Functional k-sample problem when data are density functions

Pedro Delicado*

June 30, 2006

Abstract. This paper deals with the k-sample problem for functional data when the observations are density functions. First we consider two known ANOVA tests for generic functional data and we check their suitability for this particular kind of data. Then we introduce new procedures based on distances between pairs of observed density functions, allowing us to use the L_1 distance, the most natural choice for density functions. A simulation study is carried out to compare the practical behaviour of the available tests. Theoretical derivations have been done in order to allow weighted samples in the tests procedures. The paper ends with a real data example: for a collection of European regions we estimate the regional relative income densities and then we test the significance of the *country effect*.

Some key words: distance-based inference; Functional Data Analysis; Income distribution; k-sample problem; Monte Carlo Simulation; Nonparametric density estimation; Permutation tests; Weighted sample.

1 Introduction

Over the last years, the joint development of real-time measurement instruments and data storage computer resources has made possible to observe and save complete functions as results of random experiments. For instance, continuous-time monitoring clinical diagnostics or stock market information are common nowadays. Ramsay and Silverman (2005) express it saying that random functions are the *statistical atoms* in these cases. Functional Data Analysis (FDA) deals with the statistical description and modelization of samples of random functions. A broad outlook to FDA is given in the books by Ramsay and Silverman (1997) (see also the second edition, Ramsay and Silverman (2005)) and Ramsay and Silverman (2002).

It's well worthwhile noting that random functions can also be obtained from standard random samples, by the application of nonparametric curve estimation

*Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya, Barcelona (SPAIN), pedro.delicado@upc.edu. Research partially supported by the Spanish Ministry of Education and Science and FEDER, MTM2005-02370, and by the EU PASCAL Network of Excellence, IST-2002-506778. This work is part of an ongoing research with Magda Mercader (Departament d'Economia Aplicada, Universitat Autònoma de Barcelona), who has made helpful suggestions to improve this paper, and also provided the author with the relative equivalent disposable income data used in Section 7. Positive interchange with Adolfo Hernández is gratefully acknowledged, as well as suggestions received in the 4èmes Journées de Statistique Fonctionnelle et Opératorielle, Grenoble, 15-16 June 2006. The comments of two anonymous referees substantially improved this paper.

methods. For instance, Kneip and Utikal (2001) study the temporal evolution of income density functions in United Kingdom from 1968 to 1988. They work with yearly cross-sectional samples of households, and use nonparametric density estimation methods (kernel methods, to be specific) to obtain 21 income density functions over time, one corresponding to each year.

The one-way analysis of variance is one of the standard methods that have been generalized to be used in FDA, giving rise to *functional ANOVA*. In Ramsay and Silverman (1997) it is included in what they call functional linear models (a generalization of linear regression models). In addition to this approach, there are works devoted specifically to this topic (see Cuevas, Febrero, and Fraiman 2004 and references therein). The present paper lies within this framework.

Our interest for functional ANOVA was motivated by a real problem in regional income distribution in Europe. In a recent paper Mercader and Levy (2004) study the dependency between the regional Gini inequality indices and the country to which the region belongs, when the income distribution is considered both before and after taxes and transfers. They use standard univariate one-way ANOVA tests to show that there is no significant relationship before taxes and transfers, but the opposite happens after applying those redistributive instruments. To study to what extent this conclusion holds when considering the complete income distributions (instead of a summary inequality measure such as the Gini index) a functional ANOVA test applied to density functions is required. When using disposable income (after taxes and transfers) the effect of the country in regional income distribution is clear and expected: regions in rich countries tend to be richer than regions in poorer countries (see Perarnau 2005). Intuition is not so clear when talking about relative disposable income (every individual income data is divided by the regional median income). In doing so, the interest is centered on the role of the country factor in explaining the shape variability in regional income distributions, rather than their location variability. Relevant economic aspects, as relative poverty or inequality, depend on the income distribution shape. Figure 5 shows the estimated regional relative disposable income density functions, which are studied in detail in Section 7. See Delicado and Mercader (2006) for a partial analysis of what happens with incomes before taxes and transfers (they do not use all the tests presented here and they work with a different database). Let us note that in this example each region has a different weight, proportional to its population, and this particularity has to be taken into account in test procedures.

This work has two main objectives. The first one is to deal with the k-sample problem when data are density functions. We analyze the applicability of two known functional ANOVA techniques (one proposed by Ramsay and Silverman 1997 and the other one by Cuevas, Febrero, and Fraiman 2004). Moreover we present a distance-based ANOVA test (Gower and Krzanowski 1999) working on pairwise distances between observed data. This device allows us to use the L_1 distance, the most natural one for density functions: it is always well defined, it is invariant under monotone transformations of the argument, and it is closely related to the total variation distance between probability measures (see Devroye and Györfi 1985, Chapter 1). The same distance-based ANOVA

procedure is also applicable to other distance definitions and to functional data not being densities.

The second objective is to generalize functional ANOVA tests for weighted samples. We establish results valid for weighted data that generalize those of Gower and Krzanowski (1999) and Cuevas, Febrero, and Fraiman (2004).

The null distribution of any of the test statistics included in the paper is unknown. Thus, a functional ANOVA test using these statistics requires the use of permutations (or other specific Monte Carlo procedure, as in Cuevas, Febrero, and Fraiman 2004). There is no consensus in the literature as to what the best way of obtaining the permutation samples is (see Gower and Krzanowski 1999, and Anderson and Robinson 2001, and references therein). In this paper we consider two alternatives, both described in Section 4.1.

The paper structure is as follows. In Section 2 we set out the functional k-sample problem and discuss their connection with the functional ANOVA test. In Section 3 we summarize two known methods for functonal ANOVA and we heck if they are well suited for density functions (§3.1). In Section 4 we talk about distance-based ANOVA and permutation tests (§4.1). In Section 5 we present our proposals on the functional k-sample problem when data are density functions. A simulation study is included (§5.1) where we compare the performance in practice of six test statistics and the different ways to approximate their null distribution. Particular characteristics of weighted samples are the topic of Section 6. The example of the European regional relative income density functions is analyzed in Section 7. Last section sums up the conclusions of the paper. Proofs are deferred to an Appendix.

2 Notation and Preliminaries

In the functional k-sample problem it is assumed that n functions $f_{ri}(x)$ have been independently observed from the model

$$f_{ri}(x) \sim \mathcal{F}_r,\ i = 1, \ldots, n_r, r = 1, \ldots, k, x \in [a, b]\,, \tag{1}$$

where $\sum_r n_r = n$ and $\mathcal{F}_r$, $r = 1, \ldots, k$, are probability distributions of random processes, the trajectories of which are functions defined in $[a, b]$, with a and b real numbers or $\pm\infty$, and $a < b$. We are particularly interested in the case where the trajectories of $\mathcal{F}_r$ are density functions in $[a, b]$. The null hypothesis to be tested is the homogeneity of the k samples:

$$H_0 : \mathcal{F}_1 = \cdots = \mathcal{F}_k. \tag{2}$$

We also refer to testing this hypothesis as the functional k-sample problem.

A particularly interesting submodel of (1) is the functional ANOVA model, where the observed functions $f_{ri}(x)$ are assumed to verify

$$f_{ri}(x) = m_r(x) + e_{ri}(x), i = 1, \ldots, n_r, r = 1, \ldots, k, x \in [a, b]\,, \tag{3}$$

where $m_r(x)$ are k unknown mean functions and $e_{ri}(x)$ are independent trajectories drawn from a process with zero mean and covariance function $K(x, y) =$

$Cov(e_{ri}(x), e_{ri}(y))$. So the observations $f_{ri}(x)$ constitute k independent samples of random functions, each one with a specific function mean and a common covariance structure. This is the homoscedastic case. The heteroscedastic version allows for a different covariance function in each sample: $K_r(x, y) = Cov(e_{ri}(x), e_{ri}(y))$. The hypothesis to be tested in both cases is the equality of means:

$$H_0 : m_1(x) = \cdots = m_k(x), \quad \text{for all } x \in [a, b]. \tag{4}$$

This hypothesis coincides with (2) if we assume that in the ANOVA model (3) the noises $e_{ri}(x)$ have the same distribution in the k samples (for instance, assuming Gaussianity and homoscedasticity). In general, hypothesis (2) implies (4), but the inverse is not true (for instance, take the heteroscedastic case with equal group means).

When we work with density functions (positive and integrating 1 on $[a, b]$), this model is not appropriate because it does not automatically verify the required conditions on the zero mean additive noise $e_{ri}(x)$: it has to be such that $m_r(x) + e_{ri}(x) \geq 0$, it must belong to $L_1([a, b])$ and $\int_a^b e_{ri}(x)dx$ must be equal to 0.

A possible way to circumvent this difficulty is to consider a transformation Ψ of the density functions and replace model (3) by

$$\Psi(f_{ri})(x) = m_r^{\Psi}(x) + e_{ri}(x), i = 1, \ldots, n_r, r = 1, \ldots, k, x \in [a, b], \tag{5}$$

where Ψ is an injective functional, and the null hypothesis (4) by

$$H_0^{\Psi} : m_1^{\Psi}(x) = \cdots = m_k^{\Psi}(x), \quad \text{for all } x \in [a, b]. \tag{6}$$

A relevant example is the functional $\Psi(f) = \log(f)$. Using it there is no need of any sign restriction on $m_r^{\Psi}(x) + e_{ri}(x)$. Another useful transformation is

$$\Psi_N(f)(x) \equiv \frac{\partial}{\partial x} \log f(x)$$

that transforms the density of a $N(\mu, \sigma^2)$ random variable into the straight line $-(x - \mu)/\sigma$. Therefore it could be appropriate when observed densities f_{ri} are close to normality. Ramsay and Silverman (2002) use this transformation in the context of Functional Principal Component Analysis for data that are density functions. For the case of densities $f_{ri}(x)$, $x \in (0, \infty)$, close to log-normal density functions, a suitable functional is

$$\Psi_{lN}(f)(y) = \Psi_N(f(\exp(y)) \exp(y)), \; y \in (-\infty, \infty),$$

given that $f(\exp(y)) \exp(y)$ is the density function of $Y = \log(X)$, X having density f.

Observe that model (5) is different for different choices of Ψ. Also the corresponding null hypothesis (6) does not longer coincide in general with hypothesis (4), the original one, because $E[f_{ri}] \neq \Psi^{-1}(E[\Psi(f_{ri})])$ in general.

In this paper we borrow some statistic from functional ANOVA (see Section 3 below) and use them as test statistics for hypothesis (2).

3 Two known methods for functional ANOVA

We summarize here the proposals of Ramsay and Silverman (1997) and that of Cuevas, Febrero, and Fraiman (2004) for functional ANOVA using generic functional data. The next subsection is devoted to check their applicability in the case of data being density functions.

Ramsay and Silverman (1997) assume the additive ANOVA model (3) with homoscedastic noise. They propose to fix $x \in [a, b]$ and to compute the F-ratio statistic for the univariate ANOVA test

$$H_0^x : m_1(x) = \cdots = m_k(x).$$

Let F_x^R be the corresponding F-ratio statistic. Repeating the procedure for all x (in practice, for a grid of values $x_t, t = 1 \ldots T$) a F-ratio function $F^R(x) = F_x^R$, $x \in [a, b]$, is obtained. The values of the F-ratio function are expected to be much smaller under the null hypothesis than under the alternative. So it is sensible to take the integral of that function as the statistic for the functional ANOVA test:

$$T_F = \int_a^b F^R(x)dx.$$

Its null distribution is unknown and Ramsay and Silverman (1997) suggest to approximate it using a standard permutation-based mechanism (see Section 4.1). The use of a permutation test implies that homogeneity of the noise distribution in different groups is assumed (in general, homoscedasticity is not enough). An important theoretical drawback of this approach is that the integrability of F^R is not guaranteed.

