Some applications of the random projection method

Juan A. Cuesta-Albertos

Departamento de Matemáticas, Estadística y Computación
Universidad de Cantabria

V BoSanTouVal
June 3-5, 2009

This talk is based on some joint research with:

Manolo Febrero Bande

Universidad de Santiago de Compostela

What are we going to talk about?

I am going to present a general method which (in me opinion) is

- simple
- easy to compute
- reasonable
- ...

What are we going to talk about?

I am going to present a general method which (in me opinion) is

- simple
- easy to compute
- reasonable
- ...

OK, OK, OK,... but, where is the trick?

What are we going to talk about?

I am going to present a general method which (in me opinion) is

- simple
- easy to compute
- reasonable
- ...

OK, OK, OK,... but, where is the trick?

- This method is optimum under NO circumstance

Then?

What are we going to talk about?

I am going to present a general method which (in me opinion) is

simple

easy to compute

reasonable

...

OK, OK, OK,... but, where is the trick?

This method is optimum under NO circumstance

Then?

"...simple methods typically yield performance almost as good as more sophisticated methods to the extent that the difference in performance may be swamped by other sources of uncertainty..."

HAND, D.J., 2006. Classifier technology and the illusion of progress. *Statist. Sci.*, **21**(1) 1-14.

The basic result (in Hilbert spaces):

How many one-dimensional marginals

are required to determine a probability measure

on a separable Hilbert space?

The basic result (in Hilbert spaces):

How many one-dimensional marginals

are required to determine a probability measure

on a separable Hilbert space?

Only a one-dimensionial projection suffices

if it is randomly chosen

(under some assumptions on the moments)

Notation:

- $\boldsymbol{H}$ will denote a separable Hilbert space

  $\| - \|$ and $\langle \cdot, \cdot \rangle$ its norm and scalar product

- Given P a probability on $\boldsymbol{H}$ and $v \in \boldsymbol{H}$

  P_v is the marginal of P on the subspace generated by v

- Given P, Q two probabilities

$$\boldsymbol{E}(P, Q) := \{v \in \boldsymbol{H} : P_v = Q_v\}$$

The result. Separable Hilbert spaces.

Assume that:

1. *P is determined by its moments*
2. *$Q \neq P$*

Then $\mu[\boldsymbol{E}(P, Q)] = 0$ (remember: $\boldsymbol{E}(P, Q) = \{v : P_v = Q_v\}$)

Here μ is any continuous distribution

For instance:

μ absolutely continuous w.r.t. the Lebesgue measure

μ Gaussian, with non-degenerate 1-dimensional marginals

The result. Separable Hilbert spaces.

Assume that:

1. *P is determined by its moments*
2. *$Q \neq P$*

Then $\mu[\boldsymbol{E}(P,Q)] = 0$ (remember: $\boldsymbol{E}(P,Q) = \{v : P_v = Q_v\}$)

Extension to Banach spaces in Cuevas & Fraiman (2009)

The result. How to apply it.

Assume that:

1. *P is determined by its moments*
2. $Q \neq P$

Then $\mu[\boldsymbol{E}(P,Q)] = 0$ (remember: $\boldsymbol{E}(P,Q) = \{v : P_v = Q_v\}$)

If you want to test $H_0 : P = Q$

only select v at random at test $H_0^v : P_v = Q_v$

because, with probability one, H_0 and H_0^v are equivalent

How to choose v in practice?

How to choose v in practice?

1. On $\mathbb{R}^d$ Simulate v (for instance) from $N_d(0, Id)$.

How to choose v in practice?

1. On $\boldsymbol{R}^d$ Simulate v (for instance) from $N_d(0, Id)$.
2. Functional case.
   We need v with a Gaussian distribution μ

How to choose v in practice?

1. On $\boldsymbol{R}^d$ Simulate v (for instance) from $N_d(0, Id)$.
2. Functional case.

   We need v with a Gaussian distribution μ

   Data are discretized $\Rightarrow$ data belong to $\boldsymbol{R}^d$ with, perhaps, $d \gg 1$

How to choose v in practice?

