An anova test for functional data

Antonio Cuevas\textsuperscript{a,}\textsuperscript{*}, Manuel Febrero\textsuperscript{b}, Ricardo Fraiman\textsuperscript{c}

\textsuperscript{a}Departamento de Matemáticas, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid 28049, Spain
\textsuperscript{b}Departamento de Estadística, Universidad de Santiago de Compostela, Santiago de Compostela, Spain
\textsuperscript{c}Centro de Matemática, Universidad de la República, Montevideo Uruguay

Received 1 April 2003; received in revised form 29 October 2003; accepted 29 October 2003

Abstract

Given $k$ independent samples of functional data the problem of testing the null hypothesis of equality of their respective mean functions is considered. So the setting is quite similar to that of the classical one-way anova model but the $k$ samples under study consist of functional data. A simple natural test for this problem is proposed. It can be seen as an asymptotic version of the well-known anova $F$-test. The asymptotic validity of the method is shown. A numerical Monte Carlo procedure is proposed to handle in practice the asymptotic distribution of the test statistic. A simulation study is included and a real-data example in experimental cardiology is considered in some detail.

\copyright 2003 Elsevier B.V. All rights reserved.

Keywords: Comparison of curves; Equality of functional means; Functional data; Functional one-way anova; Longitudinal data

1. Introduction

Functional data (also called longitudinal data in health and social sciences) arise in a number of scientific fields, associated with continuous-time monitoring processes whose final outputs are samples of functions. These situations arise with increasing frequency due to the current fast development of very precise real-time measurement instruments. The books by Ramsay and Silverman (1997, 2002) offer a broad perspective of the available methods and case studies in functional data. The current theoretical development in this field is clearly behind the needs suggested by a number of potential

\textsuperscript{*}Corresponding author.

E-mail address: antonio.cuevas@uam.es (A. Cuevas).
applications of interest for practitioners. As a consequence, a considerable effort is being made in order to adapt some standard statistical methods for functional data. This is the case, for example, of principal component analysis (Boente and Fraiman, 2000; Dauxois et al., 1982; Locantore et al., 1999; Pezzulli and Silverman, 1993; Silverman, 1996), discriminant analysis (Ferraty and Vieu, 2003) and regression (Cardot et al., 1999; Cuevas et al., 2002; Ferraty and Vieu, 2002). Other interesting references are Bosq (1991), Brumback and Rice (1998), Ramsay and Dalzell (1991), and Rice and Silverman (1991).

This paper is concerned with the classical one-way anova problem for functional data. Our research has been partially motivated by a problem in experimental cardiology (Ruiz-Meana et al., 2003) which is briefly analyzed in Section 4. Roughly speaking, the problem is to compare the mean levels of a variable, the mitochondrial calcium overload, measured in two samples (control and treatment) of cells. As the variable of interest is evaluated for each cell every 10 s, the resulting data can be seen as genuinely functional. The formal statement is as follows. Let $X_{ij}(t)$, $j = 1, \ldots, n_i$, $t \in [a, b]$ be $k$ independent samples of trajectories drawn from $L^2$-processes $X_i$, $i = 1, \ldots, k$, such that $E(X_i(t)) = m_i(t)$. With respect to the covariance function, we will consider both the classical “homoscedastic” case, where all the $X_i$ processes are assumed to have a common covariance function $K_i(s,t) = \text{Cov}(X_i(s), X_i(t)) := K(s, t)$ and the general “heteroscedastic” case where the $K_i(s,t)$ are not necessarily equal.

We want to test $H_0 : m_1 = \cdots = m_k$.

This problem has been already considered by Fan and Lin (1998) in the case when the sampling information is in a “discrete” format $X_{ij}(t)$, $t = 1, \ldots, T$. They propose a useful HANOVA (high dimensional anova) test, relying on wavelet thresholding techniques. A different approach is given by Muñoz-Maldonado et al. (2002) who use a method based on permutations and provide an interesting example in neurophysiology. Other related references are Kneip and Gasser (1992), Munk and Dette (1998), and Dette and Derbort (2001).