Cuevas, Febrero, and Fraiman (2004) also consider the (now maybe heteroscedastic) ANOVA model (3) but their approach is different. They assume that f_{ri} are trajectories of an L_2-process and argue as follows. The classical F-ratio statistic for the univariate one-way ANOVA computes the ratio of variability *between* samples and *intra* sample. The functional version would be

$$F_n = \frac{\sum_{r=1}^k n_r \left\| \bar{f}_{r\bullet} - \bar{f}_{\bullet\bullet} \right\|^2 / (k-1)}{\sum_{r,i} \left\| f_{ri} - \bar{f}_{r\bullet} \right\|^2 / (n-k)} \tag{7}$$

where $\bar{f}_{\bullet\bullet} = \bar{f}_{\bullet\bullet}(x)$ is the global mean function, $\bar{f}_{r\bullet} = \bar{f}_{r\bullet}(x)$ is the mean function in the r-th sample, and $\|f\| = \left(\int_a^b f^2(x)dx \right)^{1/2}$ is the usual L_2 norm. The null hypothesis H_0 should be rejected if the numerator of F_n (a measure of the differences between groups) is too big, compared with the denominator of F_n (a measure of the variability of the noise process generating $e_{ri}(x)$). Cuevas, Febrero, and Fraiman (2004) indicate that it is enough to only consider the numerator of F_n when you are comparing values of the statistic coming from functional ANOVA models with noise processes having the same variability (all the denominators are estimating the same quantity). This is the case when an observed F_n value is compared with Monte Carlo simulated values and the

simulation is done to produce data according to the null hypothesis and having the same noise variability as the observed data. Technical reasons lead Cuevas, Febrero, and Fraiman (2004) to measure differences between groups using the statistic

$$V_n = \sum_{r<s} n_r \left\| \bar{f}_{r\bullet} - \bar{f}_{s\bullet} \right\|^2,$$

that is equivalent in practice to use the numerator of F_n (in the balanced case they only differ by a multiplicative constant). Their Theorem 1 establishes that the asymptotic distribution of V_n under H_0 coincides with that of the statistic

$$V = \sum_{r<s} \left\| Z_r - C_{rs} Z_s \right\|^2,$$

where $C_{rs} = (p_r/p_s)^{1/2}$, $(n_r/n) \to p_r$ as $n \to \infty$, and $Z_r = Z_r(x)$, $r = 1, \ldots, k$, are independent Gaussian processes with 0 mean and covariance function $K_r(x, y)$, that can be consistently estimated by

$$\hat{K}_r(x, y) = \sum_{i=1}^{n_r} \frac{1}{n_r - 1} \left(f_{ri}(x) - \bar{f}_{r\bullet}(x) \right) \left(f_{ri}(y) - \bar{f}_{r\bullet}(y) \right).$$

In the homoscedastic case the natural estimator of the common covariance function is

$$\hat{K}(x, y) = \frac{1}{n - 1} \sum_{r=1}^{k} (n_r - 1) \hat{K}_r(x, y).$$

This theoretical result offers an asymptotic Monte Carlo procedure to tabulate the null distribution of V_n: a large number N of values of statistic V are simulated (say $V_l^*, l = 1, \ldots, N$) and the p-value corresponding to the observed V_n is computed as the proportion of simulated values V_l^* greater than V_n.

Assuming homoscedasticity, an alternative way to obtain a valid approximation to the null distribution of V_n is to use a permutation mechanism.

3.1 Applicability when functions are density functions

The proposals of Ramsay and Silverman (1997) and Cuevas, Febrero, and Fraiman (2004) rely on the additive ANOVA model (3) and we have said in Section 2 that this model is not well suited for density functions. There is an additional difficulty related with integrability issues.

When applying the test of Cuevas, Febrero, and Fraiman (2004) it is required to have k samples of functions in L_2. This may not be the case when we are comparing samples of density functions because some densities do not belong to L_2 (they would do if they were bounded, for instance). We could consider as primary functional data the squared root of the density functions: $\Psi_{\sqrt{}}(f_{ri})(x) = \sqrt{f_{ri}(x)}$. These new functions are always in L_2 and the Cuevas, Febrero, and Fraiman (2004) functional ANOVA test is applicable to them. Unfortunately, the problem of lack of positiveness appears again.

Other functionals Ψ could also lead to a model (5) with trajectories $\Psi(f_{ri})$ not in $L_2([a,b])$. For instance, if f_{ri} is the density function of a $N(\mu_r+\delta_{ri},\sigma^2)$, δ_{ri} is drawn from a zero mean random variable, and $\Psi=\Psi_N$, then $\Psi_N(f_{ri}(x))=-(\mu_r+x-\delta_{ri})/\sigma$, $x\in(-\infty,\infty)$, is not in $L_2(-\infty,\infty)$. A natural solution is to limit the analysis to a shorter compact interval $[a^*,b^*]\subseteq[a,b]$ where the function inside the integral are bounded. This choice may lead to a lowering of power. Moreover, the test conclusion could depend on the choice of the new interval.

Regarding the Ramsay and Silverman (1997) proposal, let us consider the integrability problems of the F-ratio function F^R. A way to assure integrability follows. Observe that the denominator of the F-ratio F_x^R is an estimation of $\sigma^2(x)=V(e_{ri}(x))$. Then the statistic T_F is approximately equal to

$$\sum_{r=1}^{k}\int_a^b(\bar{f}_{r\bullet}(x)-\bar{f}_{\bullet\bullet}(x))^2\frac{1}{\sigma^2(x)}dx.$$

Thus a sufficient condition for the integral of F^R being finite is that the observed trajectories f_{ri} verify $\int_a^b f_{ri}^2(x)/\sigma^2(x)dx<\infty$. These integrals are squares of weighted L_2-norms of functions f_{ri}. In fact we could say that the test based on the F-ratio function F^R corresponds to the test of Cuevas, Febrero, and Fraiman (2004) whith a different norm definition. When a transformation Ψ is used, determining whether or not they are finite is at least as difficult as in the usual L_2-norm case. If no transformation is used, a sufficient condition is that $\sigma^2(x)\geq f_{ri}(x)$ for all $x\in[a,b]$.

There are two natural ways to avoid the F-ratio function integrability problems. The first one is to limit the analysis to a shorter compact interval $[a^*,b^*]\subseteq[a,b]$ such that $f_{ri}(x)/\sigma(x)\leq M<\infty$ for all $x\in[a^*,b^*]$. We have mentioned before the drawbacks of such a practice. The second one is to argue as in Cuevas, Febrero, and Fraiman (2004) and then consider as test statistic the integral of just the numerator of the F-ratio function. It is easy to see that doing this the resulting statistic is equivalent to that used by Cuevas, Febrero, and Fraiman (2004).

We conclude that approaches of Ramsay and Silverman (1997) and Cuevas, Febrero, and Fraiman (2004) are not completely satisfactory when dealing with density functions neither under the original model (3) nor under the transformed one (5). The main two problems are that both the additivity assumption and the use of L_2 distances are not natural when working with density functions.

In the next section we introduce a framework to test hypothesis (2) that overcomes these difficulties. In particular, we will be able to construct the functional k-sample problem for density functions based on L_1 distances between observed densities. We also whow that statistics equivalent to those proposed by Ramsay and Silverman (1997) and Cuevas, Febrero, and Fraiman (2004) can be obtained in this new framework.

4 ANOVA based on distances

The geometrical concept of distance between individuals or populations has an important role in Statistics. Techniques as multidimensional scaling are entirely based on distances between observations. Other statistical methods admit versions taking exclusively a matrix of inter-individual distances as the relevant information from the data. Examples are the regression based on distances (Cuadras and Arenas 1990), the ANOVA based on distances (Gower and Krzanowski 1999; see below for more details) or the early MRPP tests (Mulri-Response Permutation Procedures; see Mielke, Berry, Brockwell, and Williams 1981, for instance). Arenas and Cuadras (2002) present a survey on statistical methods based on distances. Distance-based statistical methods work for a very huge variety of observed objects, (for instance, individuals where non-numerical or mixed variables have been observed) because the only requirement is to be able to define a metric between objects.

Gower and Krzanowski (1999) provide a framework for the analysis of any data set whose structure conforms to that of an ANOVA model, but is not analyzable with this technique because ANOVA assumptions are not fulfilled (the case where some or all of the variables are categorical is an example). They assume that there are n individuals (divided into k groups of sizes $n_1, \ldots, n_k$) and that a distance function between individuals is available. Let $d_{ij} \geq 0$ be the dissimilarity between individuals i and j. It is assumed that $d_{ij} = d_{ji}$ and that $d_{ii} = 0$ for all $i, j = 1 \ldots, n$. They define the $n \times n$ matrix D with element (i, j) equal to $d_{ij}^2/2$. Let Δ be the $n \times n$ matrix with element (i, j) equal to d_{ij}. Let G be the $n \times k$ matrix indicating to which group every individual belongs: $g_{ir} = 1$ if individual i is in group r and $g_{ir} = 0$ otherwise.

Let us assume that for $q \leq n$ there exists a $n \times q$ data matrix X such that the Euclidean distance between the i-th and j-th rows of X is d_{ij}. We would say that X is an *Euclidean configuration* of Δ. Such a configuration does not always exist. When it does, the distance matrix Δ is said to be *Euclidean*. Consider the usual ANOVA analysis with the rows of X as dependent variables and groups indicated by G. Let T, W and B be the *total*, the *within-group* and the *between-group* sums of squares respectively. Gower and Krzanowski (1999) prove that

$$T = \frac{1}{n}\mathbf{1}'D\mathbf{1},\ W = \sum_{r=1}^{k} \frac{1}{n_r}\mathbf{1_r}'D_{rr}\mathbf{1_r},\ B = \frac{1}{n}\mathbf{n}'D_B\mathbf{n}, \tag{8}$$

where $\mathbf{1} = (1, \overset{(n)}{\ldots}, 1)'$, $\mathbf{1_r} = (1, \overset{(n_r)}{\ldots}, 1)'$, $\mathbf{n} = (n_1, \ldots, n_k)'$, D_{rr} is the diagonal block of matrix D corresponding to the r-th group of individuals, and D_B is the $k \times k$ matrix with element (r, s) equal to $(\delta_{rs}^w)^2/2 = (2f_{rs} - f_{rr} - f_{ss})/2$, f_{rs} being the element (r, s) of matrix

$$F = \mathbf{N}^{-1}G'DG\mathbf{N}^{-1}$$

and $\mathbf{N} = \text{diag}(\mathbf{n})$. They also prove that the fundamental ANOVA identity $T = W + B$ remains valid even if the distance matrix Δ is not Euclidean and

T, W and B are directly defined by equation (8). Observe that the value δ_{rs}^w represents the distance between groups r and s.