1. On $\boldsymbol{R}^d$ Simulate v (for instance) from $N_d(0, Id)$.
2. Functional case.

   We need v with a Gaussian distribution μ

   Data are discretized $\Rightarrow$ data belong to $\boldsymbol{R}^d$ with, perhaps, $d \gg 1$

   Two possibilites.

   2.1 Simulate v from $N_d(0, Id)$.

       Multiply v by the appropriate matrix and add a function m

How to choose v in practice?

1. On $\boldsymbol{R}^d$ Simulate v (for instance) from $N_d(0, Id)$.
2. Functional case.

   We need v with a Gaussian distribution μ

   Data are discretized $\Rightarrow$ data belong to $\boldsymbol{R}^d$ with, perhaps, $d \gg 1$

   Two possibilites.

   2.1 Simulate v from $N_d(0, Id)$.

   Multiply v by the appropriate matrix and add a function m

   2.2 Apply a property of μ

   Assume that (theoretically) $v \in L_2[0,1]$

   that we have measured the data at points $t_1 < \ldots < t_d$

   that μ is the distribution of the standard Brownian motion

   take $\delta_i, i = 1, \ldots, d$ i.i.d. N(0,1)

   define

$$
\begin{aligned}
v(t_0) &= 0, \text{ where } t_0 = 0, \\
v(t_i) &= v(t_{i-1}) + (t_i - t_{i-1})^{1/2}\delta_i, \; i = 1, \ldots, d
\end{aligned}
$$

Two-way factorial ANOVA for functional data

We have two factors with R and S levels respectively

Thus, for every $r = 1, \ldots, R$ and $s = 1, \ldots, S$ we have
$\mathbf{X}_i^{r,s}, i = 1, ..., n_{r,s} \in \boldsymbol{N}$ random functions in $L_2[0, 1]$

Two-way factorial ANOVA for functional data

We have two factors with R and S levels respectively

Thus, for every $r = 1, \ldots, R$ and $s = 1, \ldots, S$ we have

$\mathbf{X}_i^{r,s}, i = 1, ..., n_{r,s} \in \boldsymbol{N}$ random functions in $L_2[0,1]$

$$\mathbf{X}_i^{r,s}(t) = m(t) + f^r(t) + g^s(t) + h^{r,s}(t) + \epsilon_i^{r,s}(t), \; t \in [0,1],$$

1. $m \in L_2[0,1]$ is non random. Describes the overall shape of the process
2. $f^r, g^s, h^{r,s} \in L_2[0,1]$ are non random. Account for the main effects of the factors and for the interaction between them; and

$$\sum_r f^r(t) = \sum_s g^s(t) = \sum_r h^{r,s_0}(t) = \sum_s h^{r_0,s}(t) = 0, \forall t, r_0, s_0$$

Two-way factorial ANOVA for functional data

We have two factors with R and S levels respectively

Thus, for every $r = 1, \ldots, R$ and $s = 1, \ldots, S$ we have

$$\mathbf{X}_i^{r,s}, i = 1, ..., n_{r,s} \in \boldsymbol{N} \text{ random functions in } L_2[0,1]$$

$$\mathbf{X}_i^{r,s}(t) = m(t) + f^r(t) + g^s(t) + h^{r,s}(t) + \epsilon_i^{r,s}(t),\ t \in [0,1],$$

1. $m \in L_2[0,1]$ is non random. Describes the overall shape of the process
2. $f^r, g^s, h^{r,s} \in L_2[0,1]$ are non random. Account for the main effects of the factors and for the interaction between them;
3. $\epsilon_i^{r,s} \in L_2[0,1]$, are random, independent and $E[\epsilon_i^{r,s}] = 0$ for each r, s, $\epsilon_i^{r,s}$ are i.d.

