Chapter 11
Permutation tests in the two-sample problem for functional data

Alejandra Cabaña, Ana M. Estrada, Jairo Peña and Adolfo J. Quiroz

Abstract We propose two kind of permutation tests for the two sample problem for functional data. One is based on nearest neighbours and the other based on functional depths.

11.1 Introduction

Let $X_1(t), \cdots, X_m(t)$ denote an i.i.d. sample of real valued curves defined on some interval J. Let $\mathscr{L}(X)$ be the common probability law of these curves. Likewise, let $Y_1(t), \cdots, Y_n(t)$, be another i.i.d. sample of curves, independent of the X sample and also defined on J, with probability law $\mathscr{L}(Y)$.

We want to test the null hypothesis, $H_0 : \mathscr{L}(X) = \mathscr{L}(Y)$ against the general alternative $\mathscr{L}(X) \neq \mathscr{L}(Y)$.

We discuss three different permutation tests: a functional Schilling test, Wilcoxon type test, and another test based on depths, that uses meta analysis ideas to assess significance. We compare their performance with the classical test by Kokoszka and Horváth, based on the principal components of the pooled covariance operator of the two samples in a simulated experiment and with real data.

Alejandra Cabaña (✉)
Universidad Autónoma de Barcelona, Barcelona, Spain, e-mail: acabana@mat.uab.cat

Ana M. Estrada, Jairo Peña and Adolfo J. Quiroz
Universidad de Los Andes, Bogotá, Colombia e-mail: am.estrada213@uniandes.edu.co, ji.pena45@uniandes.edu.co, aj.quiroz1079@uniandes.edu.co

G. Aneiros et al. (eds.), *Functional Statistics and Related Fields*,
Contributions to Statistics, DOI 10.1007/978-3-319-55846-2_11

11.2 Schilling's type test

This is an adaption to functional data of the k -nearest neighbours multivariate two-sample test of Schilling's tests based on nearest neighbors [10].

Let $N = m+n$ and $Z_1, \ldots, Z_N$ be the combined sample obtained by concatenating the X and Y samples.

Define the indicator function $I_i(r) = 1$ if Z_i and its r-nn belong to the same sample and else, $I_i(r) = 0$. Nearest neighbours are based on L^2 distance, and with probability 1 the are uniquely defined. In case of ties, we would decide at random.

In practice, if the functions have been registered in a common grid, say $0 = t_0 < t_1 < \ldots t_L = T$, a reasonable approximation to the distance between functions Z_i and Z_j is

$$d_{i,j} = \sum_{l=1}^{L} \Delta_l (Z_i(t_l) - Z_j(t_l))^2, \text{ where } \Delta_l = t_l - t_{l-1}$$

If the grid used is equally spaced, Δ_l can be omitted and the curves can be treated as points in R^L in order to compute faster the k-nearest-neighbours. When no common grid is available, represent the functions in the joint sample in terms of local polynomials, or some other basis functions, and the k-nearest-neighbours are identified by a quadratic algorithm (in the joint sample size N).

Define the statistic:

$$T_{N,k} = \frac{1}{Nk} \sum_{i=1}^{N} \sum_{r=1}^{k} I_i(r)$$

Under H_0 we expect $T_{N,k}$ to be small.

Observe that the expected value $\mathbf{E}\, T_{N,k} = \mathrm{e}\, I_i(r) = \frac{m(m-1)+n(n-1)}{N(N-1)}$ while the variance depends on the amount of pairs of points that are mutual neighbours and the amount of pairs that share a common neighbour.

Under the null hypothesis of equal distribution of $\{X_m(t)\}$ and $\{Y_n(t)\}$, permutations on the labelling of the Z_i do not alter the distribution of $T_{N,k}$ and hence can be computed with a standard permutation procedure.

Algorithm

1. For the concatenation of the samples Z, keeping the natural ordering, compute the $m \times n$ matrix of distances between its elements.
2. Assume the number of neighbours is fixed to k. Build a $N \times k$ matrix that in the i-th row contains the indices of the k nearest elements to the curve Z_i.
3. In order to compute the statistic it is enough to count how many of the indices in each of the first m rows are equal or less than m, (i.e. are originally X) and how many of the indices corresponding to $m+1, \ldots, N$ are greater than m.
4. Obtain the distribution of $T_{N,k}$ using the permutation procedure.