The distance-based ANOVA statistic is based on the ratio B/W. Alternatively, arguing as Cuevas, Febrero, and Fraiman (2004) do to consider only the numerator of F_n in equation (7), it is enough to consider the between group sum of squares B as test statistic if the reference distribution under the null hypothesis is computed through a permutation test (this is the suggestion of Gower and Krzanowski 1999 and it is what we do in this paper). The MRPP test of Mielke, Berry, Brockwell, and Williams (1981) is based on W, that is equivalent to B given that $W + B = T$.

The previous arguments can be entirely reproduced with the following modification. Instead of assuming that there exist n elements in $\mathbb{R}^q$ (the rows of X) with Euclidean distances d_{ij} between them, consider now that there are n elements $x_1, \ldots, x_n$ in a linear space $\mathcal{X}$ with inner product such that $d_{ij} = \|x_i - x_j\|$. The main consequence of this observation is that if we consider a sample of functions in $L_2([a, b])$ and we use their L_2 distances to define the distance matrix Δ, then $B/(k-1)$ and $W/(n-k)$ are equal to the numerator and the denominator of F_n in equation (7), respectively. Therefore in this case it is equivalent to work with the test statistic proposed by Cuevas, Febrero, and Fraiman (2004) or with the statistic B computed from the distance matrix. Small differences may appear in the unbalanced case and when using permutations or Monte Carlo simulation to compute p-values. The same applies for the test proposed by Ramsay and Silverman (1997) and distances computed as weighted L_2-norms.

Let us note that in distance-based ANOVA, as it happens in the test proposed by Ramsay and Silverman (1997), the use of a permutation mechanism to approximate the null distribution of the test statistic implies the assumption of some kind of homoscedasticity. But it is not apparent what the meaning of homoscedasticity is when working only with inter-individual distances.

It is clear from equation (8) that the only required information from the data in order to apply a distance-based ANOVA test is the inter-individual distance matrix. So this framework allows us to analyze ANOVA models for any kind of statistical objects as long as a distance measure between objects is available. In particular we can analyze functional data using any reliable distance between the observed functions. In this paper we apply this procedure to the specific case of the ANOVA model for density functions. See Section 5 for more details.

4.1 Two ways to do permutation tests

The standard way to carry out an ANOVA permutation test is as follows (see for instance Manly 1997, Chapter 7). Let f_{ri} be data following the model (1). The subscript r is the label indicating to which group the data f_{ri} belongs. The procedure to obtain pseudo functional data sets according to the null hypothesis (2) consists in randomly permute the group label r of the observed functions. Under the null hypothesis the original sample and the permuted sample are interchangeable, and so they are the observable statistic T and the value T^p that it takes in a permuted sample. Let T_0 be the observed value of the ANOVA

test T in the actual sample. The permutation mechanism is repeated a large number N of times in order to generate values $T_1, \ldots, T_N$ from the permutation distribution of T. The p-value of the permutation test is defined as $\#\{T_s \geq T_0 : s = 1, \ldots, N\}/N$.

When a distance-based ANOVA test is used, the data are used through a $n \times n$ distance matrix Δ (or equivalently, through the transformed matrix D) and a $n \times k$ matrix G, whose i-th row indicates to which group the i-th observed data belongs. In this context a permuted sample is obtained by randomly permuting the rows of G, whereas matrix Δ remains unaltered. Then equation (8) is used to obtain the value of the between-group variability in the permuted sample.

This permutation test presents a drawback under the alternative hypothesis. The between group variability is translated by the permutation procedure to noise variability in the permuted samples. Therefore the artificial samples verify the null hypothesis of groups homogeneity, but the noise variability is greater than the corresponding to the original data. The main consequence of the increment in noise variability is the reduction of the test power: small deviation from the null hypothesis would not be detected because of the precision loss.

The preceding procedure can be modified to obtain a more powerful permutation test, as it was suggested by ter Braak (1992) in the context of the standard linear model. Consider the additive model (3). Instead of permuting the observed functions, we can permute the estimated residuals and define the artificial functions as the sum of the global mean plus a permuted estimated residual,

$$f^p_{ri}(x) = \bar{f}_{\bullet\bullet}(x) + \hat{e}^p_{ri}(x),$$

and $\hat{e}^p_{ri}(x)$ is selected from the estimated residuals, $\hat{e}_{rl}(x) = f_{rl}(x) - \bar{f}_{r\bullet}(x)$, by random permutation. When the null hypothesis is false this modified permutation test guarantees that the artificial data verify the null hypothesis and have approximately the same noise variability as the original sample. Therefore this modified test (that we can call *permuting residuals*, in contrast with *permuting observations*, the standard one) would detect deviations from the null hypothesis that would overlooked by the standard permutation test.

Manly (1997) points out that doing the ANOVA test permuting residuals is very close to doing a bootstrap test. In the context of bootstrap tests Hall and Wilson (1991) give two general recommendations (see Delicado and del Río 1994 for their application in the one-way ANOVA model): to do the resampling in a way that reflects the null hypothesis even if it is false (a serious reduction in power can occurs otherwise), and to use asymptotic pivotal test statistics (what improves the level accuracy of the test). The first guideline can also be expressed as follows: estimate the model assuming that the data were generated under the alternative hypothesis, and generate the bootstrap data under the null hypothesis. The permuting residuals scheme complies with this recommendation completely, whereas permuting observations follows only its second part. This argument suggests that the permuting residual test will be more powerfull than the other one.

Despite the rationale behind the permuting residual test, there are references

in the literature indicating that the claimed increase in power (with respect to permuting observations) does not seem to occur. For instance, the permuting residuals procedure is analyzed in Manly (1997), Sections 6.5 and 7.2, for the standard ANOVA test. Manly points out that it is not an exact test (there is no guarantee that a nominal α level test will have probability of Type I error not greater than α) but he also indicates that the behaviour of the permuting residuals test under the null hypothesis is acceptable in practice. Regarding the performance of both alternatives under the null hypothesis, the limited Monte Carlo experiments done by Manly (1997) indicate that they give similar results. In a more complex model (testing of partial regression coefficients in a multiple linear regression model) Anderson and Robinson (2001) prove that permuting observations and permuting residuals have the same asymptotic power under local alternatives, and their simulations support the theoretical result.

In order to find out whether or not permuting residuals improves the power of the standard permutation test in the functional ANOVA model, we have included both procedures in the simulation study presented in Section 5.1.

We finish this section indicating how the permuting residual ANOVA test can be done when the test statistic is a distance-based statistic. The inter-residual distance matrix Δ^R (or equivalently, the transformed matrix D^R) is required. The following Proposition shows that Δ^R can be computed directly from the original distance matrix Δ and the group indicator matrix G. The proof is deferred to the Appendix.

Proposition 1 *Assume that the $n \times n$ distance matrix Δ is Euclidean, that the $n \times q$ matrix X is an Euclidean configuration of Δ and that X^* is the $n \times q$ matrix of residuals obtained by fitting an ANOVA model to X where the groups are defined by the $n \times k$ matrix G. Then the Euclidean distance d_{ij}^R between rows i and j of X^* is given by*

$$(d_{ij}^R)^2 = 2e_{ij}^R - e_{ii}^R - e_{ij}^R,$$

where e_{ij}^R is the entry (i,j) of the $n \times n$ matrix E^R defined as

$$E^R = \left(I_n - G\mathbf{N}^{-1}G'\right) D \left(I_n - G\mathbf{N}^{-1}G'\right),$$

where I_n is the $n \times n$ identity matrix, and $\mathbf{N}$ and D were defined in Section 4.

The matrix Δ^R having element (i,j) equal to d_{ij}^R is the inter-residual distance matrix, and associated to it we define the matrix D^R with element (i,j) equal to $(1/2)(d_{ij}^R)^2$. It can be seen that Δ^R does not depend on what particular Euclidean configuration X has been chosen in Proposition 1. So we define the inter-residual distance matrix associated to the distance matrix Δ as Δ^R, even if Δ is not Euclidean.

When permuting residuals, the reference statistic values (to which the between-group variability B in the original sample has to be compared) are generated taking Δ^R as distance matrix, randomly permuting the rows of G, and applying the definition of between-group variability given in equation (8).

5 Functional ANOVA for density functions

Given that density functions are always in L_1, we propose to base functional ANOVA tests on L_1 distances between observed densities. Let Δ^{L_1} be the $(n \times n)$ matrix of pairwise L_1 distances $\int_a^b |f_i(x) - f_j(x)|dx$. The distance-based ANOVA of Gower and Krzanowski (1999) is directly applicable to this distance matrix.

Other distances can be the base of this methodology: for instance, L_2 distance between densities (assuming they are well defined) or L_2 distance between squared root of densities (this is also known as Hellinger distance). As we pointed out in Section 4, in this two specific cases the resulting distance-based tests are equivalent to using directly the proposal of Cuevas, Febrero, and Fraiman (2004) on the observed functions or on their squared root, respectively. A similar connection exists between the Ramsay and Silverman (1997) proposal and the weighted L_2-norm $\|f\|_\sigma = \left(\int_a^b f^2(x)/\sigma^2(x)dx\right)^{1/2}$. This gives us an argument for using these test statistics with density functions.

From now on, we use the proposals of Ramsay and Silverman (1997) and that of Cuevas, Febrero, and Fraiman (2004) as two more tentative test statistics for testing hypothesis (2) in model (1). Their use here does not imply that we are assuming that the observed density functions follow model (3). It is also possible to use distances between transformed densities $\Psi(f)$. In the framework of distance-based ANOVA, the use of a specific distance or transformation Ψ implies the definition of a specific test statistic for testing hypothesis (2). Some of them would be better than other, but we are always testing the same null hypothesis. Remember that in the additive ANOVA model (Section 3) changes in the choice of functional $\Psi(f)$ entail changes in the assumed model and in the null hypothesis to be tested.

In practice it is not possible to observe real density functions and we are only able to work with estimates of them. Let us note that working with estimated densities has special characteristics. First, in order to have a true null hypothesis (2) applied to estimates $\hat{f}_{ri}$ it is required that this hypothesis is true for the unknown densities f_{ri}, that the samples from each f_{ri} used to compute the nonparametric density estimators have equal sizes and that the bandwidth choice method is always the same. Otherwise the probability distribution of $\hat{f}_{ri}$ would depend either on the specific sample size or on bandwidth. An alternative requirement for sample sizes is to consider them as random variables following the same law.

Secondly, the additive ANOVA model (3) is harder to be accepted than in the case of known densities. Even if $f_{ri}(x) = m_r(x) + e_{ri}(x)$, with $e_{ri}(x)$ having zero mean, the biased nature of the usual non-parametric density estimators (kernel methods among them) implies that the working densities are

$$\hat{f}_{ri}(x) = m_r(x) + e_{ri}(x) + h_{ri}(x),$$

and the error term $e_{ri}(x) + h_{ri}(x)$ is no longer a centered process. In order to have homoscedasticity, same conditions on sample sizes and bandwidth choice

as before are required.

When working with distance-based tests, the observed distance between estimated densities $\hat{f}_{ri}$ and $\hat{f}_{sj}$ is a kind of random dissimilarity between f_{ri} and f_{sj}: as a way to evaluate whether f_{ri} and f_{sj} are close or not, we take random samples from both densities, we compute nonparametric density estimators from the samples, then we apply a distance formula to the estimators and we annotate this value as a dissimilarity between f_{ri} and f_{sj}. This argument supports the use of tests based on distances between estimated densities, even if you want to test the hypothesis of groups homogeneity for true densities. In the next subsection a part of the simulation study is devoted to evaluate the effect of replacing true densities by estimates of them.

