bayes factors or posterior odds; for instance, Spiegelhalter and Smith (1982), Aitkin (1991), O'Hagan (1995), Berger and Pericchi (1996). Other papers have considered a predictive approach; for instance, Geisser and Eddy (1979), San Martini and Spezzaferri (1984), Gelfand et al. (1992), Gelfand (1995), Laud and Ibrahim (1995), Gelfand and Ghosh (1998), Gutiérrez-Peña and Walker (2001), Trottini and Spezzaferri (2002), De la Horra and Rodríguez-Bernal (2005, 2006).

[1] Departamento de Matemáticas. Universidad Autónoma de Madrid. 28049 Madrid (Spain). julian.delahorra@uam.es

[2] Departamento de Estadística e Investigación Operativa. Universidad Complutense de Madrid. 28040 Madrid (Spain). mayter@mat.ucm.es

Received: October 2010
Accepted: February 2012

A different and easy method was proposed and analysed in De la Horra (2008). This method was based on the well-known property that, under the true model, the cumulative distribution function is distributed as a uniform distribution over the interval (0, 1). A suitable discrepancy measure between the sample and the Bayesian model is needed. In De la Horra (2008), the chi-square (χ^2) discrepancy was used but, of course, this is not the only discrepancy measure we may consider. The main aims in this article are:

1. To carry out a comparison between the performance of the χ^2 discrepancy and the performance of other possible discrepancy measures.
2. To calibrate the discrepancy we find between the set of data and the selected model.

The article is organized as follows:

In Section 2, the method introduced in De la Horra (2008) is briefly explained (for any discrepancy measure). In Section 3, several discrepancy measures are proposed. In Section 4, these discrepancy measures are compared by simulation, and it is found that the χ^2 discrepancy is reasonable to use.

Once we have decided to use a discrepancy measure and we have chosen a model, we have to remember that this model is not to be understood as the true model (because nobody knows the true model) but as the best model among several possible models. The discrepancy between the data and the model is just a number, and it is very important to decide if this number indicates either a small discrepancy or a large discrepancy:

- If the discrepancy is small, the model we have chosen is a good model for our data.
- If the discrepancy is large, the model we have chosen is not a good model for our data.

This problem of calibrating discrepancy measures has been previously studied, for instance, by McCulloch (1989), Soofi et al. (1995), and Carota et al. (1996). In Section 5, an easy procedure for calibrating the discrepancy between the set of data and the selected model is considered. Some examples (control cases) are analysed for illustrating and evaluating this method.

Finally, a set of real data is analysed in Section 6.

2. A method for model selection

An easy method for Bayesian model selection was proposed and developed in De la Horra (2008). This method was based on the use of a discrepancy measure and it is briefly explained here for the continuous case (although a modification was also given for its application to the discrete case).

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample of a continuous random variable X. We have to choose among m different Bayesian models, M_i, $i = 1, \ldots, m$. Each

Bayesian model consists of two components: a sampling density, $f_i(x|\theta)$ (where $\theta \in \Theta$), and a prior density, $\pi_i(\theta)$. For the sake of simplicity, we are assuming that the parameter space, Θ, is the same for all the models, but this is not necessary. The cumulative distribution function corresponding to $f_i(x|\theta)$ will be denoted by $F_i(x|\theta)$; this cumulative distribution function will have a relevant role in the method. In short, we can write:

$$M_i = \{f_i(x|\theta), \pi_i(\theta)\},\ i = 1, \ldots, m.$$

The method is based on the following idea:

Let us assume that $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from a continuous random variable X with density function $f_i(x|\theta)$ and cumulative distribution function $F_i(x|\theta)$ (for θ fixed). It is well known that $(F_i(X_1|\theta), \ldots, F_i(X_n|\theta))$ can be considered as a random sample from a $U(0,1)$ (uniform distribution over the interval (0, 1)), because $F_i(X|\theta)$ follows a $U(0,1)$ distribution and, as a consequence, we hope that $(F_i(X_1|\theta), \ldots, F_i(X_n|\theta))$ will be well fitted by the $U(0,1)$ distribution.

We next describe the method in three steps:

(1) First of all, we measure the discrepancy between the sample we have obtained, $\mathbf{x} = (x_1, \ldots, x_n)$, and the distribution function $F_i(x|\theta)$ (for a fixed θ), by using a suitable discrepancy measure between $(F_i(x_1|\theta), \ldots, F_i(x_n|\theta))$ and the $U(0,1)$ distribution. This discrepancy will be denoted by $D_i(\mathbf{x}, \theta)$.

The idea behind this discrepancy is simple: if $F_i(x|\theta)$ (for a fixed θ) is a good model, $D_i(\mathbf{x}, \theta)$ will be close to zero; if $F_i(x|\theta)$ (for a fixed θ) is not a good model, $D_i(\mathbf{x}, \theta)$ will be far from zero.

(2) Of course, we are interested in evaluating the discrepancy between the sample we have obtained, $\mathbf{x} = (x_1, \ldots, x_n)$, and the whole Bayesian model, M_i. The Bayesian solution is easy; first of all, we compute the posterior density of the parameter,

$$\pi_i(\theta|\mathbf{x}) = \pi_i(\theta|x_1, \ldots, x_n) = \frac{f_i(x_1, \ldots, x_n|\theta)\pi_i(\theta)}{\int_\Theta f_i(x_1, \ldots, x_n|\theta)\pi_i(\theta)d\theta}$$
$$= \frac{f_i(x_1|\theta) \cdots f_i(x_n|\theta)\pi_i(\theta)}{\int_\Theta f_i(x_1|\theta) \cdots f_i(x_n|\theta)\pi_i(\theta)d\theta},$$

and then we evaluate the posterior expected discrepancy between the sample $\mathbf{x}$ and the model M_i:

$$D_i(\mathbf{x}) = \int_\Theta D_i(\mathbf{x}, \theta)\pi_i(\theta|\mathbf{x})d\theta.$$

(3) Finally, we only have to compare $D_1(\mathbf{x}), \ldots, D_m(\mathbf{x})$, and choose the Bayesian model having the smallest posterior expected discrepancy.

3. Some discrepancy measures

The χ^2 discrepancy was used and studied in De la Horra (2008). This discrepancy measure may be reasonable but, of course, it is not the only one we may consider.

In this section, several reasonable discrepancy measures are proposed. Remember that, in all the cases, we want to measure the discrepancy between $(F_i(x_1|\theta), \ldots, F_i(x_n|\theta))$ and the $U(0,1)$ distribution.

(1) $\boldsymbol{\chi^2}$ discrepancy

The discrepancy between $(F_i(x_1|\theta), \ldots, F_i(x_n|\theta))$ and the $U(0,1)$ distribution may be measured by the χ^2 discrepancy. For doing that, we partition the interval (0, 1) in k subintervals, $(0, 1/k), (1/k, 2/k), \ldots, ((k-1)/k, 1)$ and the χ^2 discrepancy is defined as usual:

$$D_i^1(\mathbf{x}, \theta) = \sum_{j=1}^{k} \frac{[O_{ij}(\theta) - n(1/k)]^2}{n(1/k)} = \sum_{j=1}^{k} \frac{[O_{ij}(\theta) - (n/k)]^2}{n/k},$$

where $O_{ij}(\theta)$ is the number of elements of $(F_i(x_1|\theta), \ldots, F_i(x_n|\theta))$ we have obtained in each subinterval.

(2) Kolmogorov-Smirnov discrepancy

Let $G_0(y)$ denote the cumulative distribution function of the $U(0,1)$, and let $G_i(y|\theta)$ denote the empirical cumulative distribution function corresponding to the sample $(F_i(x_1|\theta), \ldots, F_i(x_n|\theta))$. The Kolmogorov-Smirnov discrepancy is defined as usual:

$$D_i^2(\mathbf{x}, \theta) = \sup_{y \in (0,1)} |G_i(y|\theta) - G_0(y)|.$$

(3) $\boldsymbol{L^1}$ discrepancy

Let $g_0(y)$ denote the density function of the $U(0,1)$, and let $g_i(y|\theta)$ denote some reasonable density estimator obtained from $(F_i(x_1|\theta), \ldots, F_i(x_n|\theta))$. The L^1 discrepancy is defined as usual:

$$D_i^3(\mathbf{x}, \theta) = \int_0^1 |g_i(y|\theta) - g_0(y)| dy.$$

In the next section, a density estimator with an Epanechnikov kernel will be used for $g_i(y|\theta)$ (see, for instance, Silverman (1986)). Of course, other density estimators may be used.

(4) Intrinsic discrepancy

Let us consider again $g_0(y)$ (defined over $\mathcal{X}_0 = (0,1)$) and $g_i(y|\theta)$ (defined over $\mathcal{X}_i \subset (0,1)$). Bernardo and Rueda (2002) defined the intrinsic discrepancy as follows (see also Bernardo (2005), Berger et al. (2009)):

$$D_i^4(\mathbf{x},\theta) = \min\left\{ \int_{\mathcal{X}_i} g_i(y|\theta) \log \frac{g_i(y|\theta)}{g_0(y)} dy \, , \, \int_{\mathcal{X}_0} g_0(y) \log \frac{g_0(y)}{g_i(y|\theta)} dy \right\}$$
$$= \int_{\mathcal{X}_i} g_i(y|\theta) \log \frac{g_i(y|\theta)}{g_0(y)} dy \, ,$$

where the last equality follows because $\mathcal{X}_i \subset \mathcal{X}_0$ and the second integral in the first line is not finite (for general properties of the intrinsic discrepancy, see Bernardo (2005)). In the next section, a density estimator with an Epanechnikov kernel will be again used for $g_i(y|\theta)$.

4. Comparing discrepancy measures

First of all, we will compare the performance of the four discrepancy measures proposed in Section 3. For doing that, we will proceed by simulation as follows:

(1) Fix m Bayesian models, $M_i = \{f_i(x|\theta), \pi_i(\theta)\}$, $i = 1, \ldots, m$.

(2) Simulate a very large number of random samples from the Bayesian model M_i. Apply the method described in Section 2 to these samples, for the four discrepancy measures proposed in Section 3, and record the percentage of correct classification with each discrepancy.

(3) Repeat Step (2) for each model M_i, $i = 1, \ldots, m$. Construct a double entry table with the percentages of correct classification with each discrepancy measure and each model.

(4) Finally, look for the discrepancy measure having the best performance.

This algorithm is next applied to two examples. These examples are simple to describe but quite interesting, as explained below.

Example 1. We consider the three following Bayesian models:

$$M_1 = \{f_1(x|\theta) \sim N(\theta, \sigma = 1); \pi_1(\theta) \propto 1\}$$
$$M_2 = \{f_2(x|\theta) \sim N(\theta, \sigma = 2); \pi_2(\theta) \propto 1\}$$
$$M_3 = \{f_3(x|\theta) \sim N(\theta, \sigma = 3); \pi_3(\theta) \propto 1\}$$

In the three models, the sampling model is the normal distribution (with different standard deviations) and the prior density is the reference prior. We are considering three similar Bayesian models because, if the method has a good performance when similar models are compared, the performance of the method will be still better when the models are quite different.

We now apply the algorithm described at the beginning of this section. For doing that, we generate, for instance, 1000 random samples with 50 elements each from one of the sampling densities in Model 1 (for instance, from the $N(0, \sigma = 1)$ distribution). The improper prior in Model 1 is used for obtaining the posterior density:

$$\pi(\theta|x_1, \ldots, x_{50}) \sim N(\bar{x}; \sigma = 1/\sqrt{50}\,) \qquad (\text{model } M_1).$$

The same procedure is then carried out for Model 2 and Model 3. For these models, posterior densities are:

$$\pi(\theta|x_1, \ldots, x_{50}) \sim N(\bar{x}; \sigma = 2/\sqrt{50}\,) \qquad (\text{model } M_2)$$
$$\pi(\theta|x_1, \ldots, x_{50}) \sim N(\bar{x}; \sigma = 3/\sqrt{50}\,) \qquad (\text{model } M_3)$$

The percentages of correct classification for each discrepancy measure and each model are shown in Table 1.

In these examples, χ^2 discrepancies are computed by partitioning the interval (0, 1) into $k = 4$ subintervals (the number of subintervals must not be too small, but each subinterval must contain a reasonable number of observations). ■

Table 1: *Percentages of correct classification in Example 1.*

	χ^2	$K-S$	L^1	Intrinsic
M_1	100%	100%	100%	100%
M_2	99%	96%	77%	76%
M_3	92%	89%	43%	45%

Example 2. We consider the three following Bayesian models:

$$M_1 = \{f_1(x|\theta) \sim N(\theta, \sigma = 1); \pi_1(\theta) \propto 1\}$$
$$M_2 = \{f_2(x|\theta) \sim N(\theta, \sigma = 5); \pi_2(\theta) \propto 1\}$$
$$M_3 = \{f_3(x|\theta) \sim N(\theta, \sigma = 10); \pi_3(\theta) \propto 1\}$$

In the three models, the sampling model is again the normal distribution (with different standard deviations) and the prior density is again the reference prior, but now the models are more different than in Example 1, because the standard deviations are more different. So, it is expected that the percentages of correct classification will be better than in Example 1 for all the discrepancy measures.

We now apply the algorithm described at the beginning of this section, in a similar way to Example 1. The percentages of correct classification with each discrepancy measure and each model are shown in Table 2. ■

Table 2: *Percentages of correct classification in Example 2.*

	χ^2	$K-S$	L^1	Intrinsic
M_1	100%	100%	100%	100%
M_2	99%	100%	100%	100%
M_3	98%	98%	79%	66%

Main conclusions

- The global performance of the χ^2 discrepancy is the best one in these examples.
- Of course, another discrepancy measure may have a better performance in other cases, but the point here is that the χ^2 discrepancy is reasonable to use. As a consequence, the χ^2 discrepancy will be used in the following sections.