Although of course the functional data come often in practice in a discrete fashion, the discretization mechanism appears in many cases just as a practical device to approximate the data functions whose values are in fact available at points $t_j$ arbitrarily close. On the other hand, it is also conceivable that the increasing interest on functional data could lead to measurement devices whose outputs provide “true” functions, with “analytical” expressions (obtained maybe by nonparametric smoothing) instead of finite dimensional approximations. For these reasons our approach here is purely functional in the sense that our test statistic is a functional of the sample trajectories and its motivation is also given in functional terms. However, as we will see below, discretization will be in fact used in both our simulation study and in the case study with real data. So our functional approach should be seen just as an asymptotic procedure based on a simple and direct motivation leading to a method where discretization appears only as a final approximating device. This step could be avoided in some cases if the data come expressed as true functions. The simulation study in Section 4 suggests that our approach (inspired by the classical $F$-test) could perform quite well in practice.
2. The proposed test

2.1. The test statistic and its asymptotic behavior

In a first approximation to the problem, in the homoscedastic case, we might consider the possibility of using a direct analog of the classical $F$ statistic for the one-way anova problem as given by

$$F_n = \frac{\sum_{i=1}^{k} n_i \|\bar{X}_i - \bar{X}._i\|^2/(k - 1)}{\sum_{i,j} \|X_{ij} - X_{i.}\|^2/(n - k)},$$

where

$$\bar{X}_i(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}(t), \quad \bar{X}._i = \frac{1}{n} \sum_{i=1}^{k} n_i \bar{X}_i(t), \quad n = \sum_{i=1}^{k} n_i,$$

and $\|\cdot\|$ stands for the usual $L^2$ norm,

$$\|x\| = \left( \int_a^b x^2(t) \, dt \right)^{1/2}.$$

Likewise the classical anova $F$-statistic, we can see here that whereas the numerator accounts for the “external” variability between the different samples, the denominator measures the “internal” variability within the samples. In fact, expression (1) would coincide exactly with that the classical $F$-test if the functions $X_{ij}$ are replaced by numerical data $x_{ij}$ and the squared norm $\|X\|^2$ is replaced by $x^2$.

Of course, the idea would be to reject $H_0$ at a prescribed level $\alpha$, when $F_n > F_{n,\alpha}$, where $F_{n,\alpha}$ is such that $P_{H_0} \{F_n > F_{n,\alpha} \} = \alpha$. Unfortunately, the close analogy with the classical anova test for univariate random variables must stop here since the distribution of (1) under $H_0$, required to obtain $F_{n,\alpha}$, does not seem easy to derive, even if the $X_i$ are assumed to be Gaussian. However, the $F$-test is just one among many possible ways of formalizing a very simple idea: $H_0$ should be rejected whenever the “external” variability between groups, as measured by the differences between their “sample means” is large enough at a prescribed level. The universal use of the $F$ statistic in the univariate setup is due to the fact that the exact sampling distribution under $H_0$ of $F_n$ is known (assuming some standard conditions) so that one gets an exact test for each sample size $n$. In the functional setup, the task of finding such an exact test seems hopeless. However, if we are willing to use an asymptotic test, thus renouncing to exact significance levels and requiring large samples, the structure of the classical $F$-statistic suggests an alternative which could be easily adapted to the functional setup.

More specifically, let us recall that the denominator $Q_2$ of the anova $F$-statistic can be expressed as $Q_2 = \sum_{i}(n_i - 1)s_i^2/(n - k)$, $s_i^2$ being the (quasi-) variance of the $i$th sample. Under the usual homoscedasticity assumption $s_i^2$ tends (a.s.) to the common variance $\sigma^2$ as $n_i$ tends to infinity. So $Q_2$ can be seen as a consistent estimator of $\sigma^2$. This entails that asymptotically we could replace $Q_2$ by $\sigma^2$. This parameter could be incorporated to the numerator $Q_1$ so that it is only necessary to calculate the asymptotic distribution of $Q_1/\sigma^2$ and replace $\sigma^2$ by an estimator in that distribution. This is
essentially the same situation arising when the standard exact $t$-test is replaced by the asymptotic normal alternative (which requires also to estimate the variance). Note also that homoscedasticity is not essential here. A similar reasoning could be used in the heteroscedastic case.