5.1 Simulation study

A simulation study is carried out to compare the practical behaviour of six test statistics:

1. V_n, from the test of Cuevas, Febrero, and Fraiman (2004) applied directly to observed density functions. (We have labeled it as V in tables and figures below).
2. $V_n^{\surd}$, from the test of Cuevas, Febrero, and Fraiman (2004) applied to the squared root of observed density functions. (Labeled as Vsqrt below).
3. T_F, from the test based on the F-ratio function. (Labeled as Frat below).
4. T_{L_1}, from the distance-based test for L_1 distances between observed density functions. (Labeled as L1 below).
5. T_{L_2}, from the distance-based test for L_2 distances between observed density functions. (Labeled as L2 below).
6. $T_{L_2}^{\surd}$, from the distance-based test for L_2 distances between the squared root of observed density functions. (Labeled as L2sqrt below).

We also pay attention to what extent the different available Monte Carlo methods to reproduce the null distribution of these statistics are adequate. Two models are considered.

Model 1. We have simulated random density functions following model (5) with functional Ψ_{lN}, so they are close to log-normality. To be precise, we generate density functions $f(x)$ in such a way that

$$\Psi_{lN}(f)(y) = \eta(y) = -\frac{y-\mu}{\sigma} + \varepsilon(y),\ y = \log(x),\ x \in (0,\infty),$$

for some $\mu \in I\!R$, $\sigma > 0$, where $\varepsilon(y)$ is a smooth noise process. Therefore, the densities we are working with have the expression

$$f(x) = \frac{\exp\left(\int_{-\infty}^{\log(x)} \eta(u)du\right)}{\int_{-\infty}^{\infty} \exp\left(\int_{-\infty}^{y} \eta(u)du\right) dy} \frac{1}{x},\ x \in (0,\infty). \tag{9}$$

Specifically, $\varepsilon(y)$ is the function of y obtained as the local linear fit of the data $(y_i, e_i), i = 1, \ldots, n_e$, where y_i are evenly spaced points between $\mu - 3\sigma$ and $\mu + 3\sigma$, and e_i are i.i.d. observations from a $N(0, \sigma_e^2)$. In our implementation (done in the package `R`, R Development Core Team 2005), we have chosen $n_e = 51$, $\sigma_e = 0.5$, and the function `loess` (with `span`=0.25) as local linear fitting method.

The number of groups is $k = 3$, and $n_r = 10$, $r = 1, 2, 3$, so $n = 30$. Let f_{ri} be one of the simulated density functions. The simulation parameters have been taken to have

$$\Psi_{lN}(f_{ri})(y) = -\frac{y - \mu}{\sigma_r} + \varepsilon_{ri}(y),$$

with $\mu = 0$, $\sigma_1 = s$, $\sigma_2 = 1$, $\sigma_3 = 1/s$, $s = 1 + .03 \times i$, $i = 1 \ldots 5$. When $s = 1$, data follow the null hypothesis of groups homogeneity. The other five values of parameter s allow for tests power evaluation. The numerical integrals have been done in the interval $[a, b]$, $a = \exp -3 \times 1.15$, $b = \exp 3 \times 1.15$.

We have simulated 200 samples for each value of parameter s. For each sample, the six statistics listed above are computed. For statistics V_n and $V_n^{\surd}$ there are three alternative ways to approximate their null distribution (simulating Gaussian processes as the asymptotic result of Cuevas, Febrero, and Fraiman 2004 indicates, permuting observations and permuting residuals). For the other statistics there are only two ways (permuting observations and permuting residuals). The number of simulated Gaussian processes and the number of pseudo-samples obtained by permutations is always 200 in our simulation study.

In order to have an idea about the behaviour of the statistics independent of the resampling mechanism used to approximate their null distribution, we have generated 1000 extra samples following H_0. This way we have samples of size 1000 for the six statistics under the null hypothesis, the empirical distributions of which are a good approximation to the statistics null distribution. Observe that in a real case we are not able to compute such empirical distribution functions because we do not know the real data generating process.

Figure 1 shows the six statistics power functions (s varying from 1 to 1.15) when these empirical distributions functions are used as reference distributions. The theoretical level is fixed in $\alpha = 0.05$. We observe that the power functions corresponding to tests based on T_{L_2} and V_n coincide, as it was expected to happen because the design is balanced. The same comment applies for tests based on $T_{L_2}^{\surd}$ and $V_n^{\surd}$. It is clear that the most powerful test is that based on T_F (F-ratio), followed by the test based on $T_{L_2}^{\surd}$ (or $V_n^{\surd}$). We also observe that tests based on T_{L_1} and T_{L_2} (or V_n) perform similarly. Obviously, the empirical level (the ordinate of the power function for $s = 1$) is close to the theoretical one in all cases.

Figure 2 compares the unfeasible power functions (those labeled as H0, already displayed in Figure 1) with the feasible ones (those based on permutations, or on simulated Gaussian processes). The interest here is to determine for each statistic what feasible procedure gives the closest power function to the unfeasible one.

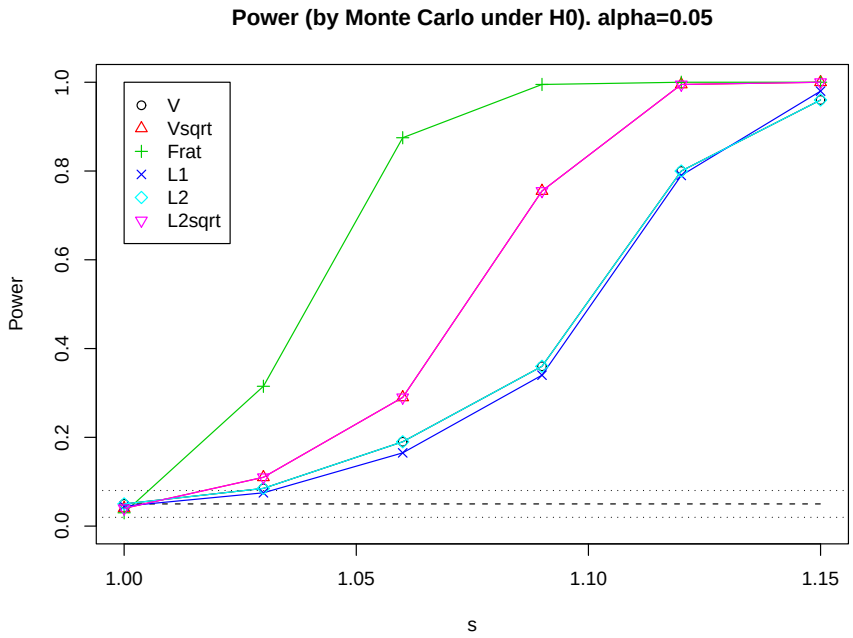

Figure 1: Power functions based on the empirical distribution function of size 1000 samples simulated under the null hypothesis. Horizontal doted lines are pointwise acceptance bands for the null hypothesis that the true power is $\alpha = 0.05$.

In the case of the test based on the F-ratio function (T_F, label as Frat in the figure) permuting observations and permuting residuals are very close to the unfeasible one. For other test statistics, permuting residuals leads to empirical levels greater than the nominal ones and they tend to over-estimate the test power. Moreover, permuting observations is always closer to the true power function than permuting residuals. For statistics V_n and $V_n^{\sqrt{}}$ the procedures based on simulated Gaussian processes (labeled as CLT) performs worse than permuting residuals and better than permuting residuals under the null hypothesis. Under the alternative they are comparable to permuting observations.

The high power reported in Figure 1 by the test based on the F-ratio function, points this functional ANOVA test as the most advisable one in Model 1. We also conclude that in this case permuting observations is a better practice than permuting residuals.

Model 1. Unbalanced case. We have repeated the preceding simulation plan varying the sample sizes. Now we use $n_1 = 5$, $n_2 = 10$ and $n_3 = 15$. Figure 3 (similar to Figure 1) shows the differences between power functions for the balanced design and the unbalanced one. All the methods lose power in a similar way when the design is unbalanced. It can also be seen that in the unbalanced case the tests based on Cuevas, Febrero, and Fraiman (2004) differ slightly from the corresponding distances based tests.

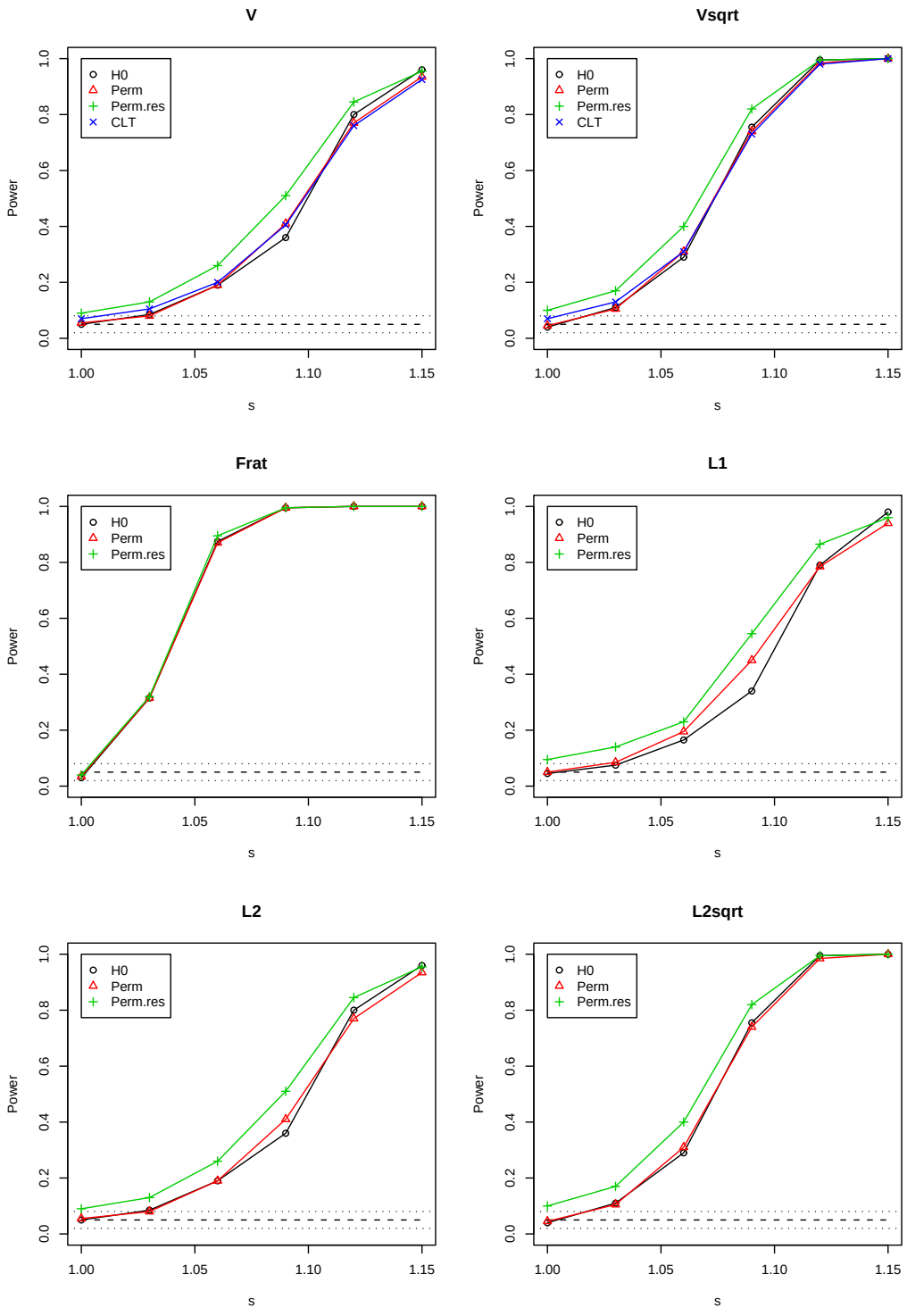

Figure 2: Power functions based on permutations (Perm, Perm.res) and on simulated Gaussian processes (CLT), compared with those based on simulation under the null hypothesis (H0).