Two-way factorial ANOVA for functional data

We have two factors with R and S levels respectively

Thus, for every $r = 1, \ldots, R$ and $s = 1, \ldots, S$ we have

$\mathbf{X}_i^{r,s}, i = 1, ..., n_{r,s} \in \boldsymbol{N}$ random functions in $L_2[0,1]$

$$\mathbf{X}_i^{r,s}(t) = m(t) + f^r(t) + g^s(t) + h^{r,s}(t) + \epsilon_i^{r,s}(t), \;\; t \in [0,1],$$

1. $m \in L_2[0,1]$ is non random. Describes the overall shape of the process
2. $f^r, g^s, h^{r,s} \in L_2[0,1]$ are non random. Account for the main effects of the factors and for the interaction between them;
3. $\epsilon_i^{r,s} \in L_2[0,1]$, are random, independent and $E[\epsilon_i^{r,s}] = 0$ for each r, s, $\epsilon_i^{r,s}$ are i.d.

We want to test the null hypotheses:

$H_0^A:$	$f^1 = \ldots = f^R = 0$	the first factor has no effect
$H_0^B:$	$g^1 = \ldots = g^S = 0$	the second factor has no effect
$H_0^I:$	$h^{1,1} = \ldots = h^{R,S} = 0$	there is no interaction between factors

The theorem

Theorem (Cuesta-Albertos and Febrero-Bande, 2009)

Let us assume the previous model. If μ is Gaussian, then

1. *If H_0^A fails, then* $\mu\left\{v \in L_2[0,1] : \langle v, f^1\rangle = \ldots = \langle v, f^R\rangle\right\} = 0$
2. *If H_0^B fails, then* $\mu\left\{v \in L_2[0,1] : \langle v, g^1\rangle = \ldots = \langle v, g^S\rangle\right\} = 0$
3. *If H_0^I fails, then* $\mu\left\{v \in L_2[0,1] : \langle v, h^{1,1}\rangle = \ldots = \langle v, h^{R,S}\rangle\right\} = 0$

The theorem

Theorem (Cuesta-Albertos and Febrero-Bande, 2009)

Let us assume the previous model. If μ is Gaussian, then

1. *If H_0^A fails, then* $\mu\left\{v \in L_2[0,1] : \langle v, f^1\rangle = \ldots = \langle v, f^R\rangle\right\} = 0$
2. *If H_0^B fails, then* $\mu\left\{v \in L_2[0,1] : \langle v, g^1\rangle = \ldots = \langle v, g^S\rangle\right\} = 0$
3. *If H_0^I fails, then* $\mu\left\{v \in L_2[0,1] : \langle v, h^{1,1}\rangle = \ldots = \langle v, h^{R,S}\rangle\right\} = 0$

PROOF.- Let $r \in \{1, ..., R\}$, and let P^r be such that $P^r[f^r] = 1$

Obviously, P^r is determined by its moments

Thus, we can apply the result on random projections to every pair of probability distributions P^{r_1} and P^{r_2}

The proofs of 1 and 2 are identical.

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

But, what does it happen if the data are heteroscedastic?

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \ \ i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

But, what does it happen if the data are heteroscedastic?

Nothing!

We only need a (one-dimensional) procedure valid for heteroscedastic data

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

But, what does it happen if the data are not gaussian?

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \;\; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

But, what does it happen if the data are not gaussian?

Nothing!

We only need a (one-dimensional) procedure valid for non-gaussian data

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

But, what does it happen if there are covariables?

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

But, what does it happen if there are covariables?

Nothing!

We only need a (one-dimensional) procedure allowing covariables

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

But, what does it happen if ...?

Two-way factorial ANOVA. The procedure

To test H_0^A:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANOVA procedure to test the null hypothesis

$$H_0^{A,v} : \langle v, f^1 \rangle = \ldots = \langle v, f^R \rangle = 0$$

But, what does it happen if ...?

Nothing!

Well, at least if we have a (one-dimensional) procedure allowing ...

Therefore

The random ANOVA for functional data is a procedure which is

Therefore

The random ANOVA for functional data is a procedure which is

- Simple

Therefore

The random ANOVA for functional data is a procedure which is

- Simple
- Easy to compute

Therefore

The random ANOVA for functional data is a procedure which is

- Simple
- Easy to compute
- Flexible (it can be applied to many situations and designs)

Therefore

The random ANOVA for functional data is a procedure which is

- Simple
- Easy to compute
- Flexible (it can be applied to many situations and designs)
- ...

Therefore

The random ANOVA for functional data is a procedure which is

- Simple
- Easy to compute
- Flexible (it can be applied to many situations and designs)
- ...