Now, coming back to the functional framework and bearing in mind this motivation, we could think of using a test based on the numerator of the statistic (1) which, up to the known constant $(k - 1)$, is

$$T_n = \sum_{i=1}^{k} n_i \| \bar{X}_i - \bar{X} \|^2.$$  \hfill (2)

An equivalent way to measure the “between groups” variability (which we prefer for technical reasons) is given by

$$V_n = \sum_{i<j} n_i \| \bar{X}_i - \bar{X}_j \|^2.$$  \hfill (3)

For example in the balanced case when the $n_i$ are equal, it can be easily seen that $T_n$ and $V_n$ only differ by a multiplicative constant. The weighting factor $n_i$ in (3) is, in a way, arbitrary and introduces some asymmetry as it could be replaced by $n_j$. However, as we will see, the limit is in both cases the same, up to known multiplicative constants. By the way, let us notice that the use of a test based on $T_n$ or $V_n$ with the asymptotic motivation given by Theorem 1 below, will allow us to get rid of the homoscedasticity assumption, typically present in the classical univariate anova models. Of course, the price will be to use an approximate (asymptotic test). The possibly different covariance functions $K_i(s, t)$ will of course appear in the asymptotic distribution, but each $K_i$ can be consistently estimated from the $i$th sample $X_{ij}(t)$, $j = 1, \ldots, n_i$.

**Theorem 1.** Assume that $n_i, n \to \infty$ in such a way that $n_i/n \to p_i > 0$ for $i = 1, \ldots, k$. Assume that we have observations $X_{ij}(t)$, with $j = 1, \ldots, n_i$, corresponding to $k$ independent samples of sizes $n_i$ from $k$ $L^2$-processes with mean 0 and covariance functions $K_i(s, t) = \text{Cov}(X_i(s), X_i(t))$. Then, the asymptotic distribution of $V_n$ under $H_0$ coincides with that of the statistic

$$V := \sum_{i<j} \| Z_i(t) - C_{ij} Z_j(t) \|^2,$$  \hfill (4)

where $C_{ij} = (p_i/p_j)^{1/2}$ and $Z_1(t), \ldots, Z_k(t)$ are independent gaussian processes with mean 0 and covariance functions $K_i(s, t)$.

**Proof.** Let us first assume that $n_i/n = p_i$. Then, the result is a direct consequence of the Central Limit Theorem for random variables taking values in a Hilbert space (see, e.g., Laha and Rohatgi, 1979, p. 474). According to this theorem we have, under $H_0: m_1(t) = \cdots = m_k(t) := m(t)$, the weak convergence (in the space of probability measures on $L^2 \times L^2$)

$$(\sqrt{n_1}\bar{X}_1(t) - m), \ldots, \sqrt{n_k}\bar{X}_k(t) - m) \overset{d}{\to} (Z_1(t), \ldots, Z_k(t)),$$
where $Z_i(t)$ has the same covariance function as $X_i(t)$. Now, the conclusion of the theorem follows from the continuous mapping theorem (see, e.g., Billingsley, 1968), since
\[
V_n = g(\sqrt{n_1}(\bar{X}_1 - m), \ldots, \sqrt{n_k}(\bar{X}_k - m)),
\]
where $g$ is real continuous a function, defined on $L^2 \times \mathbb{R}^k$ by
\[
g(x_1(t), \ldots, x_k(t)) := \sum_{i<j} \|x_i(t) - C_{ij}x_j(t)\|^2.
\]
Finally, the general case $n_i/n \to p_i$ follows from the well-known Slutsky’s theorem, applied to the case of random variables taking values in a metric space, see, e.g., Billingsley (1968, p. 34).

It is not difficult to see that the natural test based on the statistic $V_n$ would be consistent as $V_n$ would tend to infinity if $H_0$ is not true.