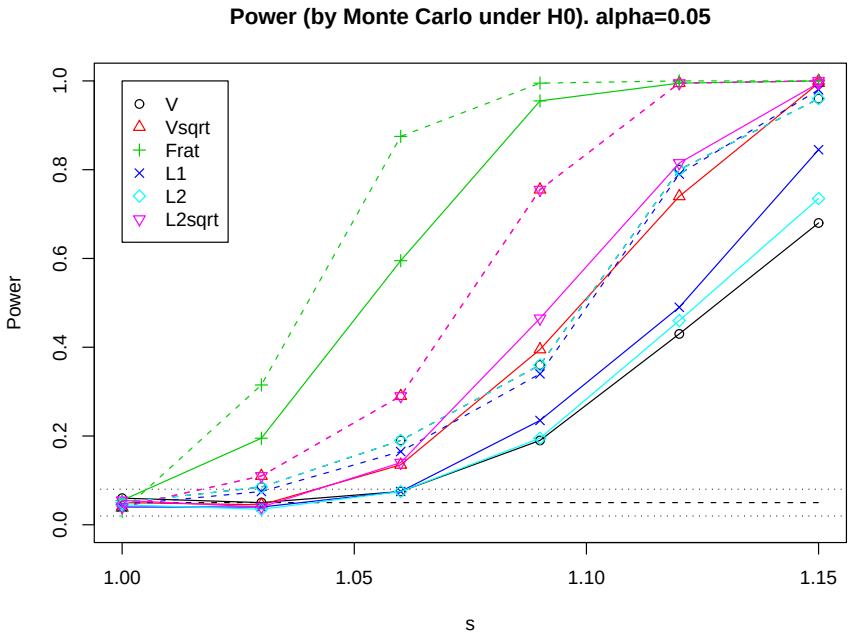

Figure 3: Power functions under the null hypothesis. Comparison between balanced (dashed lines) and unbalanced (solid lines) designs.

Let us now talk about the performance of the feasible methods to approximate the statistic null distributions. A figure analogous to Figure 2 (not shown here) allows us to say that in the unbalanced case permuting observations works well in all cases, that permuting residuals lead to power functions always over the true ones, and that the procedures based on simulated Gaussian processes (for statistics V_n and $V_n^{\surd}$) are worse than permuting residuals (and consequently worse than permuting observations).

Model 2. Estimated densities. Now we want to explore the effect of nonparametric density estimation in the tests results. We reproduce the simulation plan of the balanced Model 1 and, instead of working with density functions f_{ri} as in equation (9), we generate from them 500 random numbers and we compute kernel estimators $\hat{f}_{ri}$. We use these estimated density functions as our primary functional data.

The kernel estimation of f_{ri} is done as follows. We generate random data from $Y = \log(X)$, X having density f_{ri} (Y is close to normality). We take the usual kernel estimation of the density of Y, choosing the bandwidth by the normal reference rule. Changing the variable back, we obtain $\hat{f}_{ri}$.

Figure 4 (similar to Figure 1) shows the differences between power functions for Model 1 (dashed lines), where the true densities were known, and Model 2 (solid lines), where only kernel estimators are available. It can be seen that the most remarkable power decrease is for the test based on the F-ratio function.

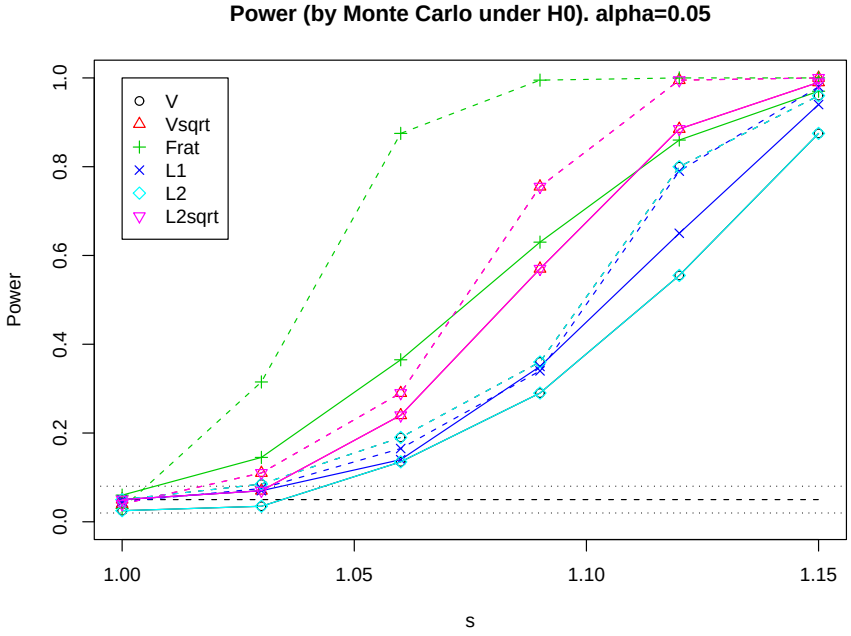

Figure 4: Power functions under the null hypothesis. Comparison between using true (dashed lines) and estimated (solid lines) densities.

Observe that the effect of having to estimate densities is greater than the effect of having an unbalanced design (compare Figures 4 and 3). The other tests lose power less markedly here than in the case of unbalanced design. In fact, Figure 4 shows that the test based on $T_{L_2}^{\surd}$ (or $V_n^{\surd}$, because the design is balanced) is comparable to that based on T_F (F-ratio).

With respect to the performance of the feasible methods to approximate the statistic null distributions, simulation results not shown here indicate that conclusions drawn from Model 1 are also valid when using estimated densities.

We conclude that estimated densities are useful for testing the null hypothesis (2) of groups homogeneity for the unknown true density functions. The most advisable test are those based on T_F or $T_{L_2}^{\surd}$ using a permuting observations resampling scheme to obtain the p-value.

6 Methods for weighted data

Let us now discuss an important question arising in economic and social micro data bases: in these contexts it is usual that each observation has a different weight, proportional to the amount of people in the population it is representing. The six functional ANOVA tests listed in Section 5 are not directly applicable to such weighted samples. In this section we provide valid versions of those tests for weighted samples.

The method based on the F-ratio function is easily adapted to this case: the univariate F-ratio statistics are now defined as the quotient between two weighted sums of squares, and the definition of the statistic T_R as the integral of the F-ratio function does not change.

The following Proposition establishes the analogue result to Theorem 1 in Cuevas, Febrero, and Fraiman (2004) for weighted samples. The proof is deferred to the Appendix.

Proposition 2 *Assume that each observed function $f_{ri}(x)$ has an associated weight $w_{ri} > 0$. Let $\bar{f}^w_{r\bullet} = \bar{f}^w_{r\bullet}(x)$ be the weighted mean function in the r-th sample. Let $w_{r\bullet} = \sum_{i=1}^{n_r} w_{ri}$ be the total weight of the i-th sample. Define*

$$\alpha^2_{nr} = \frac{w^2_{r\bullet}}{\sum_{i=1}^{n_r} w^2_{ri}}$$

and assume that $(\alpha^2_{nr} / \sum_{s=1}^{k} \alpha^2_{ns}) \to p^w_r \in (0,1)$ for all r, as $n \to \infty$. Then the statistic

$$V^w_n = \sum_{r<s} \alpha^2_{nr} \left\| \bar{f}^w_{r\bullet} - \bar{f}^w_{s\bullet} \right\|^2 \tag{10}$$

has the same asymptotic distribution under H_0 that the statistic

$$V^w = \sum_{r<s} \left\| Z_r - C^w_{rs} Z_s \right\|^2,$$

where $C^w_{rs} = (p^w_r / p^w_s)^{1/2}$, and $Z_r = Z_r(x), i = 1, \ldots, k$ are independent Gaussian processes with 0 mean and covariance function $K_r(x, y)$.

The statistic V^w_n is the version of V_n appropriate for weighted samples. Its null distribution can be approximated by Monte Carlo simulation based on Proposition 2. The coefficients C^w_{rs} can be approximated by $C^w_{nrs} = (\alpha_{nr}/\alpha_{ns})$ to compute V^w in practice. Permutations method is also a suitable way to approximate the null distribution of V^w_n in the homoscedastic case.

Let us now give appropriate versions of the distance-based ANOVA test when data are weighted. First we will provide a decomposition of the total variability generalizing equation (8), and then we will give the weighted version of Proposition 1.

The next proposition states the result that extends that of Gower and Krzanowski (1999). The proof is deferred to the Appendix. We need some extra notation. Let $\mathbf{w} = (w_1, \ldots, w_n)'$ be the vector of individual weights, let $\mathbf{w_r}$ be the piece of this vector with length n_r corresponding to individuals in group r, and let $w_{r\bullet} = \mathbf{1_r}'\mathbf{w_r}$, $r = 1, \ldots, k$. Let $W_D = \mathrm{diag}(\mathbf{w})$, $\mathbf{n}_w = (w_{1\bullet}, \ldots, w_{k\bullet})'$, $\mathbf{N}_w = \mathrm{diag}(\mathbf{n}_w)$, and $\mathbf{1}_k = (1, \overset{(k)}{\ldots}, 1)'$.

Proposition 3 *Assume that the $n \times n$ distance matrix Δ, with element (i, j) equal to d_{ij}, is Euclidean, that the $n \times q$ matrix X is an Euclidean configuration of Δ and that the i-th row of X has weight w_i. Let D be the matrix with element*

(i,j) *equal to* $d_{ij}^2/2$. *Then the* weighted total, weighted within-group *and the* weighted between-group *sums of squares are, respectively,*

$$\left.\begin{array}{rcl} T_w & = & (\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}'D\mathbf{w}(\mathbf{1}'\mathbf{w})^{-1} \\ W_w & = & \sum_{r=1}^{k}\left((\mathbf{1_r}'\mathbf{w_r})^{-1}\mathbf{w_r}'D_{rr}\mathbf{w_r}(\mathbf{1_r}'\mathbf{w_r})^{-1}\right)(\mathbf{1_r}'\mathbf{w_r})(\mathbf{1}'\mathbf{w})^{-1} \\ B_w & = & (\mathbf{1}_k'\mathbf{n}_w)^{-1}\mathbf{n}_w'D_B^w\mathbf{n}_w(\mathbf{1}_k'\mathbf{n}_w)^{-1} \end{array}\right\} \quad (11)$$

where D_B^w *is the* $k \times k$ *matrix with element* (r,s) *equal to* $(\delta_{rs}^w)^2/2 = (2f_{rs}^w - f_{rr}^w - f_{ss}^w)/2$, f_{rs}^w *being the element* (r,s) *of matrix*

$$F^w = \mathbf{N}_w^{-1}G'W_{D}DW_{D}G\mathbf{N}_w^{-1}.$$

Moreover it is verified that $T_w = W_w + B_w$, *even if the distance matrix* Δ *is not Euclidean and* T_w, W_w *and* B_w *are directly defined by equation (11).*

Observe that the value δ_{rs}^w is the distance between the weighted means of groups r and s. The ANOVA test statistic we propose is the weighted between-group variability B_w. The standard permutation procedure (permuting observations) to approach the null distribution of this statistic is analogous to what we introduced in Section 4.1 for unweighted samples, with the only difference that now we use equation (11) instead of (8). A permuted sample is obtained by randomly permuting the rows of G, whereas both matrix Δ_w and weight vector $\mathbf{w}$ remain unaltered.