5. Calibrating the discrepancy

Now, we have to choose among m different Bayesian models, M_i, $i = 1, \ldots, m$, trying to find the best model for our data. It is important to remark that the model we choose is not to be understood as the true model (nobody knows the true model), but as the best model for our data we can find among several possible models.

Therefore, we have to answer the following question: is the model we have finally chosen good enough for our data? A reasonable procedure for answering this question is given next.

Once we have chosen M_i as the best model among $M_1,\ldots,M_m$, $D_i(\mathbf{x})$ is just a number giving the posterior expected discrepancy between our data, $\mathbf{x}$, and the model, M_i. Now, it is important to calibrate this number:

- If the discrepancy is small, the model we have chosen is a good model for our data.
- If the discrepancy is large, the model we have chosen is not a good model for our data.

For deciding if the discrepancy, $D_i(\mathbf{x})$, between our data, $\mathbf{x}$, and the model, M_i, is either large or small, we may proceed as follows:

(1) Simulate a very large number of random samples from the Bayesian model, M_i, and compute the posterior expected discrepancies between each of these samples and M_i.

(2) Compare $D_i(\mathbf{x})$ to the posterior expected discrepancies we have computed in Step (1), for obtaining in what percentile $D_i(\mathbf{x})$ is placed.

This procedure is applied next to some examples with simulated data. The aim of these examples is to evaluate the behaviour of the procedure in these control cases.

Example 3. A random sample is simulated from a $N(0,\sigma=2)$ distribution.

Consider, as possible models, the three following Bayesian models:

$$M_1 = \{f_1(x|\theta) \sim N(\theta,\sigma=1); \pi_1(\theta) \propto 1\}$$
$$M_2 = \{f_2(x|\theta) \sim N(\theta,\sigma=2); \pi_2(\theta) \propto 1\}$$
$$M_3 = \{f_3(x|\theta) \sim N(\theta,\sigma=3); \pi_3(\theta) \propto 1\}$$

We apply the algorithm described in Section 2. The model M_2 is chosen, because the smallest posterior expected discrepancy, $D_2(x) = 8.05$, is obtained from M_2. Now, we have to calibrate this value, so we simulate 1000 random samples from model M_2. It is found that the discrepancy $D_2(x) = 8.05$ is between percentiles 12 and 13. Therefore, in this case, model M_2 is a very good model for our data. This is a very reasonable result because all the data did come from M_2. ■

Example 4. A random sample is simulated in which 5% of the elements come from a $N(0,\sigma=1)$ distribution, 90% from a $N(0,\sigma=2)$ distribution, and 5% from a $N(0,\sigma=3)$ distribution.

Consider again, as possible models, the three Bayesian models given in Example 3, and apply the algorithm described in Section 2. The model M_2 is again chosen, because

the smallest posterior expected discrepancy, $D_2(x) = 12.16$, is obtained from M_2. To calibrate this value we simulate 1000 random samples from model M_2, and it is found that the discrepancy $D_2(x) = 12.16$ is between percentiles 56 and 57. Therefore, in this case, model M_2 is still a good model for our data, although the discrepancy is larger than in Example 3. This is again a very reasonable result because, in this case, almost all the data came from M_2. ■

Example 5. Finally, we simulate a random sample in which 33% of the elements come from a $N(0, \sigma = 1)$ distribution, 34% from a $N(0, \sigma = 2)$ distribution, and 33% from a $N(0, \sigma = 3)$ distribution.

Again the three Bayesian models given in Example 3 are considered and the algorithm described in Section 2 is applied. The model M_2 is chosen again, because the smallest posterior expected discrepancy, $D_2(x) = 17.58$, is obtained from M_2.

Calibrating as before, we simulate 1000 random samples from model M_2, and it is found that the discrepancy $D_2(x) = 17.58$ is between percentiles 92 and 93. Therefore, in this case, model M_2 is not a good model for our data and, once more, this is a very reasonable result. ■

Main conclusions

- This method for calibrating the discrepancy shows a good behaviour in these controlled situations.
- As a consequence, the method can be applied to a set of real data with reasonable confidence. This is carried out in the next section.

6. Application to real data

A set of 30 failure times for air conditioners on an airplane was introduced by Proschan (1963). This set of real data was analysed first by Berger and Pericchi (1996) and then by Gutiérrez-Peña and Walker (2001). They consider three Bayesian models for explaining this set of real data, with exponential, lognormal and Weibull densities as sampling densities:

$$M_1 = \left\{ f_1(x|\theta) = \frac{1}{\theta} \exp\left\{ -\frac{x}{\theta} \right\}; \pi_1(\theta) \propto \frac{1}{\theta} \right\}$$

$$M_2 = \left\{ f_2(x|\mu, \sigma^2) = \frac{\exp\{-(\log x - \mu)^2/(2\sigma^2)\}}{\sqrt{2\pi} x \sigma}; \pi_2(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \right\}$$

$$M_3 = \left\{ f_3(x|\alpha, \beta) = \beta x^{(\beta-1)} \alpha^{-\beta} \exp\left\{ -(x/\alpha)^\beta \right\}; \pi_3(\alpha, \beta) \propto \frac{1}{\alpha\beta} \right\}$$

The corresponding posterior distributions are:

$$\pi_1(\theta|x) \sim (2n\bar{x})\chi^{-2}_{2n}$$

$$\pi_2(\mu,\sigma^2|x) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\log^2 x_i - n(\overline{\log x})^2\right)\right\} \exp\left\{-\frac{n}{2\sigma^2}(\mu-\overline{\log x})^2\right\}$$

$$\sim NIG\left(n-1, \frac{1}{n-1}\left(\sum_{i=1}^{n}\log^2 x_i - n(\overline{\log x})^2\right), n, \overline{\log x}\right)$$

$$\pi_3(\alpha,\beta|x) \propto \beta^{n-1}\left[\prod_{i=1}^{n} x_i\right]^{\beta-1} \alpha^{-n\beta-1}\exp\left\{-\frac{1}{\alpha^\beta}\sum_{i=1}^{n}x_i^\beta\right\} \quad ,$$

where χ^{-2}_{2n} denotes the "inverse chi-square distribution", $\overline{\log x} = \frac{1}{n}\sum_{i=1}^{n}\log x_i$ and NIG denotes the normal inverse gamma distribution.

Next, we show and comment the results, when different methods are applied:

(1) Method by Berger and Pericchi. They obtained that the model M_1 is preferred to M_2, and the model M_2 is preferred to M_3.

(2) Method by Gutiérrez-Peña and Walker. They obtained that the models M_1 and M_3 are preferred to M_2.

First of all, we remark that the results are different for the two methods shown above.

(3) Method in this article. We apply our method to these data (notice that the third posterior distribution is not in closed form, and so, a Markov chain Monte Carlo (MCMC) method is needed for simulations). The following discrepancies are obtained:

$$D_1(\mathbf{x}) = 6.5$$
$$D_2(\mathbf{x}) = 2.8$$
$$D_3(\mathbf{x}) = 11.8$$

Therefore, with this method, the model M_2 is preferred to M_1 and M_3. This result is also different from the results obtained with the other two methods. The calibration of the discrepancies, by using the algorithm proposed in Section 5, throws light upon these results:

- It is obtained (by simulation) that the discrepancy between the real data and the model M_1, $D_1(\mathbf{x}) = 6.5$, is between percentiles 94 and 95. Therefore M_1 is a bad (although not very bad) model for these real data.

- It is obtained (by simulation) that the discrepancy between the real data and the model M_2, $D_2(\mathbf{x}) = 2.8$, is between percentiles 74 and 75. Therefore M_2 is a reasonable (although not especially good) model for these real data.
- It is obtained (by simulation) that the discrepancy between the real data and the model M_3, $D_3(\mathbf{x}) = 11.8$, is between percentiles 90 and 91. Therefore M_3 is a bad (although not very bad) model for these real data.

Main conclusions

- The best model according to our method is M_2 (the lognormal model) and is a reasonable model for these real data, because the calibration shows that this discrepancy is between percentiles 74 and 75.
- The discrepancies for the models M_1 and M_3 are larger than the discrepancy for the model M_2. It is important to notice that their calibrations are bad, but not very bad. Possibly, this is the reason why they were chosen when other methods are used.

Acknowledgements

This research has been partially supported by grant MTM2010-17366 (the first author) and under grant MTM2008-03282 (the second author) from the Spanish ministry of Science and Innovation. We thank the referees for useful suggestions that improved the article.

References

Aitkin, M. (1991). Posterior Bayes factors (with discussion). *Journal of the Royal Statistical Society B*, 53, 11–142.

Berger, J. O., Bernardo, J. M. and Sun, D. (2009). The formal definition of reference priors. *Annals of Statistics*, 37, 905–938.

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factors for model selection and prediction. *Journal of the American Statistical Association*, 91, 109–122.

Bernardo, J. M. (2005). Intrinsic credible regions: an objective Bayesian approach to interval estimation (with discussion). *Test*, 14, 317–384.

Bernardo, J. M. and Rueda, R. (2002). Bayesian hypothesis testing: a reference approach. *International Statistical Review*, 70, 351–372.

Carota, C. Parmigiani, G. and Polson, N. G. (1996). Diagnostic measures of model criticism. *Journal of the American Statistical Association*, 91, 753–762.

De la Horra, J. (2008). Bayesian model selection: Measuring the χ^2 discrepancy with the uniform distribution. *Communications in Statistics-Theory and Methods*, 37, 1412–1424.

De la Horra, J. and Rodríguez-Bernal, M. T. (2005). Bayesian model selection: a predictive approach with losses based on distances L^1 and L^2. *Statistics & Probability Letters*, 71, 257–265.

De la Horra, J. and Rodríguez-Bernal, M. T. (2006). Prior density selection as a particular case of Bayesian model selection: a predictive approach. *Communications in Statistics-Theory and Methods*, 35, 1387–1396.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 153–160.

Gelfand, A. E. (1995). Model determination using sampling-based methods. In: Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 145–161.

Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (Eds.), *Bayesian Statistics 4*. Oxford University Press, Oxford, 147–167.

Gelfand, A. E. and Ghosh, S. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85, 1–11.

Gutiérrez-Peña, E. and Walker, S. G. (2001). A Bayesian predictive approach to model selection. *Journal of Statistical Planning and Inference*, 93, 259–276.

Laud, P. W. and Ibrahim, J. G. (1995). Predictive model selection. *Journal of the Royal Statistical Society B*, 57, 247–262.

McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association*, 84, 473–478.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society B*, 57, 99–138.

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics*, 5, 375–383.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

San Martini, A. and Spezzaferri, F. (1984). A predictive model selection criterion. *Journal of the Royal Statistical Society B*, 46, 296–303.

Soofi, E. S., Ebrahimi, N. and Habibullah, M. (1995). Information distinguishability with application to the analysis of failure data. *Journal of the American Statistical Association*, 90, 657–668.

Spiegelhalter, D. J. and Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society B*, 44, 377–387.

Trottini, M. and Spezzaferri, F. (2002). A generalized predictive criterion for model selection. *Canadian Journal of Statistics*, 30, 79–96.

SORT 36 (1) January-June 2012, 81-102

Decision-making techniques with similarity measures and OWA operators

José M. Merigó[1] and Anna M. Gil-Lafuente

Abstract

We analyse the use of the ordered weighted average (OWA) in decision-making giving special attention to business and economic decision-making problems. We present several aggregation techniques that are very useful for decision-making such as the Hamming distance, the adequacy coefficient and the index of maximum and minimum level. We suggest a new approach by using immediate weights, that is, by using the weighted average and the OWA operator in the same formulation. We further generalize them by using generalized and quasi-arithmetic means. We also analyse the applicability of the OWA operator in business and economics and we see that we can use it instead of the weighted average. We end the paper with an application in a business multi-person decision-making problem regarding production management.

MSC: 62C86

Keywords: Decision-making, OWA operator, similarity measure, Hamming distance, production management.

1. Introduction

Decision-making problems are very common in the literature (Canós and Liern 2008; Figueira et al 2005; A.M. Gil-Lafuente and Merigó 2010; Torra and Narukawa 2007; Wei 2009; Wei et al 2010; 2011). They are very useful in a lot of situations because people are almost all the time taking decisions. Sometimes, they take unconscious decisions or sometimes they simply take the usual decisions of their lives such as what to eat, what to see on TV and so on. In business and economics, people and organizations also take decisions almost all the time. Sometimes, they take decisions on how to do or

[1] Corresponding author: Tel: +34 93 402 19 62; Fax: +34 93 403 98 82. jmerigo@ub.edu
Department of Business Administration, University of Barcelona, Av. Diagonal 690, 08034, Barcelona, Spain.
Received: September 2011
Accepted: January 2012

improve their work or sometimes the decisions are more global and affect a lot of decision-makers. Obviously, in these situations we also find a lot of unconscious decisions.

For the development of the decision-making process we can use a lot of tools for taking decisions such as individual decision-making, group decision-making, multi person decision-making, sequential decision-making and different statistical techniques. Among the different statistical techniques that we can use in decision-making, a very useful one is the aggregation operator because it permits to aggregate the information and obtain a single result that permits to continue with the decision process and make the decision. It is worth noting the ordered weighted averaging (OWA) operator (Yager 1988). It is a tool that provides a parameterized family of aggregation operators between the maximum and the minimum. Since its appearance, the OWA operator has been used in a wide range of studies and applications (Merigó et al 2010; Xu 2005; Xu and Da 2003; Yager 1993; 2004a; Yager and Kacprzyk 1997; Yager, Kacprzyk and Beliakov 2011; Zhao et al 2010; Zhou and Chen 2010).

In business and economics, it is very useful to use different similarity measures that also use aggregation operators such as the Hamming distance (Hamming, 1950). The Hamming distance is a very useful tool in decision-making because it permits to compare the available results with some ideal ones that are supposed to be the best ones. This is especially useful because, depending on the particular problem we are looking at, the best results are not always the best for the decision-maker because he may have different interests. An extreme example of this would be the concept of dumping, which means that the seller is selling the product with a price that is lower than its production cost. Thus, in the decision process of fixing this price, the seller obviously is looking for an ideal that it is not the best one. In the literature, we find a lot of studies that analyse the concept of the Hamming distance (Karayiannis 2000; Kaufmann 1975; Kaufmann and Gil-Aluja 1986; 1987; Xu 2010a; 2010b).