11.3 Depths-based tests

Another way of approaching the functional data analysis is introducing the notion of depth, that is related to a generalisation of the concept of ordering for functional data. The idea is to assign an order to each element of the sample related to its centrality within the whole set.

There are many measures of depth, and there is no agreement about their advantages. We will concentrate on the Fraiman-Muñiz depth [2].

Consider first a univariate sample $U_1,...,U_n$ and $U_{(1)},...,U_{(n)}$ be the corresponding order statistics. Assuming there are no ties, the natural depth of U_i is said to be $D_n(U_i) = \frac{1}{2} - \left|\frac{1}{2} - \left(\frac{j}{n} - \frac{1}{2n}\right)\right|$. This notion of depth assigns minimal and equal depth to the two extreme values of the sample, maximum to the innermost point(s) and changes linearly with the position the datum occupies in the sample.

For the case of functional data, consider the sample $X=\{X_i(t)\}$ defined in a common interval J. For each t we can compute the natural depth, $D_n(X_i(t))$, and then the depth for each $X_i(t)$ is:

$$I(X_i,X) = \int_J D_n(X_i(t))dt,$$

where, in practice, the integral is replaced by a sum over t for the time grid.

A Wilcoxon test based on this ordering is a natural option, and has been suggested by López-Pintado and Romo in [6], [7] based on their band-depth. Observe that for univariate samples, the Wilcoxon test statistic defined in terms of Fraiman-Muñiz depth corresponds to Ansay-Bradley's statistic, and hence is suitable for detecting differences in dispersion between the samples.

11.3.1 Meta-analysis based tests

Let $\mathscr{X} = \{X_1,\ldots,X_m\}$ denote the functional X sample and $\mathscr{Y} = \{Y_1,\ldots,Y_n\}$ the functional Y sample. For each $X_i \in \mathscr{X}$, we consider its depth with respect to the Y sample with the curve X_i added. We denote this depth $I(X_i, \mathscr{Y} \cup \{X_i\})$. This is a measure of how outlying the curve X_i is with respect to the Y sample. If "many" of the X_i turn out to be outlying with respect to $\mathscr{Y}$, that would be evidence against the null hypothesis of equality of distributions. Similarly we can measure how outlying is each curve Y_j with respect to the X-sample, $\mathscr{X}$, by computing $I(Y_j, \mathscr{X} \cup \{Y_j\})$. The first question is how to combine the values of $I(X_i, \mathscr{Y} \cup \{X_i\})$, for all $i \leq m$, in a single number that combines the information in all these depths. For this purpose, we rely in an idea coming from Meta-Analysis.

To the depth $I(X_i, \mathscr{Y} \cup \{X_i\})$ we associate an empirical p-value, Let p_i be an empirical p-value related to the *centrality* of the variable Y_i on the sample X,

$$\hat{p}_* = \frac{\sum_{j=1}^{m} I\left(D\left(X_j|X\right) \leq D\left(Y_*|X\right)\right)}{m+1} \qquad \begin{cases} p_i = \hat{p}_i \text{ if } \hat{p}_i \neq 0 \\ p_i = \frac{1}{m+1} \text{ if } \hat{p}_i = 0 \end{cases}$$

Under H_0 (ignoring ties), each p_i is uniformly distributed in $\{1/m, 2/m, \ldots, 1\}$, but they are not independent.

For symmetry, we consider as well the q_i p-values related to the *centrality* of the variable X_i on the sample Y.

$$\hat{q}_* = \frac{\sum_{j=1}^{n} I\left(D\left(Y_j|Y\right) \leq D\left(X_*|Y\right)\right)}{n+1} \qquad \begin{cases} q_i = \hat{q}_i \text{ if } \hat{q}_i \neq 0 \\ q_i = \frac{1}{n+1} \text{ if } \hat{q}_i = 0 \end{cases}$$

Observe that when H_0 does not hold, the depth of a curve in a family where it does not belong will be very low, and so would be its associated p-value. In that case the corresponding statistic will be very big.

