Functional Principal Component Analysis with Long Range Dependent Errors

Jan Beran, Haiyan Liu and Klaus Telkmann

Department of Mathematics and Statistics
University of Konstanz

August 27-30, 2015
Loccum

1 Introduction

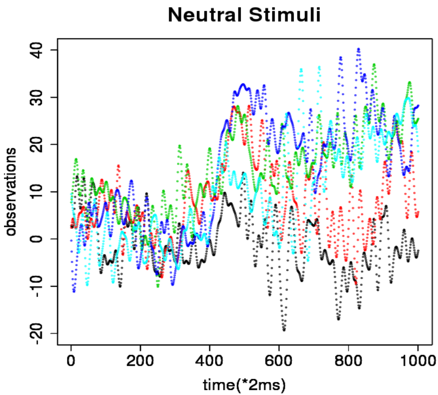


Figure 1: EEG signals in a Flanker task experiment with neutral stimuli on electrode Cz. (Dambacher et al., Uni. Konstanz)

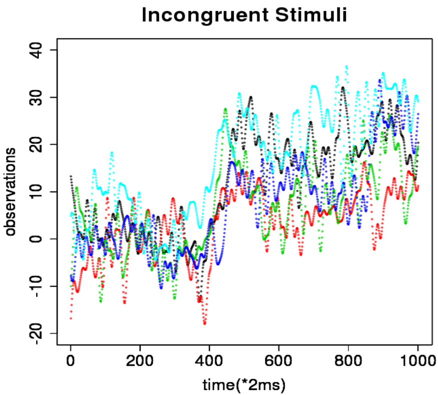


Figure 2: EEG signals in a Flanker task experiment with incongruent stimuli on electrode Cz. (Dambacher et al., Uni. Konstanz)

1 Introduction

- Existing literature:
  - In classical functional data analysis (FDA), the random curves are assumed to be observed directly, see Ramsay and Silverman (2005), Horváth and Kokoszka (2012).
  - In a more realistic setting, the random curves are discrete and may be perturbed by random noise, for the case of iid errors, see Yao, Müller and Wang (2005), Yao (2007).
- We are focusing on:
  - For certain types of observations, (long range) dependence in the error process can occur, see Beran and Liu (2014a, 2014b), Beran, Liu and Telkmann (2015).

2 Model

- Observations include two independent samples: $Y_{ij}^{(1)}$, $Y_{ij}^{(2)}$.
- For each sample, we observe n (omit superscript here) independent time series $Y_{i\cdot} = (Y_{i1}, ..., Y_{iN})(i = 1, ..., n)$ defined by

$$Y_{ij} = X_i(t_j) + \epsilon_i(j) \quad (t_j = jN^{-1}, j = 1, ..., N). \tag{1}$$

- Random curves $(X_1(t), ..., X_n(t) \sim_{i.i.d.} X(t))$ of the form

$$X_i(t) = \mu(t) + \sum_{l=1}^{p} \xi_{il}\phi_l(t) \quad (p \leq \infty), \tag{2}$$

  with $E[X(t)] = \mu(t)$ and $cov(X(s), X(t)) = C(s, t) = \sum_l \lambda_l \phi_l(s)\phi_l(t)$, where $\xi_{il} \sim_{i.i.d.} N(0, \lambda_l)$ and $\sum \lambda_l < \infty$.
- Error processes $(\epsilon_i(j))_{j\in\mathbb{N}}$ are stationary Gaussian with

$$\gamma_\epsilon(k) = cov\left(\epsilon_i(j), \epsilon_i(j+k)\right) \underset{k\to\infty}{\sim} c_\gamma |k|^{2d-1} \tag{3}$$

  for some $0 < c_\gamma < \infty$ and $d \in \left(0, \frac{1}{2}\right)$.

3 Estimators

- Let $\bar{y}_{\cdot j} = n^{-1} \sum_{i=1}^{n} Y_{ij}$, the kernel estimator of $\mu(t)$ is defined by

$$\hat{\mu}(t) = \frac{1}{Nb} \sum_{j=1}^{N} K_1 \left(\frac{t - t_j}{b} \right) \bar{y}_{\cdot j}. \tag{4}$$

- Let $C_{ijk} = (Y_{ij} - \hat{\mu}(t_j))(Y_{ik} - \hat{\mu}(t_k))$, the two-dimensional kernel estimator of $C(s,t)$ is defined as

$$\hat{C}(s,t) = \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_2 \left(\frac{s - t_j}{b}, \frac{t - t_k}{b} \right) n^{-1} \sum_{i=1}^{n} C_{ijk}. \tag{5}$$