Recently, several authors (Karayiannis 2000; Merigó 2008; Merigó and Casanovas 2010a; 2011a; Merigó and A.M. Gil-Lafuente 2007; 2010; Xu and Chen 2008) have analysed the use of the OWA operator in the Hamming distance. We can refer to this new aggregation operator as the ordered weighted averaging distance (OWAD) operator. Its main advantage is that it provides a parameterized family of distance aggregation operators between the maximum and the minimum distance. The OWAD operator can be further extended by using other types of distances such as the Euclidean distance, the Minkowski distance and the quasi-arithmetic distance (Karayiannis 2000; Merigó 2008; Merigó and Casanovas 2011b; 2011c; Merigó and A.M. Gil-Lafuente 2011).

Other similarity measures that are very useful in business and economics are the adequacy coefficient (Kaufmann and Gil-Aluja 1986; 1987) and the index of maximum and minimum level (J. Gil-Lafuente 2001; 2002). The adequacy coefficient is an extension of the Hamming distance that analyses the results that are higher than the ideal by using a t-norm. This approach can also be extended by using the OWA operator, obtaining the ordered weighted averaging adequacy coefficient (OWAAC) operator

(Merigó and A.M. Gil-Lafuente 2010). Further developments on the OWAAC can be found in Merigó (2008) and Merigó et al. (2011a). The index of maximum and minimum level is a model that uses the Hamming distance and the adequacy coefficient in the same formulation using the one that is more appropriate for each variable considered. This tool can also be extended by using the OWA operator, forming the ordered weighted averaging index of maximum and minimum level (OWAIMAM) operator.

The aim of this paper is to introduce new decision-making techniques based on the use of the OWA operator and the weighted average in order to obtain a formulation that it is able to deal with the subjective beliefs of the decision-maker and with his attitudinal character. For doing so, we use the concept of immediate probabilities (Engemann et al 1996; Yager et al 1995) applied in situations where we use weighted averages instead of probabilities. Thus, we obtain the concept of immediate weights. We suggest the use of immediate weights with the OWAD operator, the OWAAC operator and the OWAIMAM operator. Therefore, we get the immediate weighted OWAD (IWOWAD), the immediate weighted OWAAC (IWOWAAC) and the immediate weighted OWAIMAM (IWOWAIMAM) operator. The main advantage of these similarity measures is that they are able to deal with the weighted Hamming distance and with the OWAD operator in the same formulation. Thus, we are able to represent the information in a more complete way because we can consider the degree of importance of the characteristics and the degree of "orness", that is, the tendency of the aggregation to the minimum or to the maximum. Thus, we can under or over estimate the results according to the interests we have in the aggregation.

We also extend this analysis by using generalized and quasi-arithmetic means, obtaining the generalized IWOWAD (GIWOWAD), the generalized IWOWAAC (GIWOWAAC) and the generalized IWOWAIMAM (GIWOWAIMAM) operator. The main advantage of these new generalizations is that they include a wide range of particular cases, including the usual arithmetic, geometric and quadratic aggregations. Thus, we obtain a more general formulation that permits to analyse the aggregation problem from different contexts.

We also analyse a wide range of applications that can be developed. Specially, we focus on a wide range of decision-making problems that can be implemented in business and economic scenarios. We study a business decision-making problem in production management by using different multi-person decision-making techniques based on the OWA operator such as the IWOWAD operator, the IWOWAAC operator and the IWOWAIMAM operator. Thus, we are able to construct new aggregation operators including the multi-person IWOWAD (MP-IWOWAD), the multi-person IWOWAAC (MP-IWOWAAC) and the multi-person IWOWAIMAM (MP-IWOWAIMAM). The main advantage of these aggregation methods is that they are able to deal with the opinion of several persons in the analysis providing results that represents the aggregated information of the group. We see that each method provides different results depending on the interests of the decision-maker. Therefore, we see that the results may lead to different decisions depending on the particular type of aggregation operator used. Thus,

we get a general overview of the different scenarios that may occur and select the one that is in more accordance with our interests.

The paper is organized as follows. In Section 2, we briefly review some basic decision-making techniques such as the Hamming distance, the adequacy coefficient, the index of maximum and minimum level and the OWA operator. Section 3 presents new decision-making techniques based on the use of immediate weights. Section 4 summarizes different applications that can be developed with the OWA operator in business and economics. In Section 5, we present a particular problem in a decision-making problem about the selection of production strategies in a company. In Section 6 we present a numerical example and in Section 7 we summarize the main conclusions of the paper.

2. Preliminaries

In this section, we briefly review some basic concepts to be used throughout the paper such as the Hamming distance, the adequacy coefficient, the index of maximum and minimum level and their extensions with the OWA operator.

2.1. The Hamming distance

The Hamming distance (Hamming 1950) is a useful technique for calculating the differences between two elements, two sets, etc. In fuzzy set theory, it can be useful, for example, for the calculation of distances between fuzzy sets, interval-valued fuzzy sets, intuitionistic fuzzy sets and interval-valued intuitionistic fuzzy sets. For two sets A and B, the weighted Hamming distance can be defined as follows.

Definition 1 *A weighted Hamming distance of dimension n is a mapping* d_{WH}*:* $R^n \times R^n \to R$ *that has an associated weighting vector W of dimension n with the sum of the weights being 1 and* $w_j \in [0,1]$ *such that:*

$$d_{WH}(\langle x_1, y_1\rangle, \ldots, \langle x_n, y_n\rangle) = \sum_{i=1}^{n} w_i |x_i - y_i|, \tag{1}$$

where x_i *and* y_i *are the ith arguments of the sets X and Y.*

Note that the formulations shown above are the general expressions. For the formulation used in fuzzy set theory, see for example (Kaufmann 1975). Note also that if $w_i = 1/n$, for all i, then, the weighted Hamming distance becomes the normalized Hamming distance.

2.2. The adequacy coefficient

The adequacy coefficient (Kaufmann and Gil-Aluja 1986; 1987) is an index used for calculating the differences between two elements, two sets, etc. It is very similar to the Hamming distance with the difference that it neutralizes the result when the comparison shows that the real element is higher than the ideal one. For two sets A and B, the weighted adequacy coefficient can be defined as follows.

Definition 2 *A weighted adequacy coefficient of dimension n is a mapping* $K : [0,1]^n \times [0,1]^n \to [0,1]$ *that has an associated weighting vector W of dimension n with the sum of the weights 1 and* $w_j \in [0,1]$ *such that:*

$$K(\langle x_1, y_1\rangle, \ldots, \langle x_n, y_n\rangle) = \sum_{i=1}^{n} w_i[1 \wedge (1 - x_i + y_i)], \tag{2}$$

where x_i *and* y_i *are the ith arguments of the sets X and Y.*

Note that if $w_i = 1/n$, for all i, then, the weighted adequacy coefficient becomes the normalized adequacy coefficient.

2.3. The index of maximum and minimum level

The index of maximum and minimum level is an index that unifies the Hamming distance and the adequacy coefficient in the same formulation (J. Gil-Lafuente 2001; 2002). For two sets A and B, the weighted index of maximum and minimum level can be defined as follows.

Definition 3 *A WIMAM of dimension n is a mapping* $K : [0,1]^n \times [0,1]^n \to [0,1]$ *that has an associated weighting vector W of dimension n with the sum of the weights 1 and* $w_j \in [0,1]$ *such that:*

$$\eta(\langle x_1, y_1\rangle, \ldots, \langle x_n, y_n\rangle) = \sum_{u} Z_i(u) \times |x_i(u) - y_i(u)| + \sum_{v} Z_i(v) \times [0 \vee (x_i(v) - y_i(v))], \tag{3}$$

where x_i *and* y_i *are the ith arguments of the sets X and Y.*

Note that if $w_i = 1/n$, for all i, then, the weighted index of maximum and minimum level becomes the normalized index of maximum and minimum level.

2.4. The OWA operator

The OWA operator (Yager 1988) provides a parameterized family of aggregation operators which have been used in many applications (Beliakov et al 2007; Merigó 2008; Xu 2005; Yager 1993; Yager and Kacprzyk 1997). It can be defined as follows.

Definition 4 *An OWA operator of dimension n is a mapping OWA:* $R^n \to R$ *that has an associated weighting vector W of dimension n with* $w_j \in [0,1]$ *and* $\sum_{j=1}^{n} w_j = 1$*, such that:*

$$\text{OWA}(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j, \tag{4}$$

where b_j*is the jth largest of the* a_i*.*

The OWAD operator (Merigó 2008; Merigó and A.M. Gil-Lafuente 2007; 2010) is an aggregation operator that uses OWA operators and distance measures in the same formulation. In this subsection, we focus on the Hamming distance. However, it is worth noting that it is also possible to use other types of distance measures with the OWA operator such as the Euclidean or the Minkowski distance (Merigó 2008). It can be defined as follows for two sets X and Y.

Definition 5 *An OWAD operator of dimension n is a mapping OWAD:* $R^n \times R^n \to R$ *that has an associated weighting vector W, with* $\sum_{j=1}^{n} w_j = 1$ *and* $w_j \in [0,1]$ *such that:*

$$\text{OWAD}(\langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle) = \sum_{j=1}^{n} w_j D_j, \tag{5}$$

where D_j *represents the jth largest of the* $|x_i - y_i|$*.*

The OWAAC operator (Merigó and A.M. Gil-Lafuente 2010) is an aggregation operator that uses the adequacy coefficient and the OWA operator in the same formulation. It can be defined as follows for two sets X and Y.

Definition 6 *An OWAAC operator of dimension n is a mapping OWAAC:* $[0,1]^n \times [0,1]^n \to [0,1]$ *that has an associated weighting vector W, with* $w_j \in [0,1]$ *and* $\sum_{j=1}^{n} w_j = 1$*, such that:*

$$\text{OWAAC}(\langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle) = \sum_{j=1}^{n} w_j K_j, \tag{6}$$

where K_j *represents the jth largest of* $[1 \wedge (1 - x_i + y_i)], x_i, y_i \in [0,1]$.

The OWAIMAM operator (Merigó 2008; Merigó et al. 2011b) is an aggregation operator that uses the Hamming distance, the adequacy coefficient and the OWA operator in the same formulation. It can be defined as follows.

Definition 7 *An OWAIMAM operator of dimension n, is a mapping OWAIMAM:* $[0,1]^n \times [0,1]^n \to [0,1]$ *that has an associated weighting vector W, with* $w_j \in [0,1]$ *and the sum of the weights is equal to 1, such that:*

$$\text{OWAIMAM}\ (\langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle) = \sum_{j=1}^{n} w_j K_j, \tag{7}$$

where K_j *represents the jth largest of all the* $|x_i - y_i|$ *and the* $[0 \vee (x_i - y_i)]$.

3. Distance measures with immediate weights

In this section, we introduce a new approach for dealing with distance measures where we use the weighted average and the OWA operator in the same formulation. For doing so, we extend the concept of immediate probabilities (Engemann et al 1996; Merigó 2008; 2010; Yager et al 1995) for situations where we use the weighted average. Thus, instead of using immediate probabilities, we will use immediate weights in the analysis. Extending this to the use of distance measures implies the introduction of new distance and similarity measures such as the immediate weighted OWA distance (IWOWAD), the immediate weighted OWAAC (IWOWAAC) and the immediate weighted OWAIMAM (IWOWAIMAM) operator. The main advantage of these new models is that they can consider the information used in the weighted average (degree of importance) and in the OWA operator (degree of orness or optimism) in the same formulation. Thus, we get a more general formulation that is able to represent the information in a more complete way because in real world problems, it is very common that we have to combine in the same problem situations with weighted averages and with OWA operators. Before defining these three new distance aggregation operators let us recall the concept of immediate probabilities applied to the weighted average, that is, the immediate weights (IW). It can be defined as follows.

Definition 8 *An IW operator of dimension n is a mapping IW: $R^n \to R$ that has an associated weighting vector W of dimension n with $w_j \in [0,1]$ and $\sum_{j=1}^{n} w_j = 1$, such that:*

$$\text{IW}\ (a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j b_j, \tag{8}$$

where b_j is the jth largest of the a_i, each a_i has associated a WA v_i, v_j is the associated WA of b_j, and $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$.

As we can see, if $w_j = 1/n$ for all j, we get the weighted average and if $v_j = 1/n$ for all j, the OWA operator. Now, we extend the measures commented in Section 2.4., by using immediate weights. Thus, for the OWAD operator, we get the IWOWAD operator and it is defined as follows.

Definition 9 *An IWOWAD operator of dimension n is a mapping IWOWAD: $R^n \times R^n \to R$ that has an associated weighting vector W of dimension n with $w_j \in [0,1]$ and $\sum_{j=1}^{n} w_j = 1$, such that:*

$$\text{IWOWAD}(\langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle) = \sum_{j=1}^{n} \hat{v}_j b_j, \tag{9}$$

where b_j is the jth largest of the $|x_i - y_i|$, each $|x_i - y_i|$ has associated a WA v_i, v_j is the associated WA of b_j, and $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$.

In this case, if $w_j = 1/n$ for all j, we get the weighted Hamming distance and if $v_j = 1/n$ for all j, the OWAD operator. Note that the IWOWAD operator accomplishes similar properties that the OWAD operator with the exception of commutativity because the use of the weighted average does not allow the commutativity property. Note that if the weighting vector is not normalized, i.e., $\hat{V} = \sum_{j=1}^{n} \hat{v}_j \neq 1$, then, the IWOWAD operator can be expressed as: $IWOWAD/\hat{V}$.