Consider, as described in [4],

$$S_Y = \sum_{i=1}^{m} -\ln(p_i) \qquad S_X = \sum_{i=1}^{n} -\ln(q_i)$$

We want to associate a p-value to the pair (S_X, S_Y).

MA1: Meta-Analysis method 1

When the two samples display a difference in "scale", most, of the curves of the (say) X sample, turn out to be central with respect to the Y sample and S_X will not show a significant value. In such a situation, typically, several curves of the Y sample will turn out to be clearly outlying respect to the X sample, and the maximum will reach a significant value.

We propose the use of $S = \max\{S_X, S_Y\}$ as test statistic, and the use of a permutation procedure to obtain its distribution.

MA2: Meta-Analysis method 2

Better empirical performance is obtained using following result:

Lemma 1 *combining p-values Let p_X (p_Y) denote the p-value of S_X (S_Y), under the null permutation distribution, as obtained from procedure p-value if all subsets of size m were used (instead of just a sample of size B) and assuming the null hypothesis. Then*

1. $\Pr(p_X \leq t) \leq t$ *for any* $t \in (0,1)$*, and the same holds for* p_Y.
2. $\Pr(2\min(p_X, p_Y) \leq t) \leq t$ *for any* $t \in (0,1)$.

Proof. The null permutation distribution of S_X is a discrete distribution that can not be assumed uniform on its range (some values of S_X can appear more frequently than others when subsets are chosen at random). This is why part (i) of the Lemma is not obvious. Let F denote the null permutation c.d.f. of S_X and let $S_{X,\text{obs}}$ denote the observed value of S_X. Recall that large values of S_X are considered significant. Then, clearly, $p_X = 1 - F(S^-_{X,\text{obs}})$ and, for $t \in (0,1)$,

$$\Pr(p_X \leq t) = \Pr(F(S^-_{X,\text{obs}}) \geq 1-t) = \sum_{\{s:F(s)>1-t\}} \Pr(S_X = s) \leq t,$$

by definition of F.

Since p_X and p_Y are not independent, to prove (ii) we can use (i) together with the usual union bound:

$$\Pr(\min(p_X, p_Y) \leq t/2) \leq \Pr(p_X \leq t/2) + \Pr(p_Y \leq t/2) \leq t/2 + t/2 = t.$$

Part (2) of the Lemma tells us that an appropriate p-value for $2\min(p_X, p_Y)$ is the observed value of this statistic itself.

Thus, our second way of getting a p-value from S_X and S_Y is the following: Compute, approximately, p_X and p_Y for S_X and S_Y, respectively, using the permutation procedure and use $2\min(p_X, p_Y)$ as p-value.

11.4 Empirical comparison of powers and real data applications

In order to fix a standard, we have also computed the empirical power for Horvàth and Kokoszka's test for equality of mean functions. The null hypothesis is rejected for large values of the the statistic

$$U_{m,n} = \frac{mn}{m+n} \int \left(\hat{X}_m(t) - \hat{Y}_n(t)\right)^2 dt$$

Under some regularity conditions, the asymptotic distribution of $\hat{X}_m(t) - \hat{Y}_n(t)$ is a Gaussian process Γ whose covariance can be approximated by the pooled covariance operator of the two samples, hence, the distribution of $U_{m,n}$ can be approximated by the first d terms in the Karhunen-Loève expansion of $\int_0^1 \Gamma^2(s)ds \approx \sum_{i=1}^d \lambda_i N_i^2$, λ_i are the (ordered) eigenvalues of the pooled covariance estimator, and N_i are i.i.d. N(0,1).

11.4.1 A simulation experiment

We have simulated samples of functional data as realisations from a geometric Brownian motion process $f(t) = X_0 \exp\left\{rt - \frac{t\sigma^2}{2} + \sigma w_t\right\}$ where r and σ are, respectively,

the trend (drift) and volatility coefficients, w_t is a standard Wiener process and X_0 is the initial value.

We have tested $H_0 : \mathscr{L}(X) = \mathscr{L}(Y)$, against $\mathscr{L}(X) \neq \mathscr{L}(Y)$ where X was simulated with $r = 1, \sigma = 1, X_0 = 0$ and Y is any of the 'contaminated' samples with only one of the parameters variying at a time.