Where is the price we have paid for this?

Therefore

The random ANOVA for functional data is a procedure which is

- Simple
- Easy to compute
- Flexible (it can be applied to many situations and designs)
- ...

Where is the price we have paid for this?

We are replacing functions by numbers

We are losing information, this should bring some loss of power

Therefore

The random ANOVA for functional data is a procedure which is

- Simple
- Easy to compute
- Flexible (it can be applied to many situations and designs)
- ...

Where is the price we have paid for this?

We are replacing functions by numbers

We are losing information, this should bring some loss of power

A solution: Choose $v_1, ..., v_k$ at random.

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \ldots, H_0^{A,v_k}$

And base the decision on the k tests

Technical problem: How to handle multiple tests?

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \ldots, H_0^{A,v_k}$

And base the decision on the k tests

Technical problem: How to handle multiple tests?

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \ldots, H_0^{A,v_k}$

And base the decision on the k tests

Compute the p-value of each test: $p_1, \ldots, p_k$

Take $p_0 = \min(p_1, \ldots, p_k)$

Technical problem: How to handle multiple tests?

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \ldots, H_0^{A,v_k}$

And base the decision on the k tests

Compute the p-value of each test: $p_1, \ldots, p_k$

Take $p_0 = \min(p_1, \ldots, p_k)$
Correct via Bonferroni $\longrightarrow$ too conservative

Technical problem: How to handle multiple tests?

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \dots, H_0^{A,v_k}$

And base the decision on the k tests

Compute the p-value of each test: $p_1, \dots, p_k$

Take $p_0 = \min(p_1, \dots, p_k)$

Correct via Bootstrap $\rightarrow$ too time consuming

Cuesta-Albertos et al, 2007

Technical problem: How to handle multiple tests?

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \ldots, H_0^{A,v_k}$

And base the decision on the k tests

Compute the p-value of each test: $p_1, \ldots, p_k$

Use the FDR = The expected proportion of erroneous rejections when testing k null hypotheses Benjamini&Hochberg, 1995

Technical problem: How to handle multiple tests?

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \ldots, H_0^{A,v_k}$

And base the decision on the k tests

Compute the p-value of each test: $p_1, \ldots, p_k$

Use the FDR = The expected proportion of erroneous rejections when testing k null hypotheses Benjamini&Hochberg, 1995

= the level of the test in our case, all hypotheses coincide

Technical problem: How to handle multiple tests?

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \ldots, H_0^{A,v_k}$

And base the decision on the k tests

Compute the p-value of each test: $p_1, \ldots, p_k$

Use the FDR = The expected proportion of erroneous rejections when testing k null hypotheses Benjamini&Hochberg, 1995

= the level of the test in our case, all hypotheses coincide

a test at level α in our problem: Theo. 1.3, Benjamini&Yekutyeli, 2001

sort the p-values to obtain $p_{(1)} \leq \ldots \leq p_{(k)}$

reject the null hypothesis under consideration if

$$\left\{ i \in \{1, \ldots, k\} : \; p_{(i)} \leq \frac{i}{k} \frac{\alpha}{\phantom{\sum_{j=1}^{k} 1/j}} \right\} \neq \emptyset$$

if the tests are positively dependent

Technical problem: How to handle multiple tests?

Apply the ANOVA to the hypotheses $H_0^{A,v_1}, \ldots, H_0^{A,v_k}$

And base the decision on the k tests

Compute the p-value of each test: $p_1, \ldots, p_k$

Use the FDR = The expected proportion of erroneous rejections when testing k null hypotheses Benjamini&Hochberg, 1995

= the level of the test in our case, all hypotheses coincide

a test at level α in our problem: Theo. 1.3, Benjamini&Yekutyeli, 2001

sort the p-values to obtain $p_{(1)} \leq \ldots \leq p_{(k)}$

reject the null hypothesis under consideration if

$$\left\{ i \in \{1, \ldots, k\} : \; p_{(i)} \leq \frac{i}{k} \frac{\alpha}{\sum_{j=1}^{k} \frac{1}{j}} \right\} \neq \emptyset$$

always!!!

functional ANOVA. Orthosis data.