If we use immediate weights in the OWAAC operator, we get the IWOWAAC operator. In this case, we have the same expression than in Eq. (9) with the difference that now b_j is the jth largest of the $[1 \wedge (1 - x_i + y_i)]$, x_i, $y_i \in [0,1]$ and each $[1 \wedge (1 - x_i + y_i)]$ has associated a WA v_i.

As we can see, if $w_j = 1/n$ for all j, we get the weighted adequacy coefficient and if $v_j = 1/n$ for all j, the OWAAC operator. Moreover, if $x_i \geq y_i$, for all i, then, the OWAAC operator becomes the OWAD operator.

Finally, if we use the OWAIMAM operator with immediate weights, we get the IWOWAIMAM operator. Note that we get the same formulation than Eq. (9) with the difference that now b_j is the jth largest of all the $|x_i - y_i|$ and the $[0 \vee (x_i - y_i)]$; x_i, $y_i \in [0,1]$, and each $|x_i - y_i|$ and $[0 \vee (x_i - y_i)]$ has associated a WA v_i.

Furthermore, we can a present a further generalization of the previous measures by using generalized and quasi-arithmetic means (Merigó and Casanovas 2010b; 2010c; 2011d; Merigó and Gil-Lafuente 2009). Note that in this paper we will use generalized means although it is straightforward to extend it by replacing the parameter λ of the generalized mean by the strictly continuous monotonic function g of the quasi-arithmetic mean (Merigó and Gil-Lafuente 2009). By generalizing the IWOWAD operator, we get the generalized IWOWAD (GIWOWAD) operator. It can be defined as follows.

Definition 10 *An GIWOWAD operator of dimension n is a mapping GIWOWAD:* $R^n \times R^n \to R$ *that has an associated weighting vector W of dimension n with* $w_j \in [0,1]$ *and* $\sum_{j=1}^{n} w_j = 1$*, such that:*

$$\text{GIWOWAD}(\langle |x_1, y_1\rangle, \ldots, \langle x_n, y_n\rangle) = \left(\sum_{j=1}^{n} \hat{v}_j b_j^{\lambda}\right)^{1/\lambda}, \tag{10}$$

where b_j *is the jth largest of the* $|x_i - y_i|$*, each* $|x_i - y_i|$ *has associated a WA* v_i*,* v_j *is the associated WA of* b_j*,* $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$*, and* λ *is a parameter such that* $\lambda \in (-\infty, \infty) - \{0\}$.

Note that if we extend the IWOWAD operator with quasi-arithmetic means we get the quasi-arithmetic IWOWAD (Quasi-IWOWAD) operator as follows:

$$\text{Quasi-IWOWAD}\left(\langle x_1, y_1\rangle, \ldots, \langle x_n, y_n\rangle\right) = g^{-1}\left(\sum_{j=1}^{n} \hat{v}_j g\left(b_j\right)\right), \tag{11}$$

where $g(b)$ is a strictly continuous monotonic function. Further generalizations in this direction can be developed by using norm aggregations following Yager (2010). If we generalize the IWOWAAC operator, we obtain the generalized IWOWAAC (GIWOWAAC) operator. It can be defined as follows.

Definition 11 *A GIWOWAAC operator of dimension n is a mapping GIWOWAAC:* $[0,1]^n \times [0,1]^n \to [0,1]$ *that has an associated weighting vector W of dimension n with* $w_j \in [0,1]$ *and* $\sum_{j=1}^{n} w_j = 1$*, such that:*

$$\text{GIWOWAAC}\left(\langle x_1, y_1\rangle, \ldots, \langle x_n, y_n\rangle\right) = \left(\sum_{j=1}^{n} \hat{v}_j b_j^{\lambda}\right)^{1/\lambda}, \tag{12}$$

where b_j *is the jth largest of the* $[1 \wedge (1 - x_i + y_i)]$*,* x_i*,* $y_i \in [0,1]$*, each* $[1 \wedge (1 - x_i + y_i)]$ *has associated a WA* v_i*,* v_j *is the associated WA of* b_j*,* $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$*, and* λ *is a parameter such that* $\lambda \in (-\infty, \infty) - \{0\}$*.*

And if we extend the IWOWAIMAM operator by using generalized means, we get the generalized IWOWAIMAM (GIWOWAIMAM) operator. It is defined as follows.

Definition 12 *A GIWOWAIMAM operator of dimension n is a mapping GIWOWAIMAM:* $[0,1]^n \times [0,1]^n \to [0,1]$ *that has an associated weighting vector W of dimension n with* $w_j \in [0,1]$ *and* $\sum_{j=1}^{n} w_j = 1$*, such that:*

$$\text{GIWOWAIMAM}\left(\langle x_1, y_{11}\rangle, \ldots, \langle x_n, y_n\rangle\right) = \left(\sum_{j=1}^{n} \hat{v}_j b_j^{\lambda}\right)^{1/\lambda}, \tag{13}$$

where b_j *is the jth largest of all the* $|x_i - y_i|$ *and the* $[0 \vee (x_i - y_i)]$*;* x_i*,* $y_i \in [0,1]$*, each* $|x_i - y_i|$ *and* $[0 \vee (x_i - y_i)]$ *has associated a WA* v_i*,* v_j *is the associated WA of* b_j*,* $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$*, and* λ *is a parameter such that* $\lambda \in (-\infty, \infty) - \{0\}$*.*

Note that in these two cases we can also consider the dual. Additionally, if we use quasi-arithmetic means we get the quasi-arithmetic IWOWAAC (Quasi-IWOWAC) and the quasi-arithmetic IWOWAIMAM (Quasi-IWOWAIMAM) operator.

These generalizations include a wide range of particular cases by using different types of weighting vectors and values in the parameter λ. In Table 1, we present some of the main particular cases.

Note that a lot of other families could be studied following the OWA literature for obtaining OWA weights. The main advantage of using these generalizations is that they provide a more robust formulation that includes a wide range of particular cases. Thus, we get a deeper picture of the different results that may occur in the specific problem considered.

Table 1: *Families of GIWOWAD, GIWOWAAC and GIWOWAIMAM operators.*

Particular type	**GIWOWAD**	**GIWOWAAC**	**GIWOWAIMAM**
$w_i = 1/n, \forall\, i$	OWAD	OWAAC	OWAIMAM
$v_i = 1/n, \forall\, i$	WHD	WAC	WIMAM
$g(a) = a^{\lambda}$	Quasi-IWOWAD	Quasi-IWOWAAC	Quasi-IWOWAIMAM
$\lambda = 1$	IWOWAD	IWOWAAC	IWOWAIMAM
$\lambda = 2$	Quadratic IWOWAD	Quadratic IWOWAAC	Quadratic IWOWAIMAM
$\lambda \to 0$	Geometric IWOWAD	Geometric IWOWAAC	Geometric IWOWAIMAM
$\lambda = -1$	Harmonic IWOWAD	Harmonic IWOWAAC	Harmonic IWOWAIMAM
$\lambda = 3$	Cubic IWOWAD	Cubic IWOWAAC	Cubic IWOWAIMAM
$\lambda \to \infty$	Maximum distance	Maximum adequacy coefficient	Maximum IMAM
$\lambda \to -\infty$	Minimum distance	Minimum adequacy coefficient	Minimum IMAM
Etc.			

4. Applicability of the OWA operator in business and economics

The OWA operator is a very useful tool for business and economics because it permits to reflect the attitudinal character (the degree of orness or optimism) of the decision-maker in the aggregation of the information.

First of all, it is clear that the OWA operator plays a key role in decision-making problems by unifying the classical decision criteria under uncertainty, that is, the optimistic criteria, the pessimistic criteria, the Laplace and the Hurwicz criteria. Thus, we can use them in a lot of situations such as individual decision-making, group decision-making, multi-attribute decision-making, multi-criteria decision-making, multi-person decision-making, sequential decision-making and dynamic decision-making.

Moreover, we can apply it in a lot of other decision-making contexts such as probabilistic decision-making (Engemann et al 1996; Yager et al 1995), minimization of regret (Yager 2004b), Dempster-Shafer theory of evidence (Yager 1992; Merigó and Casanovas 2009), analytic hierarchy process (Yager and Kelman 1999), neural networks (Yager 1994) and game theory (Yager 1999).

Focussing on business and economic decision-making, we see that the OWA operators, combined with one or more of the previous methods can be applied in a lot of situations. For example, we could use the OWA operator in business decision-making problems such as financial management, strategic management, human resource management and product management. Inside these business areas, we could use the OWA operator in different ways depending on the particular problem we are analyzing as mentioned in the previous paragraph. For example, in human resource management, we could be looking for a selection process between directors, mid-range jobs, low-range jobs, in public administration, in sports and so on.

When using the OWA operator in economics, we could relate it with political decision-making because they are very much connected. For example, when looking for general economic decisions, these ones have a strong impact in political decision-

making. For example, the economic decisions about the selection of monetary policies, fiscal policies and commercial policies involve both the economic and the political sector. Other economic decisions that could be considered are those that affect the public sector such as decisions from the ministries, decisions from the autonomic authorities and decisions from the local authorities.

Obviously, both business and economic decision-making are very much related and the situations mentioned above could be seen as general framework inside business and economics.

The OWA operator is also useful in a lot of other situations that are not directly related with decision-making. Basically, the OWA operator is very useful in those situations where it can be seen as a statistical technique representing a new type of weighted average. Thus, a lot of business and decisions problems that use some kind of weighted average can be reformulated using the OWA operator. For example, the OWA operator is very useful in statistics and econometrics. Thus, a lot of problems that use the weighted average could be revised including linear regression, multiple regressions and a lot of its extensions and applications. Thus, we see that the OWA operator can be used in a lot of business and economic environments that uses statistical techniques such as business economics, marketing, finance, management science, actuarial science, insurance, behavioural economics, macroeconomics, microeconomics, economic policy, applied economics, accounting, public economics, entrepreneurship, social choice and welfare, economic development, industrial organization, tourism management and sport management.

5. Multi-person decision-making in production management

In the following, we are going to consider a business multi-person decision-making problem in production management. The motivation for using the OWA operator in the selection of production strategies in all different kinds of areas, appears because the decision-maker wants to take the decision with a certain degree of optimism or pessimism rather than with a neutral position. Due to the fact that the traditional methods are neutral against the attitude of the decision-maker, the introduction of the OWA operator in these models can change the neutrality and reflect decisions with different degrees of optimism and pessimism. These techniques can be used in a lot of situations but the general ideas about it is the possibility of under estimate or over estimate the problems in order to get results that reflects this change in the evaluation phase.

The process to follow in the selection of production strategies with the OWA operator, is similar to the process developed in Gil-Aluja (1998) and Kaufmann and Gil-Aluja (1986; 1987) for the selection of human resources with the difference that the instruments used will include the OWA operator in the selection process. Note that similar models that use the OWA operator have been developed for other selection processes (Merigó and A.M. Gil-Lafuente 2010). The five steps to follow are:

Step 1: Analysis and determination of the significant characteristics of the available production strategies. Let $A = \{A_1, A_2, \ldots, A_m\}$ be a set of finite alternatives, and $C = \{C_1, C_2, \ldots, C_n\}$, a set of finite characteristics (or attributes), forming the matrix $(x_{hi})_{m\times n}$. Let $E = \{E_1, E_2, \ldots, E_p\}$ be a finite set of decision-makers. Let $V = (v_1, v_2, \ldots, v_p)$ be the weighting vector of the weighted average such that $\sum_{k=1}^{p} v_k = 1$ and $v_k \in [0,1]$ and $U = (u_1, u_2, \ldots, u_p)$ be the weighting vector of the decision-makers that $\sum_{k=1}^{p} u_k = 1$ and $u_k \in [0,1]$. Each decision-maker provides their own payoff matrix $(x_{hi}^{(k)})_{m\times n}$.

Step 2: Fixation of the ideal levels of each significant characteristic in order to form the ideal production strategy. That is:

Table 2: *Ideal production strategy.*

	C_1	C_2	$\cdots$	C_i	$\cdot$	C_n
$P =$	x_1	x_2	$\cdots$	x_i	$\cdots$	x_n

where P is the ideal production strategy represented by a fuzzy subset, C_i is the ith characteristic to consider and $x_i \in [0,1]$; $i = 1, 2, \ldots, n$, is the valuation between 0 and 1 for the ith characteristic. Note that we assume that the ideal investment is given as a consensus between the opinions of the experts.

Step 3: Use the weighted average (WA) to aggregate the information of the decision-makers E by using the weighting vector U. The result is the collective payoff matrix $(x_{hi})_{m\times n}$. Thus, $x_{hi} = \sum_{k=1}^{p} u_k \left(x_{hi}^{(k)}\right)$.

Step 4: Comparison between the ideal production strategy and the different production strategies considered, and determination of the level of removal using the OWA operator. That is, changing the neutrality of the results to over estimate or under estimate them. In this step, the objective is to express numerically the removal between the ideal production strategy and the different production strategies considered. For this, it can be used the different available selection indexes such as those explained in the previous sections including the Hamming distance, the adequacy coefficient and the index of maximum and minimum level.

Step 5: Adoption of decisions according to the results found in the previous steps. Finally, we should take the decision about which production strategy select. Obviously, our decision will consist in choosing the production strategy with the best results according to the index used.

Note that when developing this decision process, we can summarize all the calculations in an aggregation process. We can do this with the IWOWAD obtaining the multi-person IWOWAD (MP-IWOWAD), with the IWOWAAC forming the multi-person

IWOWAAC (MP-IWOWAAC) and with the IWOWAIMAM obtaining the multi-person IWOWAIMAM (MP-IWOWAIMAM). They can be defined as follows.