- The eigenfunctions and eigenvalues of $C(s,t)$ are estimated as solutions of the equation

$$\hat{C}(s,t) = \sum_{l} \hat{\lambda}_l \hat{\phi}_l(s) \hat{\phi}_l(t). \tag{6}$$

4 Assumptions

- (K1) K_1 is a symmetric density function with support $[-1, 1]$,
- (K2) $\|K_1\|^2 = \int K_1^2(t)dt < \infty$,
- (K3) $K_1 \in C^1(\mathbb{R})$,
- (K4) $0 < \int K_1(t)t^2 dt < \infty$,
- (K5) $K_2(s, t) = K_1(s)K_1(t)$,
- (K6) For some $v \in \mathbb{N}$,

$$\int K_1(t)t^j dt = \begin{cases} 0 & j = 1, ..., 2v - 1, \\ C_K \in (0, +\infty) & j = 2v \end{cases}$$

5 Notations

- Here, "$\Rightarrow$" denotes weak convergence in $C[0,1]$ or $C[0,1]^2$ equipped with the supremum norm.
- Here, ζ_i, ζ_{ij}, $(i, j \in \mathbb{N})$ are iid standard normal random variables.

Theorem 1

Suppose:

0) Y_{ij} *be defined by (1), (2) and (3),*

i) *(K1), (K2), (K3), (K4) hold,*

ii) $\mu \in C^2[0,1]$,

iii) $n \to \infty$, $N \to \infty$, $b = b_N \to 0$, *such that*

$$\liminf N b^{1+2/(1-2d)} > q \in \mathbb{R}_+.$$

Then: $\sqrt{n}\left(\hat{\mu}(t) - E\left[\hat{\mu}(t)\right]\right) \Rightarrow \sum_{l=1}^{p} \sqrt{\lambda_l}\phi_l(t)\zeta_l \quad (t \in [0,1])$.

Suppose in addition:

iv) $nb^4 \to 0$.

Then: $\sqrt{n}\left(\hat{\mu}(t) - \mu(t)\right) \Rightarrow \sum_{l=1}^{p} \sqrt{\lambda_l}\phi_l(t)\zeta_l \quad (t \in [0,1])$.

Remark 1

- *Condition iii) is used to prove the tightness while obtaining the weak convergence.*
- *Condition iv) is only required to make sure that the bias of $\hat{\mu}(t)$ is in the order $o\left(n^{-\frac{1}{2}}\right)$.*
- *Conditions iii) and iv) together imply*

$$n = o\left(N^{4\frac{1-2d}{3-2d}}\right) \quad (0 < d < 0.5). \tag{7}$$

7 Contrast transformation

Denote by $\mathbf{1} = (1, ..., 1)^T$ and let $c_{\cdot 1}, ..., c_{\cdot n-1} \in \mathbb{R}^n$ ($c_{\cdot i} = (c_{1i}, ..., c_{ni})^T$)be such that $\langle \mathbf{1}, c_{\cdot i} \rangle = 0$, $\langle c_{\cdot i}, c_{\cdot i'} \rangle = \delta_{ii'}$. We then define $n-1$ contrast series Y_{ij}^c by

$$Y_{ij}^c = \langle c_{\cdot i}, Y_{\cdot j} \rangle = \sum_{s=1}^{n} c_{si} Y_{sj}.$$

Then

$$Y_{ij}^c = \sum_{l=1}^{p} \xi_{il}^c \phi_l(t_j) + \epsilon_i^c(j) \quad (i = 1, \cdots, n-1; j = 1, \cdots, N) \tag{8}$$

with

$$\xi_{il}^c = \sum_{s=1}^{n} c_{si} \xi_{sl}, \qquad \epsilon_i^c(j) = \sum_{s=1}^{n} c_{si} \epsilon_s(j).$$

7 Contrast transformation

Under the Gaussian assumption and the equidistant assumption, $Y_{ij}^c =_d Y_{ij}$ with $\mu(t) = 0$. In the following we assume that $\mu(t) = 0$, i.e.

$$Y_{ij} := \sum_{l=1}^{p} \xi_{il}\phi_l(t_j) + \epsilon_i(j), \tag{9}$$

and the covariance estimator (5) can be replaced by

$$\hat{C}(s,t) = \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) n^{-1} \sum_{i=1}^{n} Y_{ij}Y_{ik}. \tag{10}$$