2.2. Practical implementation. Numerical aspects

Provided that the $n_i$ are large enough, hypothesis $H_0$ is rejected, at a level $\alpha$, whenever $V_n > V_\alpha$, where $P_{H_0}\{V > V_\alpha\} = \alpha$. In order to approximately evaluate $V_\alpha$ let us note that the distribution of $V$ under $H_0$ is known whenever the covariance functions $K_i(s,t)$ are. This is not usually the case, but these functions can be estimated under $H_0$ by
\[
\hat{K}_i(s,t) = \frac{1}{n_i-1} \sum_{j=1}^{n_i} (X_{ij}(s) - \bar{X}_i(s))(X_{ij}(t) - \bar{X}_i(t)).
\]
If the homoscedastic assumption (H) looks reasonable, the natural estimate of the common covariance function would be
\[
\hat{K}(s,t) = \frac{1}{n-1} \sum_{i,j} (X_{ij}(s) - \bar{X}_i(s))(X_{ij}(t) - \bar{X}_j(t)).
\]
In any case, if the covariance functions are replaced by estimators, the distribution of $V$ is not easy to handle in practice as it is a complicated function of $k$ Gaussian processes.

Thus an asymptotic Monte Carlo procedure is implemented as follows: a large number $N$ (in the simulations below $N = 2000$) of discretized artificial trajectories are simulated, for each process $Z_i(t)$ appearing in the limit statistic $V$ defined in (4). In other words, we draw for each $i = 1, \ldots, k$, $N$ iid observations
\[
Z_{il} = (Z_{il}(t_1), \ldots, Z_{il}(t_m)), \quad l = 1, \ldots, N
\]
from a $m$-dimensional gaussian random variable with mean 0 and covariance matrix $(\hat{K}_i(t_p,t_q))_{1 \leq p, q \leq m}$. This amounts to approximate the continuous trajectories of $Z_i(t)$ by step versions evaluated in a grid of values $a \leq t_1 < \cdots < t_m \leq b$. The functional $L^2$-distances $\|Z_i(t) - C_{ij}Z_j(t)\|^2$ are then approximated by the $\mathbb{R}^m$-Euclidean distances.
∥Z_{il} - C_{ij}Z_{jl}^*∥^2. More precisely, we construct artificial replications \( \tilde{V}_l \), \( l = 1, \ldots, N \) of \( V \) defined by

\[
\tilde{V}_l = \sum_{i < j} \|Z_{il}^* - C_{ij}Z_{jl}^*\|_2.
\]

The distribution of \( V \) under \( H_0 \), and in particular the value \( V/VT \), is finally approximated from the empirical distribution corresponding to the artificial sample \( \tilde{V}_1, \ldots, \tilde{V}_N \). This procedure can be considered as a sort of asymptotic parametric bootstrap. Let us also observe that, given the structure of \( U \)-statistic of \( V_n \), a direct result of bootstrap validity for \( V_n \) could be obtained along the lines of Arcones and Giné (1992).

3. A simulation study

We have considered an artificial example with \([a, b] = [0, 1]\) and 3 levels (that is, \( k = 3 \)) in four cases

- (M1) \( m_i(t) = t(1 - t), \ i = 1, 2, 3 \),
- (M2) \( m_i(t) = t^i(1 - t)^{6-i}, \ i = 1, 2, 3 \),
- (M3) \( m_i(t) = t^{i/5}(1 - t)^{6-i/5}, \ i = 1, 2, 3 \),
- (M4) \( m_i(t) = 1 + i/50, \ i = 1, 2, 3 \).

Case M1 corresponds to a situation where \( H_0 \) is true; M2 and M3 provide examples, with \( H_0 \) false, of monotone functions with different increase patterns. Whereas in M2 the \( m_i \) are quite separated, in M3 the differences are less apparent (so the testing problem should be harder). Finally, M4 is an example where the functional approach is unnecessarily complicated as the functions are in fact constant. In this case a classical anova \( F \)-test for real variables would be more appropriate. The reason for including such a degenerate case is to check the performance of our functional anova test in a situation where an optimal (at least in the independent case) alternative is available. Of course the comparison between both cases, one-dimensional and functional, is not completely fair as the former incorporates the additional information that the \( m_i \) are constant.

The test has been applied 5600 times for each of these cases, under different conditions, corresponding to subcases to be specified below. More precisely, 5600 independent samples (with \( n_i = 10 \) for \( i = 1, \ldots, 3 \)) have been drawn, for each choice M1, \ldots, M4 of the mean functions, under the model

\[
X_{ij}(t) = m_i(t) + e_{ij}(t), \quad j = 1, \ldots, 10.
\]

In fact, the processes have been generated in discretized versions \( X_{ij}(t_r) \), for \( r = 1, \ldots, 25 \), where the values \( t_r \) have been chosen equispaced in the interval [0, 1].