Definition 13 *A MP-IWOWAD operator is an aggregation operator that has a weighting vector U of dimension p with* $\sum_{k=1}^{p} u_k = 1$ *and* $u_k \in [0,1]$*, and a weighting vector W of dimension n with* $\sum_{j=1}^{n} w_j = 1$ *and* $w_j \in [0,1]$*, such that:*

$$\text{MP-IWOWAD}(\langle x_1^1,\ldots,x_1^p\rangle, y_1\rangle,\ldots,\langle (x_n^1,\ldots,x_n^p), y_n\rangle) = \sum_{j=1}^{n} \hat{v}_j b_j, \tag{14}$$

where b_j *is the* $|x_i - y_i|$ *largest individual distance, each* $|x_i - y_i|$ *has associated a weight* v_i*,* v_j *is the associated weighted average (WA) of* b_j*,* $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$*,* $x_i = \sum_{k=1}^{p} u_k x_i^k$ *and* x_i^k *is the argument variable provided by each person.*

Definition 14 *A MP-IWOWAAC operator is an aggregation operator that has a weighting vector U of dimension p with* $\sum_{k=1}^{p} u_k = 1$ *and* $u_k \in [0,1]$*, and a weighting vector W of dimension n with* $\sum_{j=1}^{n} w_j = 1$ *and* $w_j \in [0,1]$*, such that:*

$$\text{MP-IWOWAAC}(\langle x_1^1,\ldots,x_1^p\rangle, y_1\rangle,\ldots,\langle (x_n^1,\ldots,x_n^p), y_n\rangle) = \sum_{j=1}^{n} \hat{v}_j b_j, \tag{15}$$

where b_j*is the jth largest of the* $[1 \wedge (1 - x_i + y_i)]$*,* x_i*,* $y_i \in [0,1]$*, each* $[1 \wedge (1 - x_i + y_i)]$ *has associated a weight* v_i*,* v_j *is the associated weighted average (WA) of* b_j*,* $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$*,* $x_i = \sum_{k=1}^{p} u_k x_i^k$ *and* x_i^k *is the argument variable provided by each person.*

Definition 15 *A MP-IWOWAIMAM operator is an aggregation operator that has a weighting vector U of dimension p with* $\sum_{k=1}^{p} u_k = 1$ *and* $u_k \in [0,1]$*, and a weighting vector W of dimension n with* $\sum_{j=1}^{n} w_j = 1$ *and* $w_j \in [0,1]$*, such that:*

$$\text{MP-IWOWAIMAM}(\langle (x_1^1,\ldots,x_1^p), y_1\rangle,\ldots,\langle (x_n^1,\ldots,x_n^p), y_n\rangle) = \sum_{j=1}^{n} \hat{v}_j b_j, \tag{16}$$

where b_j *is the jth largest of all the* $|x_i - y_i|$ *and the* $[0 \vee (x_i - y_i)]$*;* x_i*,* $y_i \in [0,1]$*, each* $|x_i - y_i|$ *and* $[0 \vee (x_i - y_i)]$ *has associated a weight* v_i*,* v_j *is the associated weighted average (WA) of* b_j*,* $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$*,* $x_i = \sum_{k=1}^{p} u_k x_i^k$ *and* x_i^k *is the argument variable provided by each person.*

The MP-IWOWAD, MP-IWOWAAC and MP-IWOWAIMAM have similar properties than those commented in Section 3. Thus, we can consider a wide range of extensions such as those that use generalized and quasi-arithmetic means obtaining the MP-GIWOWAD, the MP-GIWOWAAC and the MP-GIWOWAIMAM operators.

Furthermore, it is possible to consider a wide range of particular cases. For example, with the MP-IWOWAD we can consider the multi-person OWAD (MP-OWAD), the multi-person weighted Hamming distance (MP-WHD), the multi-person normalized Hamming distance (MP-NHD) and so on.

6. Illustrative example

In this Section, we present a simple numerical example where it is possible to see the applicability of the OWA operator in a business decision-making problem about selection of production strategies. Note that this example can be seen as a real world example although in this paper we do not use information from the real world.

Step 1: Assume an enterprise that produces cars is looking for its general strategy the next year and they consider that it should be useful for them to create a new production plant in order to be bigger and more competitive in the market. After careful evaluation of the information, the group of experts of the company constituted by three persons considers the following countries where it could be interesting to create a new production plant.

- $A_1 =$ Produce in Russia.
- $A_2 =$ Produce in China.
- $A_3 =$ Produce in India.
- $A_4 =$ Produce in Brazil.
- $A_5 =$ Produce in Nigeria.

The economic evaluation of producing in these countries can be described considering the following characteristics $C = (C_1 =$ Benefits in the short term, $C_2 =$ Benefits in the mid term, $C_3 =$ Benefits in the long term, $C_4 =$ Risk of the strategy, $C_5 =$ Subjective opinion of the group of experts, $C_6 =$ Other variables).

Step 2: With this information, the group of experts of the company establishes the ideal results that the ideal production strategy should have. These results are represented in Table 3.

Table 3: *Ideal production strategy.*

	C_1	C_2	C_3	C_4	C_5	C_6
$P =$	0.8	0.9	1	0.9	0.9	0.8

Step 3: Fixation of the real level of each characteristic for all the different production strategies considered. For each of these characteristics, the following information is

given by each expert shown in Tables 4, 5 and 6. Note that we assume that each expert has the same degree of importance. That is: $U = (1/3, 1/3, 1/3)$.

Table 4: *Available production strategies-Expert 1.*

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	0.5	0.4	0.5	0.3	0.7	0.4
A_2	0.1	0.7	0.7	0.3	0.7	0.9
A_3	0.6	0.8	0.5	0.1	0.6	0.3
A_4	0.5	0.4	0.6	0.6	0.7	0.6
A_5	0.2	0.1	0.4	0.8	0.7	0.1

Table 5: *Available production strategies-Expert 2.*

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	0.6	0.8	0.7	0.4	0.8	0.7
A_2	0.1	0.9	0.9	0.5	0.7	0.7
A_3	0.8	1	0.7	0.3	0.8	0.6
A_4	0.4	0.4	0.6	0.9	0.5	0.9
A_5	0.6	0.2	0.8	0.9	0.7	0.1

Table 6: *Available production strategies-Expert 3.*

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	0.7	0.9	0.6	0.5	0.9	0.7
A_2	0.4	0.8	0.8	0.1	0.7	0.8
A_3	0.7	0.9	0.6	0.2	0.7	0.6
A_4	0.3	0.4	0.6	0.9	0.6	0.6
A_5	0.7	0	0.6	0.7	0.7	0.1

With this information, we can aggregate the information of the three experts in order to obtain a collective result of the available production strategies. The results are presented in Table 7.

Table 7: *Available production strategies-Collective results.*

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	0.6	0.7	0.6	0.4	0.8	0.6
A_2	0.2	0.8	0.8	0.3	0.7	0.8
A_3	0.7	0.9	0.6	0.2	0.7	0.5
A_4	0.4	0.4	0.6	0.8	0.6	0.7
A_5	0.5	0.1	0.6	0.8	0.7	0.1

Step 4: Comparison between the ideal production strategy and the different production strategies considered, and determination of the level of removal using the OWA operator. By using the Hamming distance, we will consider the normalized Hamming distance, the weighted Hamming distance, the OWAD, the AOWAD the median-OWAD

and the IWOWAD operator. In this example, we assume that the company decides to use the following weighting vectors: $W = (0.1, 0.1, 0.1, 0.2, 0.2, 0.3)$ and $V = (0.3, 0.2, 0.2, 0.1, 0.1, 0.1)$. Note that in the literature we have a wide range of methods for determining the weights (Merigó, 2010; Merigó and Gil-Lafuente, 2009; 2010; Yager, 1993). Thus, when using immediate weights for the IWOWAD, IWOWAAC and IWOWAIMAM, we use the following weights obtained by using Eq. (9), (10) and (11), shown in Table 8.

Table 8: *Immediate weights.*

A_1	0.066	0.133	0.2	0.266	0.133	0.2
A_2	0.1875	0.0625	0.0625	0.25	0.25	0.1875
A_3	0.055	0.111	0.055	0.111	0.333	0.333
A_4	0.1428	0.2142	0.1428	0.1428	0.1428	0.2142
A_5	0.125	0.0625	0.125	0.375	0.125	0.1875

Note that we have to calculate the individual distances of each characteristic to the ideal value of the corresponding characteristic forming the fuzzy subset of individual distances for each strategy. Once, we have the individual distances, we aggregate them with the appropriate aggregation operator. The results are shown in Table 9.

Table 9: *Aggregated results with the OWAD operator.*

	NHD	*WHD*	*OWAD*	*AOWAD*	*Median*	*IWOWAD*
A_1	0.266	0.26	0.22	0.32	0.2	0.2266
A_2	0.283	0.32	0.2	0.37	0.2	0.2375
A_3	0.283	0.23	0.2	0.38	0.25	0.1555
A_4	0.3	0.35	0.24	0.36	0.35	0.2927
A_5	0.316	0.43	0.24	0.42	0.25	0.35

If we develop the selection process with the adequacy coefficient, we will get the following. First, we have to calculate how close the characteristics are to the ideal production strategy. Once we have calculated all the different individual values, we will construct the aggregation. In this case, the arguments will be ordered using Eq. (6) and Eq. (12). The results are shown in Table 10.

Table 10: *Aggregated results with the OWAAC operator.*

	NAC	*WAC*	*OWAAC*	*AOWAAC*	*Median*	*IWOWAAC*
A_1	0.733	0.74	0.68	0.78	0.8	0.7734
A_2	0.716	0.68	0.63	0.8	0.6	0.7625
A_3	0.716	0.77	0.62	0.8	0.75	0.8445
A_4	0.7	0.65	0.64	0.76	0.65	0.7072
A_5	0.683	0.57	0.58	0.76	0.75	0.65

Finally, if we use the index of maximum and minimum level in the selection process as a combination of the normalized Hamming distance and the normalized adequacy

coefficient, we get the following. In this example, we assume that the characteristics C_1 and C_2 have to be treated with the adequacy coefficient and the other four characteristics have to be treated with the Hamming distance. Its resolution consists in the following. First, we calculate the individual removal of each characteristic to the ideal, independently that the instrument used is the Hamming distance or the adequacy index. Once calculated all the values for the individual removal, we construct the aggregation using Eq. (7) and Eq. (13). Here, we note that in the reordering step, it will be only considered the individual value obtained for each characteristic, independently that the value has been obtained with the adequacy coefficient or with the Hamming distance. The results are shown in Table 11.

Table 11: *Aggregated results with the OWAIMAM operator.*

	NIMAM	*WIMAM*	*OWAIMAM*	*AOWAIMAM*	*Median*	*IWOWAIMAM*
A_1	0.266	0.26	0.22	0.42	0.2	0.2266
A_2	0.283	0.32	0.2	0.37	0.2	0.2375
A_3	0.283	0.23	0.2	0.38	0.25	0.1555
A_4	0.3	0.35	0.24	0.36	0.35	0.2927
A_5	0.316	0.43	0.24	0.42	0.25	0.35

In order to see the optimal production strategies depending on the particular types of OWA aggregations used, we establish the following table with the ordering of the production strategies. Note that this is very useful when the decision-maker wants to consider more than one alternative. The results are shown in Table 12.

Table 12: *Ordering of the production strategies.*

	Ordering		Ordering
NHD	$A_1\}A_2 = A_3\}A_4\}A_5$	AOWAAC	$A_2 = A_3\}A_1\}A_4 = A_5$
WHD	$A_3\}A_1\}A_2\}A_4\}A_5$	Median	$A_1\}A_3 = A_5\}A_4\}A_2$
OWAD	$A_2 = A_3\}A_1\}A_4 = A_5$	IWOWAAC	$A_3\}A_1\}A_2\}A_4\}A_5$
AOWAD	$A_1\}A_4\}A_2\}A_3\}A_5$	NIMAM	$A_1\}A_2 = A_3\}A_4\}A_5$
Median	$A_1 = A_2\}A_3 = A_5\}A_4$	WIMAM	$A_3\}A_1\}A_2\}A_4\}A_5$
IWOWAD	$A_3\}A_1\}A_2\}A_4\}A_5$	OWAIMAM	$A_2 = A_3\}A_1\}A_4 = A_5$
NAC	$A_1\}A_2 = A_3\}A_4\}A_5$	AOWAIMAM	$A_1\}A_4\}A_2\}A_3\}A_5$
WAC	$A_3\}A_1\}A_2\}A_4\}A_5$	Median	$A_1 = A_2\}A_3 = A_5\}A_4$
OWAAC	$A_1\}A_4\}A_2\}A_3\}A_5$	IWOWAIMAM	$A_3\}A_1\}A_2\}A_4\}A_5$

As we can see, we get different orderings depending on the aggregation operator used. The main advantage of this analysis is that the company gets a more complete view of the different scenarios that could happen in the future depending on the method used. Although it will select the alternative that it is in accordance with its interests, it will be concerned on other potential results that could happen in the uncertain environment. Note that in this specific problem, we see that A_1 or A_3 seems to be the optimal choices.

7. Conclusions

We have studied the usefulness of the OWA operator in business and economics. For doing so, we have given special attention to business and economic decision-making problems. We have used some practical decision-making techniques that use similarity measures in the decision-making process such as the Hamming distance, the adequacy coefficient and the index of maximum and minimum level. We have reviewed the use of the OWA operator in these techniques, obtaining the OWAD operator, the OWAAC operator and the OWAIMAM operator. We have seen that these aggregation operators are very useful for decision-making because they permit to under or over estimate the results according to the attitudinal character of the decision-maker in the particular problem considered.

We have suggested new techniques by using immediate weights. That is, by using a framework that is able to deal with the weighted average and the OWA operator in the same formulation. We have presented the IWOWAD, the IWOWAAC and the IWOWAIMAM operator. Furthermore, we have generalized them by using generalized aggregation operators obtaining the GIWOWAD, the GIWOWAAC and the GIWOWAIMAM operator. The main advantage of these measures is that they include a wide range of particular cases that can be used in the aggregation process depending on the particular interests in analysis.

We have also seen that the OWA operator can be also used in a lot of other problems in business and economics, especially when we see it as a statistical (or aggregation) technique similar to the weighted average. We have mentioned different potential areas where we could use it and we have seen that the applicability is very broad because we can implement it in a lot of business problems such as in finance, marketing, production and tourism. We have also presented different applications in economics and we have seen that it has a strong connection with politics because national economic decisions are usually related with political ones.