In order to apply the permuting residual version of the permutation test, we need an additional result extending Proposition 1 to the weighted sample case. The following proposition states such a result. The proof is deferred to the Appendix.

Proposition 4 *Assume that the* $n \times n$ *distance matrix* Δ *is Euclidean, that the* $n \times q$ *matrix* X *is an Euclidean configuration of* Δ *and that the* i*-th row of* X *has weight* w_i. *Assume that* X_w^* *is the* $n \times q$ *matrix of residuals obtained by fitting an ANOVA model to the weighted rows of* X, *where the groups are defined by the* $n \times k$ *matrix* G. *Then the Euclidean distance* $d_{w,ij}^R$ *between rows* i *and* j *of* X_w^* *is given by*

$$(d_{w,ij}^R)^2 = 2e_{w,ij}^R - e_{w,ii}^R - e_{w,ij}^R,$$

where $e_{w,ij}^R$ *is the entry* (i,j) *of the* $n \times n$ *matrix* E_w^R *defined as*

$$E_w^R = \left(I_n - G\mathbf{N}_w^{-1}G'W_{D}\right)D\left(I_n - W_{D}G\mathbf{N}_w^{-1}G'\right).$$

The inter-residual distance matrix is Δ_w^R having element (i,j) equal to $d_{w,ij}^R$, and the associated matrix D_w^R has element (i,j) equal to $(1/2)(d_{w,ij}^R)^2$. As it happened in Proposition 1, Δ_w^R does not depend on X, and the associated inter-residual distance matrix can be defined as Δ_w^R even for a non-Euclidean Δ.

7 European regional income densities

In this section we analyze European regional income distributions taking into account the country to which each region belongs, as it was introduced in Section 1. Let $f_{ri}(x)$ be the relative equivalent disposable income (after taxes and benefits) density function of region i in country r, one of the $k = 15$ countries forming the European Union before May 2004. The total number of regions is $n = 88$. The null hypothesis (2) establishes that regional relative equivalent disposable income densities have the same mean value in each country. Another way of wording it is to say that under the null hypothesis there is no country effect in the observed variability of regional relative equivalent disposable income densities. Given that the true densities $f_{ri}(x)$ are not available, we are working with nonparametric estimates of them. In Sections 2 and 5 we offer arguments and simulation results backing the use of density estimates.

The used incomes are *disposable* (or *net*) because they are the result of applying taxes and social benefits to the household gross income. They are *equivalent incomes* in the sense that the household incomes are divided by the equivalent number of adults living in there, according to the modified OECD scale: one adult (person aged 14 or plus), plus one half of the additional number of adults, plus 0.3 times the number of children. Finally they are *relative* because in each region the observed equivalent incomes are divided by the regional median. So the data x corresponding to a household represents that this household has a equivalent disposable income equal to x times the median regional equivalent disposable income.

The information about the income distribution in European countries and regions comes from the 8th wave of the European Community Household Panel (ECHP-w8) corresponding to year 2001. We work with the households relative equivalent disposable income data. This information is summarized by the nonparametric estimation of the probability density function representing the relative income distribution for each region. We use the logarithmic transformation and kernel density estimation techniques to estimate income density functions (see Simonoff 1996, for instance). Observe that any household in the sample has a specific weight and this characteristic has to be taken into account in the estimation process.

A well known characteristic of income distributions is their marked right asymmetry, that classically leads to model them as log-normal random variables. From a nonparametric point of view, asymmetry implies that different degree of smoothness should be used in different levels of income (typically, more smoothness is needed in the right tail of the distribution, with low density, corresponding to high incomes). A way to apply different degree of smoothness to different zones is based on transformations. Data $x_1, \ldots, x_n$ are transformed by a known function g (we use the logarithm function) achieving that the transformed data are almost symmetric. Then usual kernel estimation is done in the transformed scale, and a change-of-variable formula is used to recover a density estimation in the original scale.

An additional problem should be noted here. The logarithm transformation

has to be applied to positive data, but not all the income data we are analyzing are positive. Then a positive constant c has to be added to each observation before taking logs. In our example we have chosen $c = 1$ (all the observed values x are greater than -1).

Finally, the nonparametric density estimator we are using is as follows:

$$\hat{f}(x) = \hat{f}_w^{lg}(\log(x+c))\frac{1}{x+c}, \text{ for incomes } x > -c,$$

where $\hat{f}_w^{lg}$ is the weighted kernel density estimator derived from $y_i = \log(x_i + c)$, $i = 1, \ldots, n$:

$$\hat{f}_w^{lg}(y) = \sum_{i=1}^{n} \frac{w_i}{\sum_j w_j} \frac{1}{h} K\left(\frac{y - y_i}{h}\right),$$

where the *kernel* K is a unimodal density function symmetric around 0 (a standard normal density, for instance), and h is the *bandwidth* or *smoothing parameter*.

We select the bandwidth using the *normal reference rule* for weighted data. It is well known that this rule is appropriate only when data are near normality (that is the case for $\log(x_i+c)$) and that it tends to over-smooth (to produce too high values for h). In order to correct the over-smoothing, a common practice is to multiply the proposed values by a positive constant lower than 1. In our case, we always take 2/3 times the values provided by the normal reference rule (Luxembourg, where the normal reference rule is respected, is an exception because otherwise the estimated density would be very bumpy). The constant 2/3 was chosen by visual inspection. The same applies for the constant c choice. In order to evaluate the sensitiveness of our result with respect to the constant 2/3, we have repeated the computation using other values: .5, .75, 1, 1.5. In all cases the results are very similar to those obtained with 2/3.

An alternative bandwidth choice rule, nowadays accepted as the most satisfactory one, is the *plug-in* method (Sheather and Jones 1991) that consists in replacing in the theoretical expression of optimal bandwidth h the unknown terms (because they depend on derivatives of the unknown density that is being estimated) by estimations (based on kernel estimation of the derivatives). The involved computations are far from being trivial, and in fact there are not available implementations covering the possibility of weighted samples. This is an additional reason that leads us to use the normal reference rule, jointly with the fact that data $\log(x_i + c)$ are almost normally distributed.

The density estimation has been done using the library `sm` (Bowman and Azzalini 2001) in the package `R` (R Development Core Team 2005), that implements the normal reference bandwidth choice rule and kernel estimation for weighted data. All densities are evaluated in 51 points evenly spaced from -1 to 5. The estimated regional densities are shown in Figure 5. They are grouped by countries. The number of regions in each country is indicated next to the country name.

Now we present the results of the no country effect tests. We have used the six test procedures listed in Section 5. Table 1 shows the test statistics observed

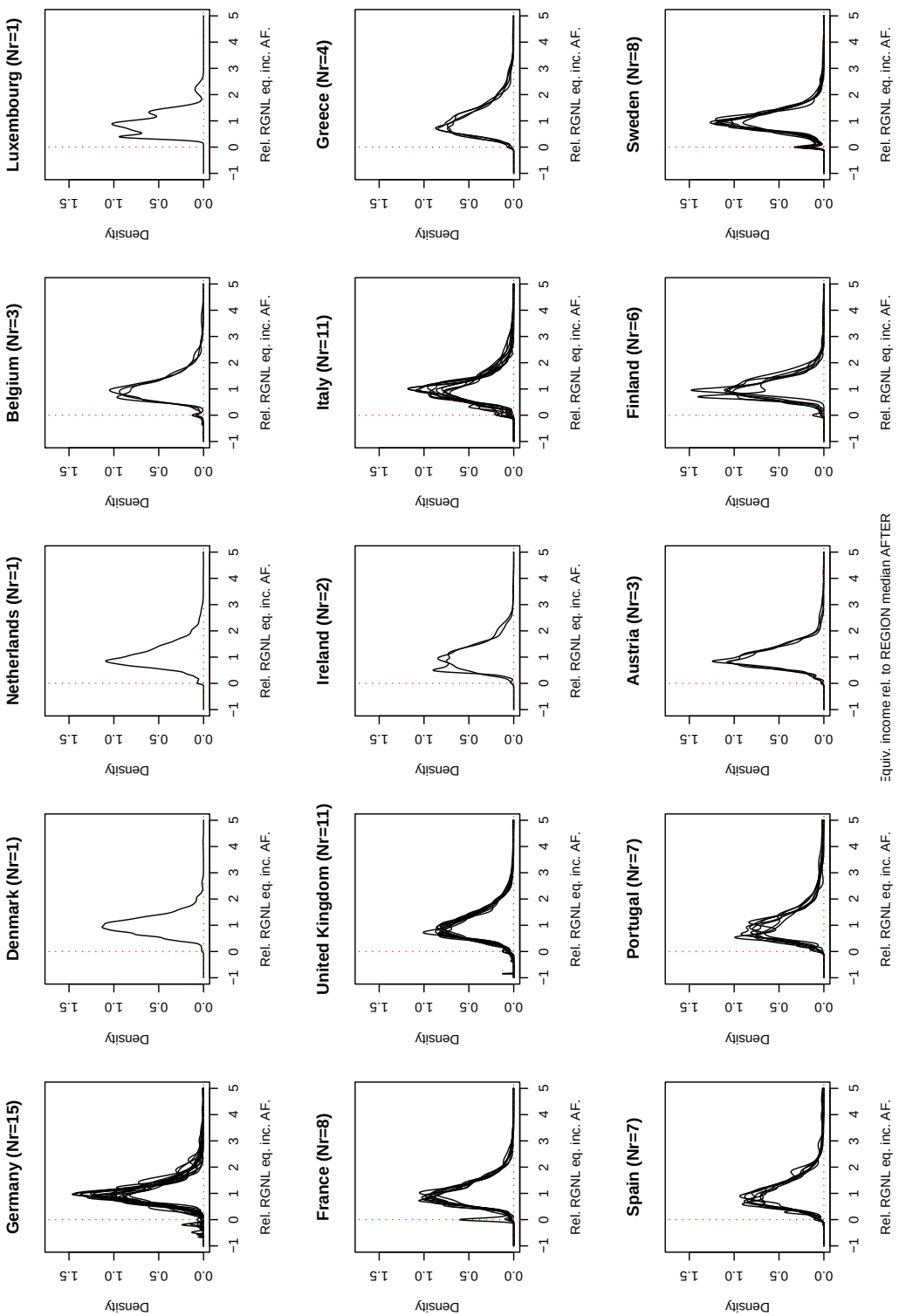

Figure 5: Density estimation for regional equivalent income after taxes and benefits, grouped by countries.