The 5600 samples of every case have been divided into two subgroups of 2800 samples, according to the chosen structure for the error processes \( e_{ij}(t) \). In the first subgroup the \( e_{ij}(t_r) \) are iid random variables \( N(0, \sigma) \). Therefore, in this case the vector \( (e_{ij}(t_1), \ldots, e_{ij}(t_{25})) \) is a “white noise” that can be seen as a discrete approximation
to the (physically unfeasible) continuous-time white noise. This is in fact a quite unfavorable situation, in the sense that the “contamination” effect given by the process $e_{ij}(t)$ is particularly severe.

In the second subgroup, the $e_{ij}(t)$ is a standard brownian process with dispersion parameter $\sigma$. We thus have 8 subgroups, each one with 700 realizations of the test, corresponding to all possible choices M1/M2/M3/M4, independent/Brownian. Finally, each subgroup is in turn divided into 7 additional subgroups of 100 runs, with different values of the dispersion parameter. These values are, for the brownian case, $\sigma_1 = 0.2$, $\sigma_2 = 1$, $\sigma_3 = 1.8$, $\sigma_4 = 2.6$ $\sigma_5 = 3.4$, $\sigma_6 = 4.2$ and $\sigma_7 = 5$. In the case of independent errors the values of the dispersion parameter are $\sigma^*_k = \sigma_k/25$, for $k = 1, \ldots, 7$. Note that in the brownian case the effect of the error process, with a given value of $\sigma$, is smaller than that obtained in the analogous independent case (with the same $\sigma$), because of the dependence structure. For this reason the values $\sigma^*_k$ and $\sigma_k$ have been chosen in order to get approximately comparable situations in both cases.

In each of these 100 runs we have evaluated (by using 2000 Monte Carlo replications) the corresponding $p$-value provided by our test as well as the probability of rejecting the null hypothesis $H_0 : m_1 = m_2 = m_3$ at a significance level $\alpha = 0.05$. In summary, the performance of our procedure has been evaluated in the 56 different situations obtained by combining the underlying models M1, M2, M3, M4 the error structure (independent or brownian) and the value of the parameter $\sigma$.

In Table 1 we indicate the average (over the 100 runs) $p$-values obtained in the cases of independent errors. We also give, as a more direct indication of the test performance, the observed acceptance proportions (at a significance level 0.05) under every considered model. In the model M4, where the functions are constant, we have performed the classical $F$-test for real data (which is meaningful here as no proper function is involved), together with our procedure. The results for the $F$-test are given in the last two rows of the table.

### Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>$p$-value</th>
<th>$\sigma^*_1$</th>
<th>$\sigma^*_2$</th>
<th>$\sigma^*_3$</th>
<th>$\sigma^*_4$</th>
<th>$\sigma^*_5$</th>
<th>$\sigma^*_6$</th>
<th>$\sigma^*_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>$p$-value</td>
<td>0.463</td>
<td>0.543</td>
<td>0.449</td>
<td>0.505</td>
<td>0.496</td>
<td>0.472</td>
<td>0.465</td>
</tr>
<tr>
<td></td>
<td>accept. 0.05</td>
<td>0.980</td>
<td>0.980</td>
<td>0.970</td>
<td>0.970</td>
<td>0.980</td>
<td>0.960</td>
<td>0.960</td>
</tr>
<tr>
<td>M2</td>
<td>$p$-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.048</td>
<td>0.152</td>
<td>0.215</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>accept. 0.05</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.210</td>
<td>0.590</td>
<td>0.810</td>
<td>0.820</td>
</tr>
<tr>
<td>M3</td>
<td>$p$-value</td>
<td>0.000</td>
<td>0.171</td>
<td>0.338</td>
<td>0.412</td>
<td>0.479</td>
<td>0.448</td>
<td>0.417</td>
</tr>
<tr>
<td></td>
<td>accept. 0.05</td>
<td>0.000</td>
<td>0.770</td>
<td>0.950</td>
<td>0.970</td>
<td>0.950</td>
<td>0.990</td>
<td>0.970</td>
</tr>
<tr>
<td>M4</td>
<td>$p$-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>0.062</td>
<td>0.137</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>accept. 0.05</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.050</td>
<td>0.370</td>
<td>0.620</td>
<td>0.770</td>
</tr>
<tr>
<td></td>
<td>$p$-value for $F$-test</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>accept. 0.05, $F$-test</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Table 2
Average $p$-values and acceptance proportions (at level 0.05) over 100 runs in the case of brownian errors