In this paper, we have focussed on a business multi-person decision-making application concerning production management by using the IWOWAD, the IWOWAAC and the IWOWAIMAM operator. Thus, we have obtained the MP-IWOWAD, the MP-IWOWAAC and the MP-IWOWAIMAM operators. We have seen that they are very practical because we can assess the information of several persons (experts) in an efficient way. We have analysed a company that it is planning its production strategy for the next year. We have seen that depending on the particular type of aggregation operator used, the results may lead to different decisions.

Acknowledgements

We would like to thank the anonymous referees for valuable comments that have improved the quality of the paper. This work is partly supported by the Spanish Ministry

of Foreign Affairs and Cooperation, "*Agencia Española de Cooperación Internacional para el Desarrollo*" (AECID) (project A/016239/08) and the Spanish Ministry of Education (Project JC2009-00189).

References

Beliakov, G., Calvo, T. and Pradera, A. (2007). *Aggregation Functions: A Guide for Practitioners*. Springer, Berlin.

Canós, L. and Liern, V. (2008). Soft computing-based aggregation methods for human resource management. *European Journal of Operational Research*, 189: 669–681.

Engemann, K. J., Filev, D. P. and Yager, R. R. (1996). Modeling decision-making using immediate probabilities. *International Journal of General Systems*, 24: 281–294.

Figueira, J., Greco, S. and Ehrgott, M. (2005). *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer, Boston.

Gil-Aluja, J. (1998). *The Interactive Management of Human Resources in Uncertainty*. Kluwer Academic Publishers, Dordrecht.

Gil-Lafuente, A. M. (2005). *Fuzzy Logic in Financial Analysis*. Springer, Berlin.

Gil-Lafuente, A. M. and Merigó, J. M. (2010). *Computational Intelligence in Business and Economics*. World Scientific, Singapore.

Gil-Lafuente, J. (2001). The index of maximum and minimum level in the selection of players in sport management (in Spanish). In: *Proceedings of the 10th International Conference of the European Academy of Management and Business Economics* (AEDEM), Reggio Calabria, Italy, pp. 439–443.

Gil-Lafuente, J. (2002). *Algorithms for Excelence. Keys for Being Succesful in Sport Management* (in Spanish). Ed. Milladoiro, Vigo.

Hamming, R. W. (1950). Error-detecting and error-correcting codes. *Bell Systems Technical Journal*, 29: 147–160.

Karayiannis, N. (2000). Soft learning vector quantization and clustering algorithms based on ordered weighted aggregation operators. *IEEE Transactions on Neural Networks*, 11: 1093–1105.

Kaufmann, A. (1975). *Introduction to the Theory of Fuzzy Subsets*. Academic Press, New York.

Kaufmann, A. and Gil-Aluja, J. (1986). *Introduction to the theory of fuzzy subsets in business management* (In Spanish). Ed. Milladoiro, Santiago de Compostela.

Kaufmann, A. and Gil-Aluja, J. (1987). *Business Techniques for the Treatment of the Uncertainty* (in Spanish). Ed. Hispano-europea, Barcelona.

Merigó, J. M. (2008). *New extensions to the OWA operator and its application in decision-making*. PhD Thesis (in Spanish). Department of Business Administration, University of Barcelona.

Merigó, J. M. (2010). Fuzzy decision-making with immediate probabilities. *Computers & Industrial Engineering*, 58: 651–657.

Merigó, J. M. and Casanovas, M. (2009). Induced aggregation operators in decision-making with Dempster-Shafer belief structure. *International Journal of Intelligent Systems*, 24: 934–954.

Merigó, J. M. and Casanovas, M. (2010a). Decision-making with distance measures and linguistic aggregation operators. *International Journal of Fuzzy Systems*, 12: 190–198.

Merigó, J. M. and Casanovas, M. (2010b). The fuzzy generalized OWA operator and its application in strategic decision-making. *Cybernetics & Systems*, 41: 359–370.

Merigó, J. M. and Casanovas, M. (2010c). Induced and heavy aggregation operators with distance measures. *Journal of Systems Engineering and Electronics*, 21: 431–439.

Merigó, J. M. and Casanovas, M. (2011a). Decision-making with distance measures and induced aggregation operators. *Computers & Industrial Engineering*, 60: 66–76.

Merigó, J. M. and Casanovas, M. (2011b). Induced aggregation operators in the Euclidean distance and its application in financial decision-making. *Expert Systems with Applications*, 38: 7603–7608.

Merigó, J. M. and Casanovas, M. (2011c). A new Minkowski distance based on induced aggregation operators, *International Journal of Computational Intelligence Systems*, 4: 123–133.

Merigó, J. M. and Casanovas, M. (2011d). The uncertain induced quasi-arithmetic OWA operator. *International Journal of Intelligent Systems*, 26: 1–24.

Merigó, J. M., Casanovas, M. and Martínez, L. (2010). Linguistic aggregation operators for linguistic decision-making based on the Dempster-Shafer theory of evidence. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 18: 287–304.

Merigó, J. M. and Gil-Lafuente, A. M. (2007). The ordered weighted averaging distance operator. *Lectures on Modelling and Simulation*, 8: 1–11.

Merigó, J. M. and Gil-Lafuente, A. M. (2009). The induced generalized OWA operator. *Information Sciences*, 179: 729–741.

Merigó, J. M. and Gil-Lafuente, A. M. (2010). Decision-making techniques for the selection of financial products. *Information Sciences*, 180: 2085–2094.

Merigó, J. M. and Gil-Lafuente, A. M. (2011). OWA operators in human resource management, *Economic Computation and Economic Cybernetics Studies and Research*, 45: 153–168.

Merigó, J. M., Gil-Lafuente, A. M. and Gil-Aluja, J. (2011a). Decision-making with the induced generalized adequacy coefficient, *Applied and Computational Mathematics*, 2: 321–339.

Merigó, J. M., Gil-Lafuente, A. M. and Gil-Aluja, J. (2011b). A new aggregation method for strategic decision-making and its application in assignment theory, *African Journal of Business Management*, 5: 4033–4043.

Torra, V. and Narukawa, Y. (2007). *Modelling Decisions: Information Fusion and Aggregation Operators*. Springer, Berlin.

Wei, G. W. (2009). Uncertain linguistic hybrid geometric mean operator and its application to group decision-making under uncertain linguistic environment. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 17: 251–267.

Wei, G. W., Wang, H. J. and Lin, R. (2011). Application of correlation coefficient to interval-valued intuitionistic fuzzy multiple attribute decision-making with incomplete weight information. *Knowledge and Information Systems*, 26: 337–349.

Wei, G. W., Zhao, X. and Lin, R. (2010). Some induced aggregating operators with fuzzy number intuitionistic fuzzy information and their applications to group decision-making. *International Journal of Computational Intelligence Systems*, 3: 84–95.

Xu, Z. S. (2005). An overview of methods for determining OWA weights. *International Journal of Intelligent Systems*, 20: 843–865.

Xu, Z. S. (2010a). A method based on distance measure for interval-valued intuitionistic fuzzy group decision-making. *Information Sciences*, 180: 181–190.

Xu, Z. S. (2010b). A deviation-based approach to intuitionistic fuzzy multiple attribute group decision-making. *Group Decision and Negotiation*, 19: 57–76.

Xu, Z. S. and Chen, J. (2008). Ordered weighted distance measure. *Journal of Systems Science and Systems Engineering*, 17: 432–445.

Xu, Z. S. and Da, Q. L. (2003). An overview of operators for aggregating the information. *International Journal of Intelligent Systems*, 18: 953–969.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multi-criteria decision-making. *IEEE Transactions on Systems, Man and Cybernetics*, B 18: 183–190.

Yager, R. R. (1992). Decision-making under Dempster-Shafer uncertainties. *International Journal of General Systems*, 20: 233–245.

Yager, R. R. (1993). Families of OWA operators. *Fuzzy Sets and Systems*, 59: 125–148.

Yager, R. R. (1994). On the RAGE aggregation method with applications to neural networks and decision-making. *International Journal of Approximate Reasoning*, 11: 175–204.

Yager, R. R. (1999). A game theoretic approach to decision-making under uncertainty. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 8: 131–143.

Yager, R. R. (2004a). Generalized OWA aggregation operators. *Fuzzy Optimization and Decision-Making*, 3: 93–107.

Yager, R. R. (2004b). Decision-making using minimization of regret. *International Journal of Approximate Reasoning*, 36: 109–128.

Yager, R. R. (2010). Norms induced from OWA operators. *IEEE Transactions on Fuzzy Systems*, 18: 57–66.

Yager, R. R., Engemman, K. J. and Filev, D. P. (1995). On the concept of immediate probabilities. *International Journal of Intelligent Systems*, 10: 373–397.

Yager, R. R. and Kacprzyk, J. (1997). *The Ordered Weighted Averaging Operators: Theory and Applications*. Kluwer Academic Publishers, Norwell, MA.

Yager, R. R., Kacprzyk, J. and Beliakov, G. (2011). *Recent Developments on the Ordered Weighted Averaging Operators: Theory and Practice*, Springer-Verlag, Berlin.

Yager, R. R. and Kelman, A. (1999). An extension of the analytical hierarchy process using OWA operators. *Journal of Intelligent and Fuzzy Systems*, 7: 401–417.

Zhao, H., Xu, Z. S., Ni, M. and Liu, S. (2010). Generalized aggregation operators for intuitionistic fuzzy sets. *International Journal of Intelligent Systems*, 25: 1–30.

Zhou, L. G. and Chen, H. Y. (2010). Generalized ordered weighted logarithm aggregation operators and their applications to group decision-making. *International Journal of Intelligent Systems*, 25: 683–707.

SORT 36 (1) January-June 2012, 103-114

Introducing migratory flows in life table construction

Jose M. Pavía[1], Francisco Morillas[2] and Josep Lledó[3]

Abstract

The purpose of life tables is to describe the mortality behaviour of particular groups. The construction of general life tables is based on death statistics and census figures of resident populations under the hypothesis of closed demographic system. Among other assumptions, this hypothesis implicitly assumes that entries (immigrants) and exits (emigrants) of the population are usually not significant (being almost of the same magnitude for each age compensating each other). This paper theoretically extends the classical solution to open demographic systems and studies the impact of this hypothesis in constructing a life table. In particular, using the data of residential variations made available to the public by the Spanish National Statistical Office (INE, Instituto Nacional de Estadística) to approximate migratory flows, we introduce in the process of constructing a life table these flows and compare, before and after graduation, the crude mortality rates and the adjusted death probabilities obtained when migratory flows are, and are not, taken into account.

MSC: 91D20, 62Q05

Keywords: General population, Lexis diagram, open demographic system.

1. Introduction

In the demographic and actuarial fields, the analysis of mortality in a population has particular relevance for their applications. Life tables, or mortality tables, are used to recreate an observed mortality situation or to present future values of the evolution of

[1] Department of Applied Economics, Facultat d'Economía, Universitat de Valencia, Campus Els Tarongers, 46022-Valencia (Spain), e-mail:pavia@uv.es

[2] Department of Applied Economics, Facultat d'Economía, Universitat de Valencia, Campus Els Tarongers, 46022-Valencia (Spain), e-mail:francisco.morillas@uv.es

[3] Actuarial Services Consultant, Ernst & Young, Torre Picasso, 1, 28020-Madrid (Spain), e-mail:jonabe@alumni.uv.es

Received: November 2011
Accepted: February 2012

mortality in certain groups, making it possible to generate demographic forecasts or to calculate premiums and/or income for life insurance and pension benefits. Medicine is another area where mortality analysis is also frequently used.

Life tables are usually drawn from the study and analysis of the intensity and rate at which mortality affects each age group in question. In general populations it is generated using information mainly from population censuses and lists of deceased, where individual records from insurance policies are the prime source of information in insured populations. To be specific, and once the sample period has been decided, the comparison between the numbers at risk and the number of deaths allows the actuary (demographer) to obtain initial (crude) estimates for the probability of death in each age group q_x. These probabilities are subjected to the corresponding graduation or adjustment processes (see, for example, Copers-Haberman, 1983; Forfar et al. 1988: or Ayuso et al., 2007) with a view to smoothing the profile of the associated stochastic process and to ultimately develop appropriate tables from a fictitious starting population of size ℓ_0.

In the construction of mortality tables for general populations (which is the subject of this paper) it is not usual to specifically consider migratory flows, making the hypothesis of closed demographic system (HCDS), which implicitly entails the assumption of certain limitations, the main ones being: (i) that migration flows (inputs and outputs) of the population by age and sex are considered not to be significant; (ii) that for each age group migration flows are random and show similar entry and exit figures; and, (iii) that immigrants acquire the same risk of death as the resident population.

These limitations, which are not always reasonable, should be checked because of their potential impact on, for example, the calculation of life expectancy, of premiums for life insurance or of estimates for the calculation of pensions. The aim of this paper is twofold, firstly, to introduce an estimator for an open demographic system and, secondly, to show the incidence of HCDS through a real case. More specifically, given the immense pressure of migration endured by Europe, and in particular Spain, in recent years, the analysis will be based on the comparison of mortality tables (by gender) obtained for Spain under HCDS and under the hypothesis of open demographic system (HODS), in which migratory flows are explicitly considered. The comparison is carried out in two ways. Firstly, the differences between the estimated crude probabilities obtained under each hypothesis are compared and, secondly, the comparison is again carried out after having graduated the crude data.

The rest of the paper is structured as follows. Section 2 explains the methodology used to obtain the mortality tables, both for the closed demographic system and the open demographic system. In Section 3, different comparisons are carried out between the HCDS and HODS tables. The last section presents the conclusions reached and indicates several issues for future research.