8 Weak Convergence of $\hat{C}(s, t)$

Theorem 2

Suppose:

0) $Y_{ij} = Y_{ij}^c$ *given by (9) be defined by (1), (2) and (3) (with* $\mu(t) \equiv 0$*),*
i) *(K1), (K2), (K3), (K4), (K5) hold,*
ii) $C(s, t) \in C^2[0, 1]^2$,
iii) $n \to \infty$, $N \to \infty$, $b = b_N \to 0$, *such that*

$$\liminf N b^{1+2/(1-2d)} > q \in \mathbb{R}_+.$$

Then:

$$\sqrt{n}\left(\hat{C}(s,t) - E\left[\hat{C}(s,t)\right]\right) \Rightarrow Z_1(s,t) + Z_2(s,t) \quad (s, t \in [0, 1]),$$

where Z_1, Z_2 *are zero mean Gaussian processes that are independent from each other.*

8 Weak Convergence of $\hat{C}(s,t)$

Remark 2

For $\sqrt{n}\left(\hat{C}(s,t) - C(s,t)\right)$, the bias term of $O(b^2)$ is more difficult to handle. Since it stems from properties of $C(s,t)$ and the $K_2(s,t)$ only, this problem can be resolved by imposing additional differentiability assumptions on $C(s,t)$ and using higher order kernels.

8 Weak Convergence of $\hat{C}(s,t)$

Theorem 3

Suppose:

0) $Y_{ij} = Y_{ij}^c$ *given by (9) be defined by (1), (2) and (3) (with* $\mu(t) \equiv 0$*),*

i) *(K1), (K2), (K3), (K4), (K5), (K6) with* $v \geq 2$ *hold,*

ii) $C(s,t) \in C^{2v+2}[0,1]^2$,

iii) $n \to \infty$, $N \to \infty$, $b = b_N \to 0$, *such that*

$$Nb^{2v+1} \to \infty,\ nb^{4v} \to 0,\ \liminf Nb^{1+2/(1-2d)} > q \in \mathbb{R}_+.$$

Then:

$$\sqrt{n}\left(\hat{C}(s,t) - C(s,t)\right) \Rightarrow Z_1(s,t) + Z_2(s,t) \quad (s,t \in [0,1]).$$

where Z_1, Z_2 *are as in Theorem 2.*

Remark 3

- *(K6)+ii)+iii) imply*

$$n = o\left(min\left\{N^{\frac{4v}{2v+1}}, N^{4v\frac{1-2d}{3-2d}}\right\}\right). \tag{11}$$

- *Condition (11) is much better than (7). For instance, for fixed d, we may choose* $v > \frac{1}{1-2d}$*, thus we have,*

$$n = o\left(N^{\frac{4v}{2v+1}}\right).$$

Theorem 4

Suppose:

0) $\hat{\lambda}_l$ *and* $\hat{\phi}_l(t)$ *are estimated from (6),*
i) *Assumptions of Theorem 3 hold,*
ii) *For some* $m \in \mathbb{N}$, $\lambda_1 > \lambda_2 > \ldots > \lambda_m > \lambda_{m+1} > 0$,
iii) $sign\left(\langle \hat{\phi}_l, \phi_l \rangle\right) = 1$ *for* $l = 1, \ldots, m$.

Then: for each $l \in \{1, \ldots, m\}$,

- $\sqrt{n}\left(\hat{\lambda}_l - \lambda_l\right) \to_d \sqrt{2}\lambda_l \zeta_{ll}$,
- $\sqrt{n}\left(\hat{\phi}_l(t) - \phi_l(t)\right) \Rightarrow$
$\sum_{k:k>l} \sqrt{\lambda_l \lambda_k}(\lambda_l - \lambda_k)^{-1} \phi_k(t) \zeta_{lk} + \sum_{k:k<l} \sqrt{\lambda_l \lambda_k}(\lambda_l - \lambda_k)^{-1} \phi_k(t) \zeta_{kl}$.