<table>
<thead>
<tr>
<th>Model</th>
<th>$p$-value</th>
<th>accept. 0.05</th>
<th>$p$-value for $F$-test</th>
<th>accept. 0.05, $F$-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.403</td>
<td>0.980</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.476</td>
<td>0.970</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.555</td>
<td>0.980</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.475</td>
<td>0.980</td>
<td>0.042</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.506</td>
<td>1.000</td>
<td>0.136</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.494</td>
<td>1.000</td>
<td>0.219</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.438</td>
<td>0.980</td>
<td>0.279</td>
<td>0.000</td>
</tr>
<tr>
<td>M2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.250</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.042</td>
<td>0.000</td>
<td>0.620</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.136</td>
<td>0.000</td>
<td>0.800</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.219</td>
<td>0.000</td>
<td>0.840</td>
<td>0.000</td>
</tr>
<tr>
<td>M3</td>
<td>0.000</td>
<td>0.164</td>
<td>0.325</td>
<td>0.780</td>
</tr>
<tr>
<td></td>
<td>0.164</td>
<td>0.325</td>
<td>0.404</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>0.404</td>
<td>0.950</td>
<td>0.443</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>0.443</td>
<td>0.970</td>
<td>0.451</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>0.435</td>
<td>0.980</td>
<td>0.435</td>
<td>0.990</td>
</tr>
<tr>
<td>M4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.014</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.014</td>
<td>0.000</td>
<td>0.068</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.068</td>
<td>0.000</td>
<td>0.161</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.161</td>
<td>0.000</td>
<td>0.203</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.203</td>
<td>0.000</td>
<td>0.740</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 2 is completely analogous for the case of brownian errors.

In our view, the correct interpretation of any simulation study with functional data must take into account the intrinsic difficulty of this field. It would be naive to draw firm conclusions from the Monte Carlo outputs when there are infinitely many, essentially different, possibilities for the model choice (given in this case by the functions $m_i$ and the error processes $e_{ij}(t)$). The simulation experiments have however some interest in order to show the practical viability of the proposed method by checking its performance in a few standard situations. From this limited perspective, Tables 1 and 2 are hopefully self-explanatory and reassuring about the behavior of our test:

(a) In the model M1 (where $H_0$ is true) the test provides satisfactory results in all considered cases. The average $p$-values range between 0.40 and 0.55 and the 0.05 estimated acceptance probabilities exceed 0.95 in all cases.

(b) In the models M2 and M3 (where $H_0$ is false) the simulation outputs give an idea of the test power for a very small sample. In general terms the results are satisfactory, getting worse as $\sigma$ increases. Of course, the test does a better job under M2 (where the functions $m_i$ are well apart) than in the “difficult” model M3.

(c) The outputs for M4 (where the $m_i$ are constant) are particularly interesting as in this case we have a reference (the optimal parametric $F$-test) to assess our results. It can be seen that although our functional anova test is obviously defeated, it performs quite well for small and moderate values of the error parameter $\sigma$.

(d) Another interesting point has to do with sample sizes. As our method relies on asymptotics plus resampling, some practical insight is needed about what is the real meaning of “asymptotic” in this setup. The results obtained for $n_i=10$ suggest a reasonable behavior for moderate sample sizes.

(e) The variability of the simulation outputs depends strongly on the error parameters. It is nearly null for the $\sigma_1$, $\sigma_1^*$, small for $\sigma_2$ and $\sigma_2^*$ (leading typically to more
than 95% of correct decisions) and relatively large in the remaining cases. In this respect note that the values of the standard deviations of the $p$-values are not particularly interesting here. We are rather concerned with the frequency of a correct decision. Thus the 0.05-acceptance frequencies given in Tables 1 and 2 can also be seen as an indirect more useful assessment of variability.