From Figure 11.1, we see that Wilcoxon's statistic performs very well against volatility variations but fails noticeably for the other alternatives considered in the experiment. On the other hand, Horvàth and Kokoszka's test (HK), being a test conceived for changes in the mean, shows the best performance against changes in the drift parameter, while its power numbers against changes in the origin (initial level) are good too. But HK results ineffective in picking the volatility changes.

The Meta Analysis methods have a power similar to HK against changes in the origin, while their power, although reasonable, is inferior to HK's when it comes to changes in drift. On the other hand, both Meta Analysis procedures display excellent power against the volatility alternatives, where HK fails. Schilling's statistic (with $k = 5$ and $K = 10$), shows very good power against all the alternatives considered in our experiment. Overall, Schilling's statistic displays the best performance in terms of power among the methods evaluated.

11.4.2 NO_2 Contamination in Barcelona

We have hourly measurements of nitrogen dioxide (a known pollutant formed in most combustion processes using air as the oxidant) in four neighbourhoods in Barcelona, namely Eixample, Palau Reial, Poblenou and Sants. The measurements were taken along the years 2014 and 2015 in automatic monitoring stations[1]

We have split the data sets into working days ($\approx$220 curves) and non-working days ($\approx$120 curves), each year.

There are many questions of interest, for instance, to assess whether the level of pollutants significantly different during working and non-working days, or if the levels of NO_2 changed from one year to the next in each of the neighbourhoods, or comparing the pollution levels among the different neighbourhoods. We include here the results of one of these many comparisons.

All tests show evidence of differences between working and non-working days in all four neigbourhoods, with Wilcoxon and Schilling-10 showing the strongest evidence of differences. Figure 11.2 shows the levels of contaminants in the neighbourhood of Sants in the years 2014 and 2015. The tests show that the level of pollutants did not change noticeably on non-working days, but significant changes are found from one year to the next on working days, with the Wilcoxon and Schilling procedures being the ones that find stronger evidence of change.

[1] available from http:// dtes.gencat.cat/icqa.

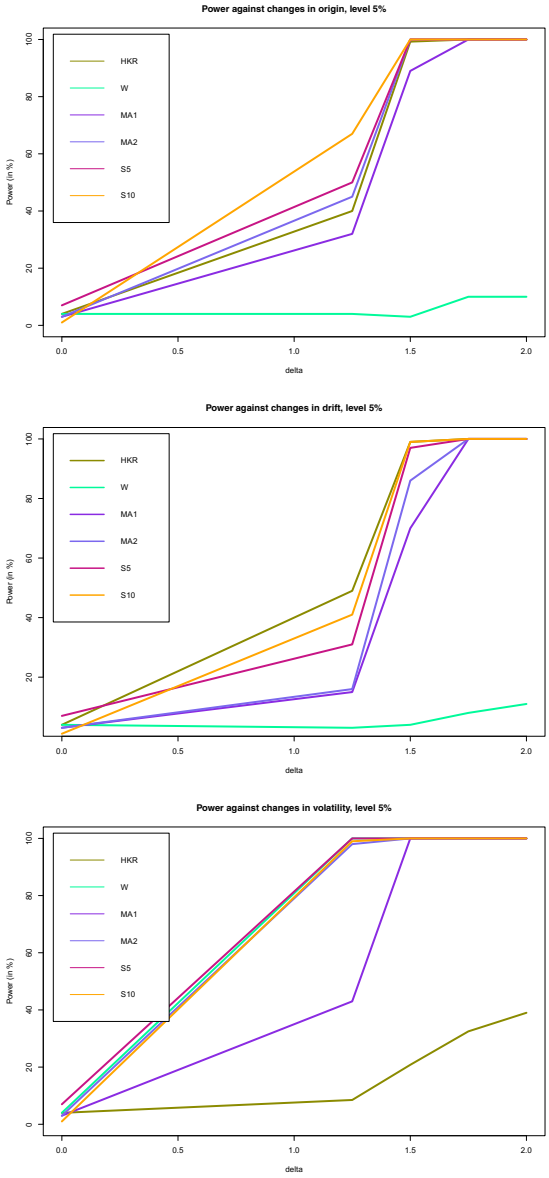

Fig. 11.1: Empirical power of the tests against changes of origen, drift and volatility in Geometric Brownian motion data; level $\alpha = 5\%$.