2. Methodology

The techniques and formulae used to estimate life tables are heavily influenced by the type of information available. When working with official statistics, the relevant data are usually offered, by and large, in aggregate form. Hence, in this research we have opted to work with aggregated figures – which come from information that the Spanish National Statistical Office (INE) has made available on its website (http://www.ine.es) – despite it being possible in the Spanish case to use some detailed (anonymized) microdata[1]. In particular, we will consider that, for each gender, aggregated figures of migrants, deaths and population are available by age and calendar year. Under these circumstances, the representation and analysis of the information in a Lexis diagram (named after the German statistician, economist and social scientist Wihelm Lexis, who adopted it in the nineteenth century to illustrate the procedures for calculating a mortality table) often greatly facilitates the reasoning and makes more manageable, after the application of a number of reasonable hypotheses, handling the flaws of detailed information which are present in aggregate data.[2]

The Lexis diagram is a two-dimensional diagram of lifelines with two-temporal dimension: calendar time and age. Calendar time is represented on the horizontal axis, while age is represented on the vertical axis. This diagram permits representation of the life events of a population from the personal history of the individuals within it. Each personal story is represented by a line segment forming an angle of 45 degrees to the horizontal axis. The classical approach (of closed demographic system) states that each personal story begins at birth, which is represented on the baseline, and ends at some point on the graph with the individual's death (Livi Bacci, 2000). In this paper, however, personal history is not determined solely by the events of birth and death. The introduction of migratory flows makes it necessary to modify the classical interpretation of the Lexis diagram since, in this case, the history of an individual could start from birth or from immigration and, likewise, it could end by death or by emigration.

Figure 1 shows a small section of a Lexis diagram which, as usual, comes divided into cells of dimensions of 1×1 so that between each pair of oblique lines are the lifelines that make up a generation of individuals and where each cell represents an observation period of one year (in which the age of the individuals also has a variation of a year).

For example, in Figure 1 (left) various individual life lines (thin lines) have been represented. Lifelines of individuals entering in the system due to immigration can originate anywhere in the diagram and are differentiated from the rest by having a $\circ$ at its origin. When an individual leaves the system, the cause of departure is differentiated graphically: if the reason is death, the tail end of the lifeline is marked with an $\times$; if the reason is emigration the end-lifeline is marked with a $\square$.

1. Such is the case of death statistics and of residential variation figures (directly available on the INE website). Indeed, the INE, unlike many other national and international statistical agencies, is characterized by its willingness to make available, in anomymized form, the detailed data generating the vast majority of its statistical operations.

2. Obviously, some of the hypotheses to be considered could be relaxed using more detailed data (see, for example, INE, 2009).

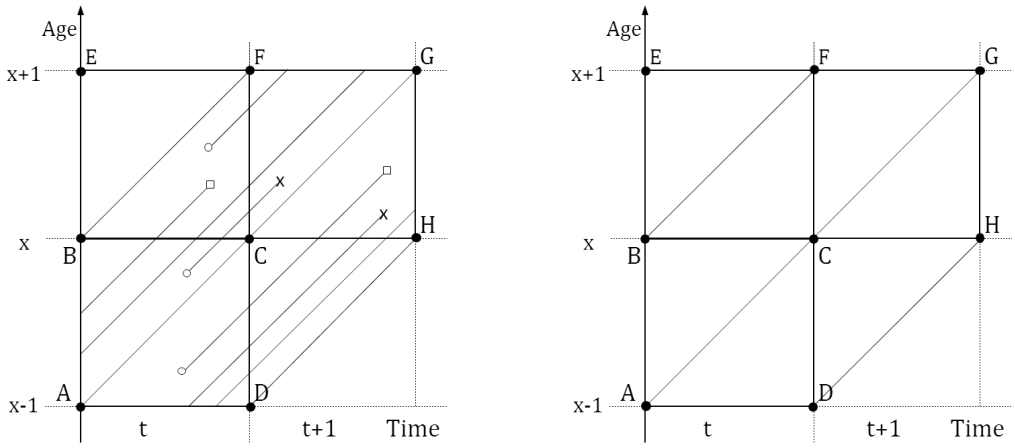


Figure 1: *Detail* (2×2) *of Lexis diagram with lifelines (left) and schematic (right).*

The aggregate information, however, does not allow for an accurate location of the lifelines of the individuals who comprise the study population. To make use of this geometric representation, therefore, the usual convention of the Lexis scheme of assigning values to segments and surfaces must be adopted. In this paper, the value of a segment is always identified by the number of lifelines that cross it, while the value(s) to be associated with each area will depend on the hypothesis under consideration.

2.1. Closed demographic system

Under HSDC, each surface is identified with a single value: the number of lifelines that end in it due to death. So, assuming for simplicity that (as in our case) for any given age x there are available the number of residents counted in January 1 of year t, C_x^t, and the number of residents who died in each age x for years t and $t+1$ (D_x^t and D_x^{t+1} respectively), we have that, as shown in Figure 1 (right), it is straightforward to draw the information. To be specific, on the one hand, the quadrilaterals $ABCD$, $BCEF$ and $CFGH$ will be identifiable respectively with D_{x-1}^t, D_x^t and D_x^{t+1}, and, on the other hand, the segment AB will be equal to C_{x-1}^t.

At this point, it is now easy to obtain an initial estimate of q_x exploiting the geometric properties of the representation and assuming uniform distribution of birth dates and deaths within each age group and year. In particular, noting that the segment BC represents the number of people reaching age x in year t, it follows that under HSDC the number of them who die before reaching age $x+1$ will come represented by the quadrilateral $BCFG$ and that therefore an estimate for q_x is obtained from:[3]

3. Given that $q_x = \frac{d_x}{\ell_x}$ is defined as the quotient of the number of deaths between ages x and $x+1$, d_x, and the number of survivors at age x, ℓ_x.

$$\widehat{q}_x = \frac{BCFG}{BC} \tag{1}$$

And from this, using the geometry of the scheme, one arrives at $BCFG = BCF + CFG$ and $BC = AB - ABC$; from which, using the hypothesis of uniform distribution, one deduces $BCFG = \frac{BCEF}{2} + \frac{CFGH}{2}$ and hence:[4,5,6]

$$\widehat{q}_x = \frac{\frac{1}{2}\left(D_x^t + D_x^{t+1}\right)}{C_{x-1}^t - \frac{1}{2}D_{x-1}^t} \tag{2}$$

2.2. Introducing migratory flows

Assuming HODS, the expression (2) becomes invalid and another approach is required to take into account the entries and exits that happen in the study group during the analysis period. At this point, it will be useful to refer to the type of reasoning usually employed for insured groups (see, for example, Benjamin and Pollard, 1992), where the number of deaths observed is separated depending on the risks of death and the time exposed to risk. That is, under HODS, the number of deaths observed, $BCFG$, will be approximately equal to the number of people that reach age x, BC, by the probability of any of them dying before reaching the age $x+1$, q_x, plus the number of people that immigrate with age $x+k$ (where $0 < k < 1$), whose lifeline starting point would be located in the surface $BCFG$, by the probability that a person of age $x+k$ dies before reaching age $x+1$, ${}_{1-k}q_x$,[7] minus the number of people that emigrate with age $x+k$ (where $0 < k < 1$), whose lifeline end point would be located in surface $BCFG$, by the probability that a person of age $x+k$ dies before reaching the age $x+1$, ${}_{1-k}q_x$.

4. This general expression, nevertheless, would not be appropriate for ages zero and one, since as it is well-known the deaths of children under one year old are concentrated in their first weeks of life. The assumption of uniform distribution cannot therefore be maintained for deaths counted with zero years: the greater part of these deaths will be located in the corresponding lower triangle. Thus, the formulae used for ages zero and one have been, respectively, $\widehat{q}_0 = \frac{0.7D_0^t + 0.3D_0^{t+1}}{B^t}$ (where B^t denotes the births in year t) and $\widehat{q}_1 = \frac{\frac{1}{2}(D_1^t + D_1^{t+1})}{C_0^t - 0.3D_0^t}$; which can be obtained assuming that the number of deaths occurring during the first half of age zero is approximately four times the number of deaths registered during the second half. Obviously, if the deaths by generation (also available on the INE website; INE, 2010a) were used, no hypothesis about how to distribute the deaths between the triangles would be necessary because of the values of these would be known exactly.

5. Unlike the expression used to estimate q_x in this paper, until recently the INE started with $BC = CF + BCF$ and arrived at a different equation, $\widehat{q}_x = \frac{\frac{1}{2}\left(D_x^t + D_x^{t+1}\right)}{C_x^{t+1} + \frac{1}{2}D_x^t}$, although equivalent under HCDS (INE, 2007). Since 2009, the INE estimates central age-specific death rates, m_x, using the detailed information available on death microdata to obtain in each age the exact time spent by those who die during the year of study (see, INE, 2009). The use of those detailed data makes it unnecessary to assume any hypotheses about the distribution of deaths within each age and calendar year.

6. A general formula to estimate q_x when the census of the population is located at any instant of the year and not necessarily at the start can be found in, for example, Pavía (2011, Ex. 91).

7. Note that using this expression entails the implicit assumption that immigrants acquire the same risk of death as the population in which they integrate.

The problem is that, unlike what happens in insured populations, the dates and specific ages at which a person immigrates or emigrates are not usually known, so to obtain a useful expression of the decomposition of $BCFG$ it is essential to extend the classic convention and assign additional variables to each surface of the Lexis diagram. To be specific, we propose to link to each area two new variables: the number of lifelines that start in the surface (immigrants) and the number of lifelines that finish in the surface for reasons other than death (emigrants). With this extension it will then be possible to obtain an operative expression for $BCFG$ from which an estimator for q_x can be derived by simply adding hypotheses (i) on the distribution of entries and exits in each surface (which, in the same way as death distribution, are assumed to be uniform, since it is reasonable for the level of information available) and (ii) on the risk of death throughout each age x (which as a rule is assumed proportional to the period of exposure to risk, that is, ${}_{1-k}q_x = (1-k)q_x$).

As usual when handling official statistics, it is assumed that the total number of people that immigrate and emigrate in any year t and with age x is only known in aggregate terms, I_x^t and E_x^t, respectively. Obviously, other more informative situations in which more precise data about the age distribution of migrants in each year were available, for example from microdata of residential variations, would also be perfectly treated with this strategy.

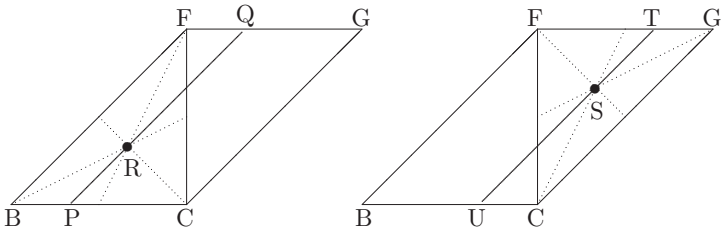


Figure 2: *Barycentres of surfaces of migratory movements.*

Under the conditions above, denoting by $N_x^t = I_x^t - E_x^t$ the net migration recorded in year t at age x, it follows that the number of people who reach age x during year t, BC, would be equal to $BC = C_{x-1}^t - \frac{1}{2}D_{x-1}^t + \frac{1}{2}N_{x-1}^t$, and that the number of entries and exits that would be registered in each of the triangles BCF and CFG would be, respectively, $\frac{I_x^t}{2}$, $\frac{E_x^t}{2}$, and $\frac{I_x^{t+1}}{2}$, $\frac{E_x^{t+1}}{2}$, and likewise, under the same hypotheses, each exit/entry produced in each triangle would be located, in average terms, in the centroid (barycentre) of the corresponding triangle (see Figure 2).[8]

8. Alternatively, it can be demonstrated that the average of the distances (across the lifelines) of each point of the corresponding triangle to segment FG is equal to the distance of the corresponding barycentre to segment FG. For example, considering the triangle BCF and, inside it, an arbitrary point K with coordinates (x,y), it is not difficult to prove that the lifeline of K intersects FG in a point, K', with coordinates $(1+x-y,1)$ —where B has been taken as the origin of the corresponding Cartesian coordinate system. Hence, the Euclidean distance from K to K' would be $(1-y)\sqrt{2}$, from which it follows that the *sum* of all the distances is $\int_0^1 \int_0^x \sqrt{2}(1-y)dydx =$

As can be seen in the representation on the left of Figure 2, the point R is in the barycentre or centroid of the triangle BCF, which is easy to prove to be at a distance of $\frac{2\sqrt{2}}{3}$ from point Q, in the same way that point S (which can be taken as representative of all points in which an entry/exit occurs in triangle CFG) is at a distance $\frac{\sqrt{2}}{3}$ from T.[9] From here, taking into account that the distances of P to Q and of U to T are equal to $\sqrt{2}$, equivalent to a year, it follows that on average the exposure to risk of each immigrant/emigrant would be, respectively, $\frac{2}{3}$ and $\frac{1}{3}$. Hence bearing in mind the previous arguments, we have that the number of deaths observed in the parallelogram $BCFG$, $\frac{1}{2}\left(D_x^t + D_x^{t+1}\right)$, could be broken down in the following way:

$$\frac{1}{2}\left(D_x^t + D_x^{t+1}\right) \approx \left(C_{x-1}^t - \frac{1}{2}D_{x-1}^t + \frac{1}{2}N_{x-1}^t\right) q_x + \frac{1}{2}N_x^t \frac{2}{3} q_x + \frac{1}{2}N_x^{t+1} \frac{1}{3} q_x$$

And consequently, an estimator for q_x, under HODS, would be obtained by way of the following expression:[10]

$$\widetilde{q}_x = \frac{\frac{1}{2}\left(D_x^t + D_x^{t+1}\right)}{C_{x-1}^t - \frac{1}{2}D_{x-1}^t + \frac{1}{2}N_{x-1}^t + \frac{1}{3}N_x^t + \frac{1}{6}N_x^{t+1}} \tag{3}$$

3. Comparative analysis

In order to analyse the impact of considering, or not, migration on the estimates of the probability of survival or death at each age x, we have constructed, using the deaths of two adjacent years, the life tables of the years 2006, 2007 and 2008 (for ages 0 to 99 years). The information handled comes from the data that INE offers directly to the public on its website. sex and age (January, 1) Population Now Cast (ePOBa) estimates (for years 2006, 2007 and 2008) have been used as population data (INE, 2012). Death

$\sqrt{2}\int_0^1 \left(x - \frac{x^2}{2}\right) dx = \sqrt{2}\left(\frac{1}{2} - \frac{1}{6}\right) = \frac{\sqrt{2}}{3}$, which coincides with the product of the area of the triangle BCF (the *number* of points in BCF), $\frac{1}{2}$, and the length of RQ, the segment of lifeline that goes from the barycentre of BCF to FG, $\frac{2\sqrt{2}}{3}$ (see next footnote).