(f) A more comprehensive simulation study could be made by checking other discretization patterns, that is, by using different values of $T$ in the simulation of the discretized versions $X_i(t_1), \ldots, X_i(t_T)$ of the sample trajectories (we have taken $T = 25$). Also other choices could be considered for the number of artificial replications and the number of runs for each considered subcase (2000 and 100, respectively, in our study).

4. A real-data example in experimental cardiology

During myocardial ischemia, alterations in the energy metabolism cause disturbances of the ionic homeostasis that ultimately lead to cell death. These alterations are closely linked to the preservation of $H^+$ gradient across mitochondria membrane. By biochemical reasons, beyond the scope of this paper (see, Piper et al. (1996), Ruiz-Meana et al. (2000, 2003)) an indirect measurement of this ionic gradient is the *mitochondrial calcium overload*, a measure of the mitochondrial calcium ion $Ca^{2+}$ levels. Higher levels indicate a better protection against the ischemia process as they suggest a higher mitochondrial capacity to buffer the toxic calcium present in the cytosol (the medium surrounding the cellular organelle). The mitochondrial calcium overload has been monitored in isolated mouse cardiac cells. In each cell, measurements were taken every 10 s (for 1 h) using a fluorescence imaging system (QuantiCell2000, Visitech, UK).

The purpose of the experiment is to check the effect of “Cariporide”, a selective blocker of $Na^+/H^+$ exchange that (according to the experimenters’ conjecture) should prevent $H^+$ influx into the mitochondria, an effect that results in a preservation of the energetic status of the cell. The potential consequence of this effect is expected to be an increase in the total $Ca^{2+}$ load within the mitochondria. Thus, the mitochondrial calcium overload was measured in two groups (control and treatment) with 45 cells each; in the treatment group the cells received Cariporide (7 muM). See Ruiz-Meana et al. (2003) for more details.

Fig. 1 shows the average curves for the control group and the treatment group as well as the global average curve.

Our test was used to check whether or not the observed differences between the average calcium overload curves are statistically significative. The results indicated a strong evidence ($p$-value = 0.005) in favor of the hypothesis that both curves are actually different.

In order to get a more precise control of the medium surrounding the mitochondria (thus avoiding the distorting effect of the cell membrane), the experiment has been repeated with “permeabilized” cells whose membrane has been removed. The resulting average curves are shown in Fig. 2.
Fig. 1. Average calcium overload curves (case of intact cells).

Fig. 2. Average calcium overload curves (case of permeabilized cells).
In this case the scale is different, due to technical reasons associated with the measurement methods used in both cases. Another more relevant difference is the structure of the curves during the initial period (say, the first three minutes) of the experiment. In the case of non-permeabilized cells, the curves show a erratic behavior (not very relevant in the experiment context) during this period. In the experiment with permeabilized cells this phenomenon is also present but it is less noticeable. In any case, the test provides again a strong statistical evidence (the \( p \)-value is about 0.004) against the null hypothesis of equality of the average curves in both groups.

We have finally tested the equality of the derivatives of both mean curves in the case of permeabilized cells. This would amount to check that both curves are essentially the same except for an additive constant. The resulting \( p \)-value has been 0.9860. Hence there is no statistical evidence against the hypothesis that both derivatives coincide. The analogous test for the case of intact cells is less interesting due to erratic irrelevant oscillations that typically arise at the beginning of the experiment.

Other alternative models could be considered. For example, the analysis could be made in terms of median functions (or even “mode functions”), although the required theory looks more involved, and so far less developed, that required for mean functions; see, however, Cadre (2001), Fraiman and Muniz (2001).

Acknowledgements

We are most grateful to David García-Dorado and Marisol Ruiz-Meana (Servicio de Cardiología, Hospital Vall d’Hebron, Barcelona) for providing us with the calcium data and explaining us their meaning.

This work has been partially supported by Grants BFM2001-0169 (A. Cuevas) and BFM2002-03213 (M. Febrero) from the Spanish Ministry of Science and Technology.

The useful and constructive comments from two anonymous referees are gratefully acknowledged.

References