How do individuals cope with a perturbation while stepping-in-place?

Seven volunteers wore a spring-loaded orthosis of adjustable stiffness

Experimental conditions: Control condition (without orthosis)

Orthosis condition (with the orthosis only)

Spring1, Spring2: a spring-loaded orthosis onto the knee joint

functional ANOVA. Orthosis data.

How do individuals cope with a perturbation while stepping-in-place?

Seven volunteers wore a spring-loaded orthosis of adjustable stiffness

Experimental conditions: Control condition (without orthosis)

Orthosis condition (with the orthosis only)

Spring1, Spring2: a spring-loaded orthosis onto the knee joint

For each of the seven subjects,

10 stepping-cycles of 20 seconds were analyzed under each condition

Moment at the knee was computed at 64 time points equally spaced and scaled so that a time interval corresponds to an individual gait cycle

functional ANOVA. Orthosis data.

How do individuals cope with a perturbation while stepping-in-place?

Seven volunteers wore a spring-loaded orthosis of adjustable stiffness

Experimental conditions: Control condition (without orthosis)

Orthosis condition (with the orthosis only)

Spring1, Spring2: a spring-loaded orthosis onto the knee joint

Antoniadis&Sapatinas, 2007 treated the subjects as random effects

We consider subjects and treatments as factors. 10 observations per cell

functional ANOVA. Orthosis data.

How do individuals cope with a perturbation while stepping-in-place?

Seven volunteers wore a spring-loaded orthosis of adjustable stiffness

Experimental conditions: Control condition (without orthosis)

Orthosis condition (with the orthosis only)

Spring1, Spring2: a spring-loaded orthosis onto the knee joint

Antoniadis&Sapatinas, 2007 treated the subjects as random effects

We consider subjects and treatments as factors. 10 observations per cell

RP	Subj.	Treat.	Inter.	Spr1&2 vs Co&Or	Cont vs Orth	Spr1 vs Spr2
5	0	0	0	0	1.86e-05	.0908
15	0	0	0	0	2.67e-05	.0231
30	0	0	0	0	3.22e-05	.0279
A&S				0	.001	.020

functional ANOVA. Orthosis data.

How do individuals cope with a perturbation while stepping-in-place?

Seven volunteers wore a spring-loaded orthosis of adjustable stiffness

Experimental conditions: Control condition (without orthosis)

Orthosis condition (with the orthosis only)

Spring1, Spring2: a spring-loaded orthosis onto the knee joint

Antoniadis&Sapatinas, 2007 treated the subjects as random effects

We consider subjects and treatments as factors. 10 observations per cell

using Bonferroni's correction:

RP	Subj.	Treat.	Inter.	Spr1&2 vs Co&Or	Cont vs Orth	Spr1 vs Spr2
5	0	0	0	0	1.86e-05	.0714
15	0	0	0	0	2.67e-05	.2141
30	0	0	0	0	3.22e-05	.1451
A&S				0	.001	.020

Comparison with MANOVA

Multidimensional data can be considered as functional

Comparison with MANOVA

Multidimensional data can be considered as functional

Data on the production of plastic film (Krzanowski, 1988):

- three characteristics: tear, gloss, opacity
- two factors: rate, additive
- with two levels each: low, high

Five measurements under each set of production conditions

$\Rightarrow$ 3-dimensional, 2-way MANOVA. 2 levels in each factor. $n_{i,j} = 5$

Comparison with MANOVA

Multidimensional data can be considered as functional

Data on the production of plastic film (Krzanowski, 1988):

- three characteristics: tear, gloss, opacity
- two factors: rate, additive
- with two levels each: low, high

Five measurements under each set of production conditions

$\Rightarrow$ 3-dimensional, 2-way MANOVA. 2 levels in each factor. $n_{i,j} = 5$

We take $k = 5, 15, 30$ random projections with $N_3(0, Id)$

We use the usual ANOVA test (Krzanowski uses Normal MANOVA)

Comparison with MANOVA

Multidimensional data can be considered as functional

Data on the production of plastic film (Krzanowski, 1988):

- three characteristics: tear, gloss, opacity
- two factors: rate, additive
- with two levels each: low, high