	Test statistics					
	V_n	$V_n^{\surd}$	T_F	T_{L_1}	T_{L_2}	$T_{L_2}^{\surd}$
Observed value	12.98	10.69	2.798	0.0134	0.0083	0.0068
p-values						
Permuting observations	$<.01$	$<.01$	$<.01$	$<.01$	$<.01$	$<.01$
Permuting residuals	$<.01$	$<.01$	$<.01$	$<.01$	$<.01$	$<.01$
CLT, homoscedastic	$<.01$	$<.01$				
CLT, heteroscedastic	$<.01$	$<.01$				

Table 1: Income density data. The observed values for the six test statistics are shown in the first row. The other rows contain the associated p-values, that have been computed using permutations procedures and, when available, simulated Gaussian processes (last two rows, first two columns).

values. The statistics null distributions have been approximated using all the available methods (Section 2) and taking into account the specific techniques for weighted data introduced in Section 6. Therefore, there are two ways (permuting observations or residuals) to compute p-values corresponding to four statistics: T_F (based on the F-ratio function), T_{L_1}, T_{L_2} and $T_{L_2}^{\surd}$ (based on distances). For the other two statistics (V_n and $V_n^{\surd}$, based on Cuevas, Febrero, and Fraiman 2004), the p-value can be computed using Proposition 2, that extends the Central Limit Theorem (CLT) result of Cuevas, Febrero, and Fraiman (2004) to the weighted case. We have considered the homoscedastic case as well as the heteroscedastic one. The number of permuted (or simulated) samples was always equal to 100. Table 1 shows the resulting p-values. In all cases p-values are lower than 0.01 (meaning that all the simulated statistics values were lower than the observed one). Based on these numbers, we conclude that the null hypothesis must be rejected.

In order to establish whether or not different ways of doing the functional ANOVA test give similar results (beyond the coincident reported p-values), in Figure 6 we look at the whole set of simulated reference statistics values. It can be seen that observed statistics values are always far from the simulated values. Moreover we observe that the observed values are considered less extreme when using the permuting observation mechanism to approximate the statistics null distribution than when permuting residuals or CLT based simulations were used.

Let us note that all the integrals involved in the test definitions have been done in $[a, b]$, for $a = -.25$ and $b = 5$. Other choices of a and b were also considered (the results are not shown here) and we conclude that only the test based on the F-ratio function is sensitive to the choice of a. When using $a = -1$ or $a = -.5$ numerical instability problems appeared leading to permuted samples with abnormally high values of T_F (the integral of the F-ratio function). A plausible explanation is that in these permuted samples for some $x \in [-1, -.25)$ the intra sample variability is zero but the between samples variability is positive. The choice of b was revealed as unimportant. We conclude that integrabil-

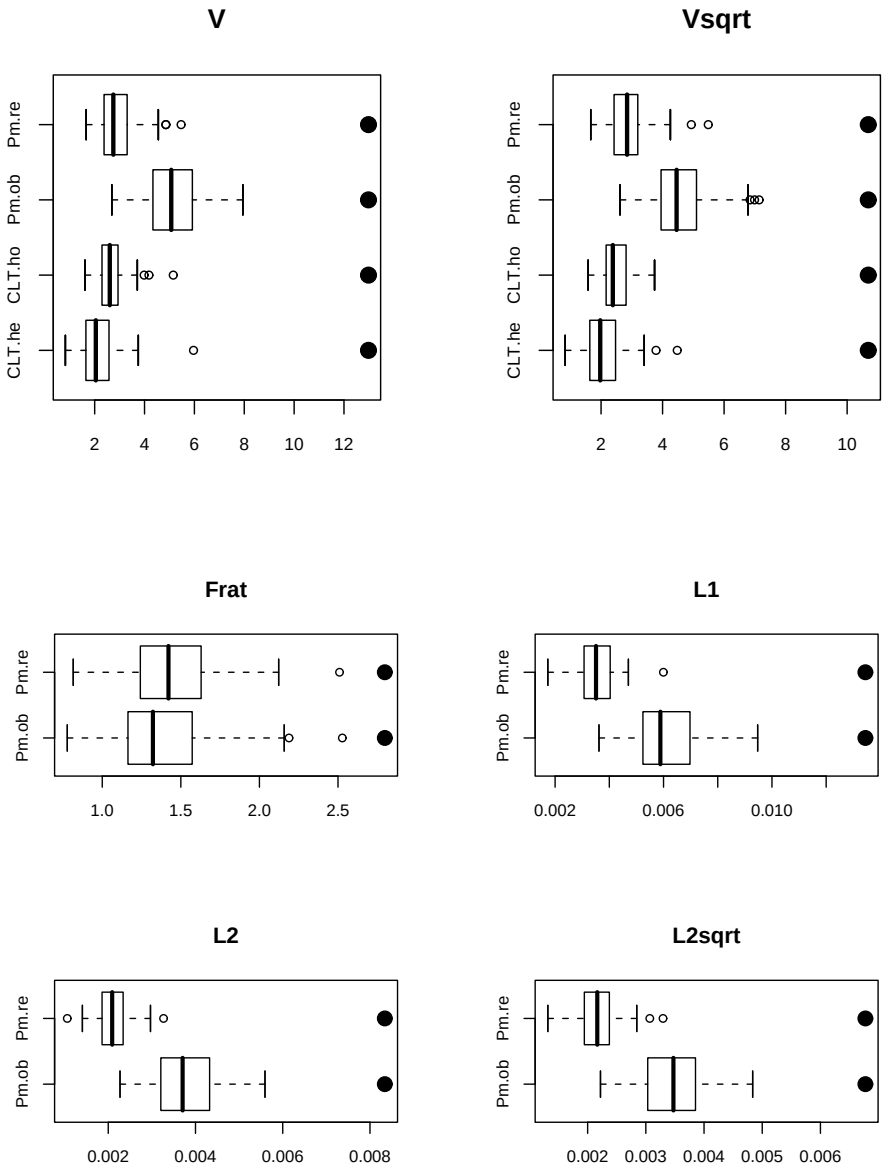

Figure 6: Income density data. Box-plots for 100 simulated test statistics values under the null hypothesis. The observed statistics values are marked as big solid circles.

ity problems of the F-ratio function (discussed in Section 3.1) are relevant in practice.

8 Conclusions

In this paper we have considered several ways to carry out the test of homogeneity between k-samples of functional data. We have focused on the case of density functions having different weights, but the presented theoretical results are also valid for more general functions. We have used methods based on between-cases distances as well as methods that were proposed for the ANOVA test with generic functional data. We have shown that the last ones can be used when data are density functions because there are equivalent test procedures based on distances.

Our simulation experiments suggests that tests based on the F-ratio function and those working with the squared root of density functions are preferable to other alternatives. When using the F-ratio test attention must be paid to eventual numerical instability. Regarding permutation tests, permuting observations gives accurate results when data follow both the null and the alternative hypothesis. Permuting residuals does not work properly. We have considered a real data example in European regional income distribution were all the tests considered here lead to accept that the country effect is important.

The main technical contributions of the paper are the derivations of formulas for doing the permuting residuals resampling method in distance-based ANOVA, and the generalization for weighted samples of two previous test methods (Gower and Krzanowski 1999 and Cuevas, Febrero, and Fraiman 2004).

Our experience working with methods based on distances allows us to suggest that statistical methods based on distances are ready to be used in a wide range of Functional Data Analysis problems, the present paper being an example.

Appendix: Proofs

Proof of Proposition 1.
This result is a particular case of Proposition 4, with equal weights for all the observations.

Proof of Proposition 2.
The proof reproduces that of Theorem 1 in Cuevas, Febrero, and Fraiman (2004), adapting it to weighted samples. It has to be taken into account that if $\bar{f}^w_{r\bullet} = \sum_{i=1}^{n_r} w_{ri} f_{ri}$ is the weighted mean function in the r-th sample, then according to the Central Limit Theorem for random variables in Hilbert spaces,

$$\alpha_{nr}(\bar{f}^w_{r\bullet} - m_r) \xrightarrow{d} Z_r,$$

where $Z_r = Z_r(x)$ is a Gaussian process with 0 mean and covariance function $K_r(x, y)$. The rest of the proof follows exactly as in Cuevas, Febrero, and

Fraiman (2004), with α_{nr}^2, $\sum_{s=1}^k \alpha_{ns}^2$, p_r^w and C_{rs}^w taking here the roles that n_r, n, p_r and C_{rs} have there, respectively.

Proof of Proposition 3.
This proof follows the lines of reasoning used in Gower and Krzanowski (1999) to establish their equation (7). Without lost of generality we can assume that the centroid of the rows of X is the origin of coordinates: $\sum_{i=1}^n w_i x_i = \mathbf{0}$, where x_i' is the i-th row of matrix X. Therefore, the weighted total sum of squares is

$$T_w = \frac{\sum_{i=1}^n w_i x_i' x_i}{\sum_{i=1}^n w_i}.$$

On the other hand,

$$(\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}'D\mathbf{w}(\mathbf{1}'\mathbf{w})^{-1} = \frac{1}{2(\sum_{i=1}^n w_i)^2}\sum_{i=1}^n\sum_{j=1}^n w_i w_j (x_i - x_j)'(x_i - x_j) =$$

$$\frac{1}{2(\sum_{i=1}^n w_i)^2}\sum_{i=1}^n\sum_{j=1}^n w_i w_j (x_i' x_i + x_j' x_j - 2x_i' x_j) =$$

$$\frac{1}{2(\sum_{i=1}^n w_i)^2} 2(\sum_{i=1}^n w_i)\sum_{i=1}^n w_i x_i' x_i = T_w.$$

Let us note that in the previous lines it can be observed the known relation between distance d_{ij} and the elements of the scalar product matrix $E = XX' = (e_{ij})$:

$$d_{ij}^2 = e_{ii} + e_{jj} - 2e_{ij}. \tag{12}$$

It is also known that X and D verify the equation

$$XX' = -(I_n - \mathbf{1}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}')D(I_n - \mathbf{w}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{1}') \tag{13}$$

(see equations (2) and (4) in Gower and Krzanowski 1999). We will be using both expressions below.