Acknowledgements The work of A. Cabaña was supported by Ministerio de Economía, Industria y Competitividad, Spain, project MTM2015-69493-R.

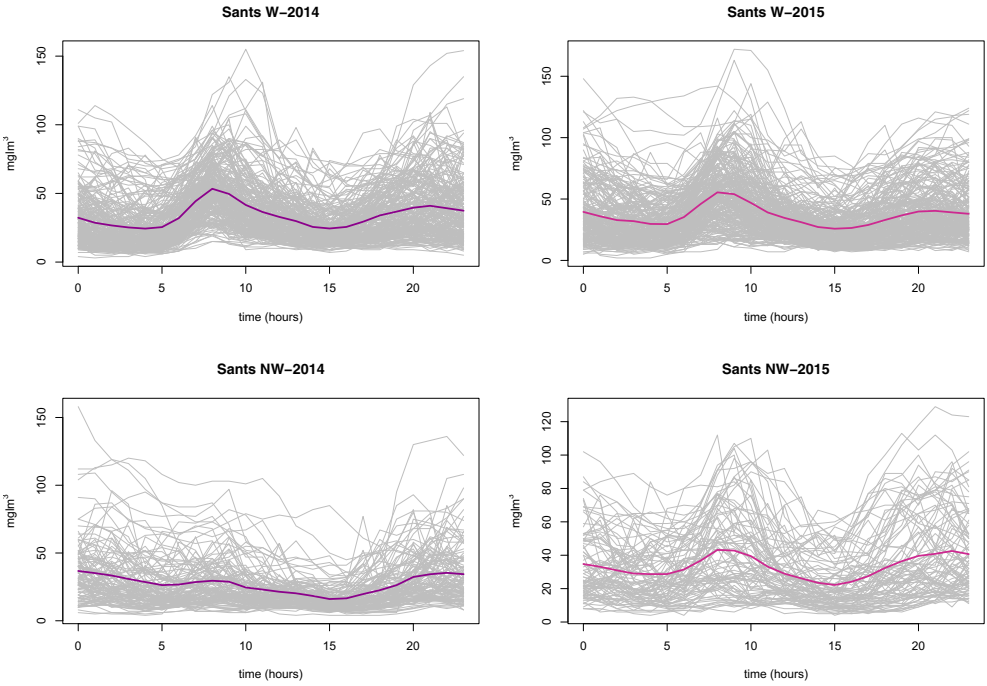

Fig. 11.2: NO_2 leves on working days (1st row) and non-working days (2nd row) in Sants, 2014-2015

References

[1] Estrada, A.M.: Nociones de Profundidad en Análisis de Datos Funcionales, MSc Thesis, Universidad de Los Andes, Colombia (2014)
[2] Fraiman, R., Muniz, G: Trimmed means for functional data. TEST **10**(2): 419–440 (2001)
[3] Good, P.: Permutation, Parametric and Bootstrap Tests of Hypothesis, Springer (2005)
[4] Hedges, L. V., Olkin, I.: Statistical Methods for Meta-Analysis, Academic Press (1985)
[5] Liu, R. Y., Parelius, J. M., Singh, K.: Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference. Ann. Statist. **27**(3), 783–840 (1999)
[6] López-Pintado, S., Romo, J.: Depth-based inference for functional data. Comput. Statist. Data Anal. **51**, 4957–4968 (2007)
[7] López-Pintado, S., Romo, J.: On the concept of depth for functional data. J. Amer. Statist. Assoc. **104**, 718–734 (2009)

[8] Peña, J.: Propuestas para el problema de dos muestras con datos funcionales., MSc Thesis, Universidad de Los Andes, Colombia (2012)
[9] Quiroz, A. J.: Graph-Theoretical Methods. In Encyclopedia of Statistical Sciences, 2nd Edition, S. Kotz, C. B. Read, N. Balakrishnan and B. Vidakovic, editors. Vol 5, pp 2910–2916, Wiley and Sons, New York (2006)
[10] Schilling, M. F. Multivariate two-sample tests based on nearest neighbours. J. Amer. Statist. Assoc. **81**:395, 799–806 (1986)