Five measurements under each set of production conditions

$\Rightarrow$ 3-dimensional, 2-way MANOVA. 2 levels in each factor. $n_{i,j} = 5$

We take $k = 5, 15, 30$ random projections with $N_3(0, Id)$

We use the usual ANOVA test (Krzanowski uses Normal MANOVA)

	Pillai test *p*-value	Random projection tests *p*-value		
		$k = 5$	$k = 15$	$k = 30$
rate	.003	.018	.007	.001
additive	.025	.005	.009	.008
interact.	.302	.263	.174	.192

Comparison with MANOVA

Multidimensional data can be considered as functional

Data on the production of plastic film (Krzanowski, 1988):

- three characteristics: tear, gloss, opacity
- two factors: rate, additive
- with two levels each: low, high

Five measurements under each set of production conditions

$\Rightarrow$ 3-dimensional, 2-way MANOVA. 2 levels in each factor. $n_{i,j} = 5$

We take $k = 5, 15, 30$ random projections with $N_3(0, Id)$

We use the usual ANOVA test (Krzanowski uses Normal MANOVA)

We have done 500 repetitions of the random ANOVA at the 0.05 level

	Pillai test p-value	Random projection tests p-value		
		$k = 5$	$k = 15$	$k = 30$
rate	.003	.018	.007	.001
additive	.025	.005	.009	.008
interact.	.302	.263	.174	.192

Comparison with MANOVA

Multidimensional data can be considered as functional

Data on the production of plastic film (Krzanowski, 1988):

- three characteristics: tear, gloss, opacity
- two factors: rate, additive
- with two levels each: low, high

Five measurements under each set of production conditions

$\Rightarrow$ 3-dimensional, 2-way MANOVA. 2 levels in each factor. $n_{i,j} = 5$

We take $k = 5, 15, 30$ random projections with $N_3(0, Id)$

We use the usual ANOVA test (Krzanowski uses Normal MANOVA)

We have done 500 repetitions of the random ANOVA at the 0.05 level

	Pillai test	Random projection tests					
		p-value			Rate of rejections		
	p-value	$k = 5$	$k = 15$	$k = 30$	$k = 5$	$k = 15$	$k = 30$
rate	.003	.018	.007	.001	.882	.998	1
additive	.025	.005	.009	.008	.772	.974	1
interact.	.302	.263	.174	.192	0	0	0

functional ANCOVA.

We have two factors with R and S levels respectively and a covariable

Thus, for every $r = 1, \ldots, R$ and $s = 1, \ldots, S$ we have
$\mathbf{X}_i^{r,s}, i = 1, ..., n_{r,s} \in \boldsymbol{N}$ random functions in $L_2[0,1]$

$$\mathbf{X}_i^{r,s}(t) = m(t) + f^r(t) + g^s(t) + h^{r,s}(t) + \epsilon_i^{r,s}(t) + \gamma Y_i^{r,s}(t), \;\; t \in [0,1]$$

functional ANCOVA.

We have two factors with R and S levels respectively and a covariable

Thus, for every $r = 1, \ldots, R$ and $s = 1, \ldots, S$ we have
$\mathbf{X}_i^{r,s}, i = 1, ..., n_{r,s} \in \boldsymbol{N}$ random functions in $L_2[0,1]$

$$\mathbf{X}_i^{r,s}(t) = m(t) + f^r(t) + g^s(t) + h^{r,s}(t) + \epsilon_i^{r,s}(t) + \gamma_i^{r,s} Y(t), \ \ t \in [0,1]$$

functional ANCOVA.