9. Indeed, taking B as the origin of a Cartesian coordinate system and using that the Lexis squares have unit sides, we have that the coordinates of the points C, F and G are, respectively, $(1,0)$, $(1,1)$ and $(2,1)$ and that, as a consequence, the coordinates of the barycentres R and S are $(\frac{2}{3}, \frac{1}{3})$ and $(\frac{5}{3}, \frac{2}{3})$, respectively. Hence, it is not difficult to see that the distance from R to Q is equal to the length of the hypotenuse of a right-angled triangle with right sides both of length $\frac{2}{3}$ and that the segment ST is the hypotenuse of a right-angled triangle of right sides $\frac{1}{3}$.

10. In this case, the formulae used for ages zero and one have been, respectively, $\widetilde{q}_0 = \frac{0.7D_0^t + 0.3D_0^{t+1}}{B^t + \frac{4}{10}N_0^t + \frac{1}{10}N_0^{t+1}}$ and $\widetilde{q}_1 = \frac{\frac{1}{2}(D_1^t + D_1^{t+1})}{C_0^t - 0.3D_0^t + \frac{1}{2}N_0^t + \frac{1}{3}N_1^t + \frac{1}{6}N_1^{t+1}}$; where for the estimation of $\widetilde{q}_0$ it has been assumed that on average the probability of death of a migrant of age zero located in the lower triangle of the period t is four times the probability of death of a migrant of the same age located in the upper triangle of the period $t+1$.

statistics come from sex and age vital statistics (INE, 2010a). And, approximations to sex and age annual immigrant and emigrant figures (for years 2006 to 2009) have been obtained from the data of residential variations (INE, 2010b)[11]. This section shows the differences obtained for the 2007 tables, before and after adjusting estimated crude probabilities. The adjustment has been carried out using nonparametric estimation; in particular, through a Gaussian kernel graduation (see, e.g., Ayuso et al., 2007, pp. 217-22).

Comparisons between values obtained for each age x with HCDS and HODS were carried out by use of two indicators of dissimilarity widely employed in the literature:

- Absolute relative error (ARE)

$$\frac{|\widehat{q}_x - \widetilde{q}_x|}{\widehat{q}_x}, x = 0,1,\ldots,99.$$

- Square relative error (SRE)

$$\frac{(\widehat{q}_x - \widetilde{q}_x)^2}{\widehat{q}_x}, x = 0,1,\ldots,99.$$

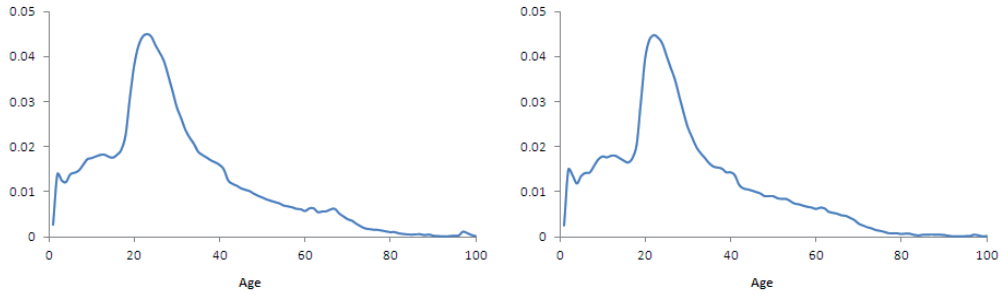


Figure 3: *Differences in crude probabilities with and without migration flows: men (left) and women (right).*

Figure 3 shows, in graphic form, ARE values obtained by comparing the estimates of crude probabilities achieved for men (left) and women (right), after applying equations (2) and (3), with and without migratory flows. As can be seen, the differences of considering HCDS or HODS are significant, reaching the greatest dissimilarities in the range of 14 to 36 years, with the maximum in both cases being reached at age 22, where the difference is close to 4.5%. Similar results are reached when *SRE* is used as measure of dissimilarity: the age range with greater differences remains the same.

11. It should be noted that the statistic of residential variations cannot be observed as a completely true source for migration flows given that this is just an account of the entrants and exits registered on the lists of the municipalities. This has been used, nevertheless, because it represents the only public source that can be used as a proxy of the migration flows occurring in Spain during the period analysed.

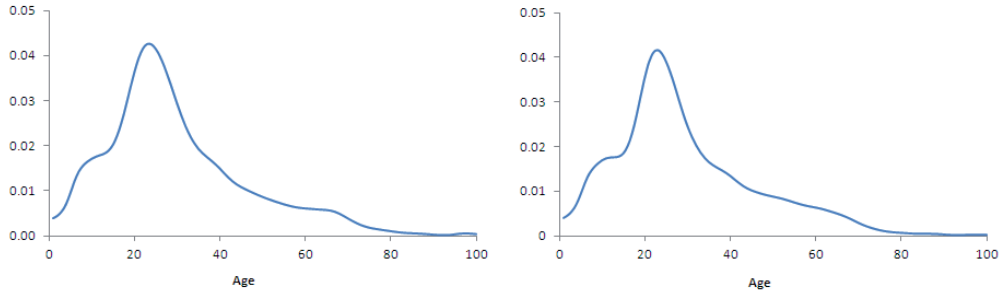


Figure 4: *Differences in graduated probabilities with and without migratory flows: (left) men and (right) women.*

The discrepancies observed reveal, at least in this case, the usual assumption of HCDS being inadequate. This provokes a systematic, not uniform overestimation in the probabilities of death for all ages; which, as can be seen below (Table 1 and Figure 5), has asymmetrical effects on the results of different actuarial calculations.

In actuarial calculations, however, the crude death estimates obtained directly from observed data, $\widehat{q}_x$ or $\widetilde{q}_x$, are not usually used without being graduated first. The objective of graduation is to soften the crude estimates in a way that eliminates (or mitigates) the random fluctuations present in empirical data. In this study, graduation has been carried out using a kernel graduation (see, for example, Ayuso et al., 2007). In particular, the kernel estimation carried out uses a Gaussian function as a kernel with a window parameter, or bandwidth, equal to *1*.[12]

Once the initial values $\widehat{q}_x$ and $\widetilde{q}_x$, were graduated, the indicators of dissimilarity ARE and SRE introduced previously were again calculated. The results obtained for men and women with the ARE measurement are shown in Figure 4. The comparison with the graduated probabilities does not change in any way the conclusions reached previously; in fact they serve to reinforce the results already obtained.

Finally, in line with Pavía and Escuder (2003), some specific probabilities have been obtained with the aim of illustrating the differences that could be derived by using one or other hypothesis on the demographic system: Table 1 shows the results. As was expected of a demographic situation such as that lived in Spain, where in recent years entries have been significantly greater than exits, the non-inclusion of migratory flows underestimates the survival probabilities and overestimates the death probabilities. Differences in every case are evident to the third significant digit in men and (at most) the fourth digit in women. The impact, therefore, is different depending on the gender. In spite of the repercussions on the individual probabilities being similar for both sexes (see Figures 3 and 4), the inclusion of migratory flows has a significantly greater impact on men than on women, at least for the range of ages and periods considered.

12. Similar results were obtained with alternative bandwidth parameters.

Table 1: *Examples of probabilities.*

	Men		Women	
	HCDS	HODS	HCDS	HODS
${}_{25}q_{40}$	.1380204	.1371103	.0584317	.0580243
${}_{15\|10}q_{50}$	.2345362	.2338640	.1112529	.1109601
${}_{85}p_{0}$	.3442274	.3452564	.5672970	.5679357
${}_{20}p_{15}$	.9872419	.9876434	.9951881	.9953219

This asymmetric impact is also clearly visible in Figure 5, where the differences in life expectancy when either migration flows have been or have not been taken into account are shown. As can be observed, the underestimation in life expectancy that entails the non-inclusion of migratory flows is, almost for all ages, double in men than in women.

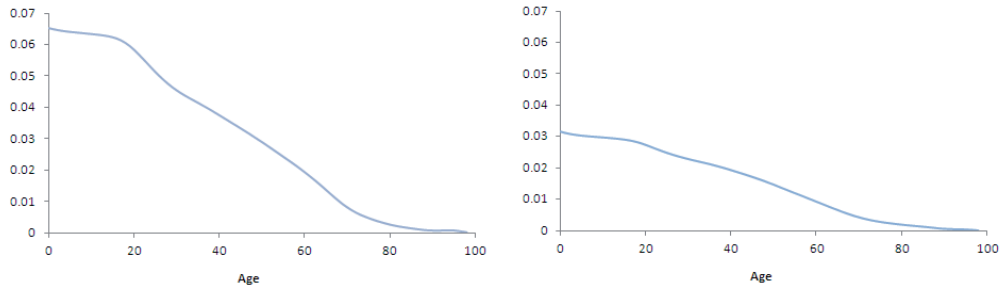


Figure 5: *Differences in life expectancy (in years) with and without migration flows: (left) men and (right) women.*

4. Conclusions

When working with general populations, the usual practice in the construction of life tables consists in ignoring the entry and exit flows that occur in the study group during the analysis period, under the assumption that these usually have little value compared to the size of the population. The use of a closed demographic system hypothesis has been consequently the norm among analysts.

In this paper, (i) the techniques used for estimating death probabilities have been extended to open demographic systems with aggregate data; and, (ii) the resulting estimator has been used, along with the classic HCDS estimator, to obtain (from 0 to 99 years) life tables of the resident population in Spain from 2006 to 2008.

Comparison of crude and graduated probabilities obtained with and without the inclusion of migratory flows shows the impact that entry and exit movements can have on actuarial and demographic calculations. In the examples considered the repercussion has been asymmetric by age and sex. On one hand, the greatest discrepancies, in relative

terms, are concentrated in the range of ages from 14 to 36 years, where the intensity of flows has been stronger in Spain in the recent years. On the other hand, by gender, it is clear that the impact on women is less, in contrast to that in men. The well-known lower probabilities of death that women suffer in the range of ages where greater probabilities of migration occur may be behind this result.

The results obtained in this paper point to the need to explicitly consider migratory flows in the estimation of life tables for general populations. The cost to introduce this information is minimal but the potential danger could be significant, especially in situations where entry movements are well above exit movements. In this type of situation, to omit migration flows would lead to an overestimation of probabilities of death and hence to an underestimation of life expectancy, with the danger that this could entail for a correct inter-generational planning that would ensure an adequate stability of the social security pension programmes characteristic of the Western welfare systems.

Certainly, beyond the possible influence that migratory flows and other relevant information might have on results, the great quantity of data provided by modern statistical systems offers an opportunity to seek new ways to exploit the available data in more efficient fashions. So, developing new methodological approaches or implementing proper analyses that help to assess the soundness of broadly used hypotheses should be included early on the statistical demographic research agenda. Along this line, in order to assess the cumulative impact of migration in mortality in Spain, it would be interesting to compare the probabilities of death and life expectancies of born-in-Spain and total (Spaniards and foreigners) populations. Likewise, death and migrant microdata should be analysed in order to ascertain the suitability of the uniform distribution hypotheses required when handling aggregate data.

Acknowledgements

The authors wish to thank M. Hodkinson for translating into English the text of the paper and the anonymous referees their valuable comments and suggestions. The authors acknowledge the support of the Spanish Ministry of Science and Innovation (MICINN) through the project CSO2009-11246.

References

Ayuso, M., Corrales, H., Guillén, M. Pérez-Marín, A. M. and Rojo, J. L. (2007). *Estadística Actuarial Vida*. Barcelona: Publicacions Universitat de Barcelona.

Benjamin, B. and Pollard, J. (1992). *The Analysis of Mortality and Other Actuarial Statistics*. London: Butterworth-Heinemann.

Copas, J. B. and Haberman, S. (1983). Non parametric graduation using kernel methods. *Journal of the Institute of Actuaries*, 110, 135-156.

Forfar, D. O, MacCutcheon, J. J. and Wilkie A. D. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, 115, 1-149.

INE (2007). *Metodología empleada en el cálculo de las tablas de mortalidad de la población de España 1992-2005*. Madrid: INE. Instituto Nacional de Estadística. http://www.ine.es/daco/daco42/mortalidad/metodo_9205.pdf

INE (2009). *Tablas de mortalidad. Metodología*. Madrid: INE. Instituto Nacional de Estadística. http://www.ine.es/daco/daco42/mortalidad/metodo_9107.pdf

INE (2010a). *Movimiento Natural de Población. Defunciones (cifras anuales)*. http://www.ine.es/jaxiBD/tabla.do?per=12&type=db&divi=MNP&idtab=49&L=0

INE (2010b). *Estadística de Variaciones Residenciales (01/02/2011)*. http://www.ine.es/prodyser/micro_varires.htm

INE (2012). *Estimaciones de la Población Actual de España*. http://www.ine.es/jaxiBD/menu.do?L=1&divi=EPOB&his=0&type=db

Livi Bacci, M. (2000). *Introducción a la Demografía*. Barcelona: Ariel.

Pavía, J. M. (2011). *101 Ejercicios Resueltos de Estadística Actuarial Vida*. Madrid: Garceta.

Pavía, J. M. and Escuder, R. (2003). El proceso estocástico de muerte. Diferentes estrategias para la elaboración de tablas recargadas. Análisis de sensibilidad. *Estadística Española*, 45, 243-274.