Let us now proceed with the weighted within-group sum of squares. Let T_w^r be the sum of squares within the r-th group, that is a weighted total sum of squares restricted to individuals in group r. Therefore we can use the above expression for the weighted total sum of squares:

$$T_w^r = (\mathbf{1_r}'\mathbf{w_r})^{-1}\mathbf{w_r}'D_{rr}\mathbf{w_r}(\mathbf{1_r}'\mathbf{w_r})^{-1}.$$

Then, the weighted within-group sum of squares is

$$W_w = \sum_{r=1}^k \frac{w_{r\bullet}}{\sum_{i=1}^n w_i} T_w^r = \sum_{r=1}^k \left((\mathbf{1_r}'\mathbf{w_r})^{-1}\mathbf{w_r}'D_{rr}\mathbf{w_r}(\mathbf{1_r}'\mathbf{w_r})^{-1}\right)(\mathbf{1_r}'\mathbf{w_r})(\mathbf{1}'\mathbf{w})^{-1}.$$

Now we deal with the weighted between-group sum of squares. The $k \times q$ matrix with the groups mean of X is

$$\bar{X} = \mathbf{N}_w^{-1} G' W_D X.$$

So, the corresponding scalar product matrix is

$$\bar{X}\bar{X}' = \mathbf{N}_w^{-1} G' W_D X X' W_D G \mathbf{N_w}^{-1}$$

and using equation (13),

$$\bar{X}\bar{X}' = \mathbf{N}_w^{-1} G' W_D \left(-(I_n - \mathbf{1}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}')D(I_n - \mathbf{w}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{1}')\right) W_D G \mathbf{N}_w^{-1}.$$

Moreover, if we take into account that $\mathbf{1}' = \mathbf{1}_\mathbf{k}' G'$ and that $\mathbf{N}_w = G' W_D G$ it follows that

$$\mathbf{1}' W_D G \mathbf{N}_w^{-1} = \mathbf{1}_\mathbf{k}' G' W_D G \mathbf{N}_w^{-1} = \mathbf{1}_\mathbf{k}'. \tag{14}$$

We conclude that

$$\bar{X}\bar{X}' = -\mathbf{N}_w^{-1} G' W_D D W_D G \mathbf{N}_w^{-1} + \mathbf{a}\mathbf{1}' + \mathbf{1}\mathbf{a}' + b\mathbf{1}\mathbf{1}',$$

where $\mathbf{a} = \mathbf{N}_w^{-1} G' W_D D (\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}$ is a k-vector and $b = (\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}' D \mathbf{w} (\mathbf{1}'\mathbf{w})^{-1}$ is a number. This expression is analogous to equation (5) in Gower and Krzanowski (1999) and, for the same reasons given there (the terms involving $\mathbf{a}$ and b cancel when applying equation (12) to the elements of $\bar{X}\bar{X}'$ to obtain inter-mean distances), the only relevant term is the first one, that is equal to matrix $-F^w$. Using the standard relation between distances and elements of the scalar product matrix, it follows that the distance between rows r and s of $\bar{X}$ is

$$\delta_{rs}^w = (2f_{rs}^w - f_{rr}^w - f_{ss}^w)^{1/2},$$

f_{rs}^w being the element (r, s) of matrix F^w. Taking into account that the weighted between-group sum of squares is also the weighted total sum of squares of the rows of $\bar{X}$, with weights in $\mathbf{n}_w$, it follows that

$$B_w = (\mathbf{1}_k' \mathbf{n}_w)^{-1} \mathbf{n}_w' D_B^w \mathbf{n}_w (\mathbf{1}_k' \mathbf{n}_w)^{-1}.$$

In order to prove that $T_w = W_w + B_w$, observe that B_w can also be expressed as

$$B_w = \frac{1}{(\sum_{i=1}^n w_i)^2} \frac{1}{2} \sum_{r=1}^k \sum_{s=1}^k w_{r\bullet} w_{s\bullet} (\delta_{rs}^w)^2$$

$$= \frac{1}{(\sum_{i=1}^n w_i)^2} \sum_{r=1}^k \sum_{s=1}^k w_{r\bullet} w_{s\bullet} f_{rs}^w - \frac{1}{\sum_{i=1}^n w_i} \sum_{r=1}^k w_{r\bullet} f_{rr}^w.$$

Moreover $F^w = \mathbf{N}_w^{-1} G' W_D D W_D G \mathbf{N}_w^{-1}$ implies that $f_{rs}^w = \mathbf{w_r}' D_{rs} \mathbf{w_s} / (w_{r\bullet} w_{s\bullet})$. Therefore,

$$B_w = \frac{1}{(\sum_{i=1}^n w_i)^2} \mathbf{w}' D \mathbf{w} - \frac{1}{\sum_{i=1}^n w_i} \sum_{r=1}^k w_{r\bullet} \mathbf{w_r}' D_{rr} \mathbf{w_r} = T_w - W_w.$$

Proof of Proposition 4.
Observe that

$$X_w^* = X - G\bar{X} = X - G\mathbf{N}_w^{-1} G' W_D X = \left(I_n - G\mathbf{N}_w^{-1} G' W_D\right) X.$$

Then, using (13),

$$X_w^*(X_w^*)' = \left(I_n - G\mathbf{N}_w^{-1}G'W_D\right) XX' \left(I_n - W_D G\mathbf{N}_w^{-1}G'\right) =$$

$$\left(I_n - G\mathbf{N}_w^{-1}G'W_D\right) \left[-(I_n - \mathbf{1}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{w})D(I_n - \mathbf{1}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{w})\right] \left(I_n - W_D G\mathbf{N}_w^{-1}G'\right) =$$

$$\left(I_n - G\mathbf{N}_w^{-1}G'W_D\right) D \left(I_n - W_D G\mathbf{N}_w^{-1}G'\right) + S,$$

where

$$S = \left(I_n - G\mathbf{N}_w^{-1}G'W_D\right) M \left(I_n - W_D G\mathbf{N}_w^{-1}G'\right)$$

and

$$M = D\mathbf{w}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{1}' + \mathbf{1}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}'D - \mathbf{1}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}'D\mathbf{w}(\mathbf{1}'\mathbf{w})^{-1}\mathbf{1}' =$$

$$\mathbf{a_M}\mathbf{1}' + \mathbf{1}\mathbf{a}_\mathbf{M}' + b_M\mathbf{1}\mathbf{1}',$$

where $\mathbf{a_M} = D\mathbf{w}(\mathbf{1}'\mathbf{w})^{-1}$ and $b_M = (\mathbf{1}'\mathbf{w})^{-1}\mathbf{w}'D\mathbf{w}(\mathbf{1}'\mathbf{w})^{-1}$.

It can be seen that $S = \mathbf{a}\mathbf{1}' + \mathbf{1}\mathbf{a}' + b\mathbf{1}\mathbf{1}'$ for some $\mathbf{a}$ and b. Effectively,

$$S = S_1 - S_2 - S_3 + S_4$$

$$= M - G\mathbf{N}_w^{-1}G'W_D M - MW_D G\mathbf{N}_w^{-1}G' + G\mathbf{N}_w^{-1}G'W_D MW_D G\mathbf{N}_w^{-1}G'.$$

Observe that

$$S_1 = M = \mathbf{a_M}\mathbf{1}' + \mathbf{1}\mathbf{a}_\mathbf{M}' + b_M\mathbf{1}\mathbf{1}',\ S_2 = \mathbf{a_2}\mathbf{1}' + \mathbf{1}\mathbf{a}_\mathbf{3}',\ S_3 = S_2' = \mathbf{a_3}\mathbf{1}' + \mathbf{1}\mathbf{a}_\mathbf{2}'$$

and then

$$S_2 + S_3 = \mathbf{a_{23}}\mathbf{1}' + \mathbf{1}\mathbf{a}_\mathbf{23}'$$

with $\mathbf{a_{23}} = \mathbf{a_2} + \mathbf{a_3}$. Moreover, if we take into account equation (14) it follows that

$$\mathbf{1}'W_D G\mathbf{N}_w^{-1}G = \mathbf{1}',\ G\mathbf{N}_w^{-1}G'W_D\mathbf{1} = \mathbf{1}.$$

Therefore

$$S_4 = G\mathbf{N}_w^{-1}G'W_D\left[\mathbf{a_M}\mathbf{1}' + \mathbf{1}\mathbf{a}_\mathbf{M}' + b_M\mathbf{1}\mathbf{1}'\right]W_D G\mathbf{N}_w^{-1}G' = \mathbf{a_4}\mathbf{1}' + \mathbf{1}\mathbf{a}_\mathbf{4}' + b_4\mathbf{1}\mathbf{1}',$$

where $\mathbf{a_4} = G\mathbf{N}_w^{-1}G'W_D\mathbf{a_M}$, and $b_4 = b_M$.

We have established that

$$X_w^*(X_w^*)' = E_w^R + \mathbf{a}\mathbf{1}' + \mathbf{1}\mathbf{a}' + b\mathbf{1}\mathbf{1}'.$$

Let v_{ij} be the (i,j)-th element of $X_w^*(X_w^*)'$. The general relationship between distances and scalar products, equation (12), is in this case

$$(d_{w,ij}^R)^2 = 2v_{ij} - v_{ii} - v_{jj} = 2e_{w,ij}^R - e_{w,ii}^R - e_{w,ij}^R,$$

where $e_{w,ij}^R$ is the entry (i,j) of matrix E_w^R. In other words, the terms $\mathbf{a}\mathbf{1}'$, $\mathbf{1}\mathbf{a}'$ and $b\mathbf{1}\mathbf{1}'$ do not contribute to the computation of distances $d_{w,ij}^R$.

References

Anderson, M. J. and J. Robinson (2001). Permutation tests for linear models. *Australian and New Zeland Journal of Statistics 43*, 75–88.

Arenas, C. and C. Cuadras (2002). Recent statistical methods based on distances. *Contributions to Science 2*, 183–191.

Bowman, A. and A. Azzalini (2001). *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press.

Cuadras, C. and C. Arenas (1990). A distance based regression model for prediction with mixed data. *Communications in Statistics A. Theory and Methods 19*, 2261–2279.

Cuevas, A., M. Febrero, and R. Fraiman (2004). An anova test for functional data. *Computational Statistics and Data Analysis 47*, 111–122.

Delicado, P. and M. del Río (1994). Bootstrapping the general linear hipothesys test. *Computational Statistics and Data Analysis 18*, 305–316.

Delicado, P. and M. Mercader (2006). The country factor on regional income distributions in Europe: A functional ANOVA approach. Technical report, Universitat Politècnica de Catalunya, http://hdl.handle.net/2117/413.

Devroye, L. and L. Györfi (1985). *Density estimation: The L_1-view*. New York: John Wiley and Sons.

Gower, J. C. and W. J. Krzanowski (1999). Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. *Applied Statistics 48*, 505–519.

Hall, P. and S. Wilson (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics 47*, 757–762.

Kneip, A. and K. Utikal (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association 96*, 519–542.

Manly, B. (1997). *Randomization, bootstrap and Monte Carlo methods in biology (Second edition)*. Chapman and Hall Ltd.

Mercader, M. and H. Levy (2004). The role of tax and transfers in reducing personal income inequality in Europes regions: Evidence from EUROMOD. Working Paper EM9/04, EUROMOD.

Mielke, P. W., K. J. Berry, P. J. Brockwell, and J. S. Williams (1981). A class of nonparametric tests based on multiresponse permutation procedures. *Biometrika 68*, 720–724.

Perarnau, J. (2005). Comparació de mostres amb dades funcionals (Comparing samples of functional data). Master's thesis, Facultat de Matemàtiques i Estadística. Universitat Politècnica de Catalunya, Barcelona. (In catalan).

R Development Core Team (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

Ramsay, J. and B. W. Silverman (1997). *Functional Data Analysis*. New York: Springer.

Ramsay, J. and B. W. Silverman (2002). *Applied Functional Data Analysis: Methods and Case Studies*. New York: Springer-Verlag.

Ramsay, J. and B. W. Silverman (2005). *Functional Data Analysis* (Second ed.). New York: Springer.

Sheather, S. J. and M. C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B, Methodological 53*, 683–690.

Simonoff, J. S. (1996). *Smoothing methods in statistics*. New York: Springer.

ter Braak, C. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In *Bootstrapping and related techniques (Trier, 1990)*, Volume 376 of *Lecture Notes in Econom. and Math. Systems*, pp. 79–85. Berlin: Springer.