We have two factors with R and S levels respectively and a covariable

Thus, for every $r = 1, \ldots, R$ and $s = 1, \ldots, S$ we have
$\mathbf{X}_i^{r,s}, i = 1, ..., n_{r,s} \in \boldsymbol{N}$ random functions in $L_2[0,1]$

$$\mathbf{X}_i^{r,s}(t) = m(t) + f^r(t) + g^s(t) + h^{r,s}(t) + \epsilon_i^{r,s}(t) + \gamma_i^{r,s} Y(t), \;\; t \in [0,1]$$

We handle the covariable exactly in the same way as the factors:

Select a vector $v \in L_2[0,1]$ (with the distribution of a Brownian motion)

Compute the (real) projections of the sample

$$\langle v, \mathbf{X}_i^{r,s} \rangle, \;\; i = 1, \ldots, n_{r,s}, r = 1, \ldots, R, s = 1, \ldots, S$$

Apply an ANCOVA procedure to test the null hypothesis

$$H_0^{C,v} : \langle v, Y \rangle = 0$$

Notice that $\gamma_i^{r,s}$ has no influence iff $Y \equiv 0$

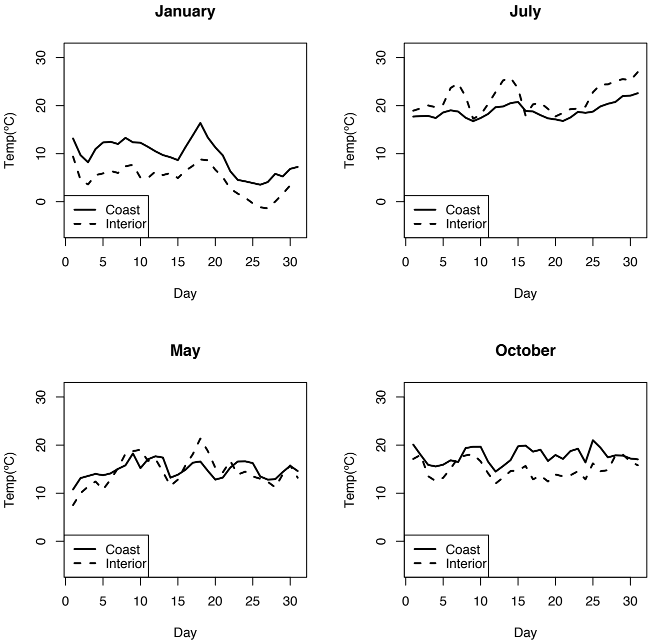
Spanish temperature data. Description of the data

Data: daily mean temp, certain locations and months. An annual cycle

* Months (4 levels):	October-06, January-07, May-07 and July-07
* Locations (2 levels):	
Coast:	A Coruña, Avilés, Bilbao, San Sebastián, Santander, Vigo
Inland:	Burgos, León, Madrid, Salamanca, Segovia, Soria, Valladolid, Vitoria and Zamora.
* Covariable (γ):	Monthly Total Amount of Rainfall
	γ is a real known r.v. which multiplies the unknown, non random function Y measuring the influence of γ each day in the month

downloaded from http://clima.meteored.com

Spanish temperature data. Means by cells



Spanish temperature data. Random projected ANCOVA

$\#$ of projections = 30

Correction method: Bonferroni and Bootstrap (B=500)

	Location	Month	Interaction	Rainfall
Bonf: p-value	$4.9 \cdot 10^{-6}$	$2.4 \cdot 10^{-33}$	$6.8 \cdot 10^{-9}$	.029
Boot: p-value	0	0	0	.037

Spanish temperature data. Random projected ANCOVA

$\#$ of projections = 30

Correction method: Bonferroni and Bootstrap (B=500)

	Location	Month	Interaction	Rainfall
Bonf: p-value	$4.9 \cdot 10^{-6}$	$2.4 \cdot 10^{-33}$	$6.8 \cdot 10^{-9}$	.029
Boot: p-value	0	0	0	.037

We have repeated the test 500 times

Spanish temperature data. Random projected ANCOVA

$\#$ of projections $=$ 30

Correction method: Bonferroni and Bootstrap (B=500)

	Location	Month	Interaction	Rainfall
Bonf: p-value	$4.9 \cdot 10^{-6}$	$2.4 \cdot 10^{-33}$	$6.8 \cdot 10^{-9}$	.029
Boot: p-value	0	0	0	.037

We have repeated the test 500 times

Proportions of rejections of the null hypotheses (level $\alpha = 0.05$)

Bonferroni	1	1	1	.804
Bootstrap	1	1	1	.808

T H A N K Y O U !!!

e-mail: cuestaj@unican.es
http://personales.unican.es/cuestaj