9 Asymptotic Properties of $\hat{\lambda}_l$ and $\hat{\phi}_l(t)$

Remark 4

- *The weak convergence of covariance operator is needed.*
- *ii) can be generalized: for* $c_\lambda > 0$, *let* $\boldsymbol{I} = \{l_1, ..., l_k\}$ *denote the set of indices such that* $\lambda_l > c_\lambda$ *and* $\lambda_l > \lambda_{l+1}$.
- *The asymptotic distribution of* $\hat{\lambda}_l$ *and* $\hat{\phi}_l(t)$ *does not depend on d.*

Remark 5

In contrast to $\hat{\lambda}_l$ *and* $\hat{\phi}_l(t)$, *the rate of convergence and the asymptotic distribution of* $\hat{\xi}_{il}$ *differ distinctly between the cases of short and long memory, see Beran and Liu (2014b).*

10 Test Statistics

- Notations and assumptions:
  - $0 < p = p^{(1)} = p^{(2)} \leq \infty$,
  - $0 < m \leq \min\left\{m^{(1)}, m^{(2)}, p\right\}$,
  - $\mathcal{U} = span\left\{\phi_1^{(1)}(t), ..., \phi_m^{(1)}(t)\right\}$,
  - $\mathcal{V} = span\left\{\phi_1^{(2)}(t), ..., \phi_m^{(2)}(t)\right\}$.
- We would like to test:
  - Null hypothesis: $H_0 : \mathcal{U} = \mathcal{V}$,
  - Alternative hypothsis: $H_A : \mathcal{U} \neq \mathcal{V}$.

10 Test Statistics

- Consider the residual functions

$$r_l(t) = \phi_l^{(2)}(t) - \sum_{i=1}^{m} a_{il}\phi_i^{(1)}(t) \ (l = 1, ..., m),$$

where $\{a_{1l}, ..., a_{ml}\} = \underset{\{a_{1l},...,a_{ml}\}}{argmin} \left\| \phi_l^{(2)} - \sum a_{il}\phi_i^{(1)} \right\|^2.$

- Under H_0, we have $r_l(t) \equiv 0$ and $\sum_{i=1}^{m} a_{il}a_{il'} = \delta_{ll'}$.
  Under H_A, there is at least one $l \in \{1, ..., m\}$ for which $\|r_l\|^2 > 0$.
- The standardized residuals are estimated by

$$\tilde{r}_{n,N;l}(t) = \sqrt{\frac{n^{(1)}n^{(2)}}{n^{(1)} + n^{(2)}}} \left[\hat{\phi}_l^{(2)}(t) - \sum_{i=1}^{m} \hat{a}_{il}\hat{\phi}_i^{(1)} \right],$$

with $\hat{a}_{il} = \left\langle \hat{\phi}_i^{(1)}(t), \hat{\phi}_l^{(2)}(t) \right\rangle$.

Theorem 5

Suppose:

0) H_0 *is true,*
i) *Assumptions of Theorem 4 hold,*
ii) *There exists an* $\eta \in (0,1)$ *such that* $\frac{n^{(1)}}{n^{(1)}+n^{(2)}} \to \eta$ *(as* $n^{(1)}, n^{(2)} \to \infty$*),*
iii) *Define* $\Lambda_{ij}^{(r)} = \sqrt{\lambda_i^{(r)}\lambda_j^{(r)}}\left(\lambda_i^{(r)} - \lambda_j^{(r)}\right)^{-1}$ *for* $i \neq j \in \{1, ..., p\}$ $r = 1, 2$.

Then

$$\tilde{r}_{n,N;l} \Rightarrow Z_{res;l,1} - Z_{res;l,2} = \sqrt{\eta} \sum_{k=m+1}^{p} \Lambda_{lk}^{(2)} \left(\phi_k^{(2)} - \sum_{i=1}^{m} a_{ik}\phi_i^{(1)}\right) \zeta_{lk}^{(2)},$$
$$- \sqrt{1-\eta} \sum_{k=m+1}^{p} \sum_{i=1}^{m} \Lambda_{ik}^{(1)} \left(a_{kl}\phi_i^{(1)} + a_{il}\phi_k^{(1)}\right) \zeta_{ik}^{(1)}.$$

10 Test Statistics

- Theorem 5 implies that under H_0, for $l \in \{1, ..., m\}$,

$$\int_0^1 \tilde{r}^2_{n,N;l}(t)dt \underset{d}{\rightarrow} U_l = \int_0^1 \left(Z_{\text{res};l,1}(t) - Z_{\text{res};l,2}(t)\right)^2 dt.$$

- Denote by $\alpha \in (0, 1)$ be the level of significance.
- Denote by F_{U_l} the distribution function of U_l and by $q_{\alpha,m;l} = F^{-1}_{U_l}(1 - \alpha/m)$ its $(1 - \alpha/m)-$quantile.
- A Bonferroni corrected rejection region can be defined by

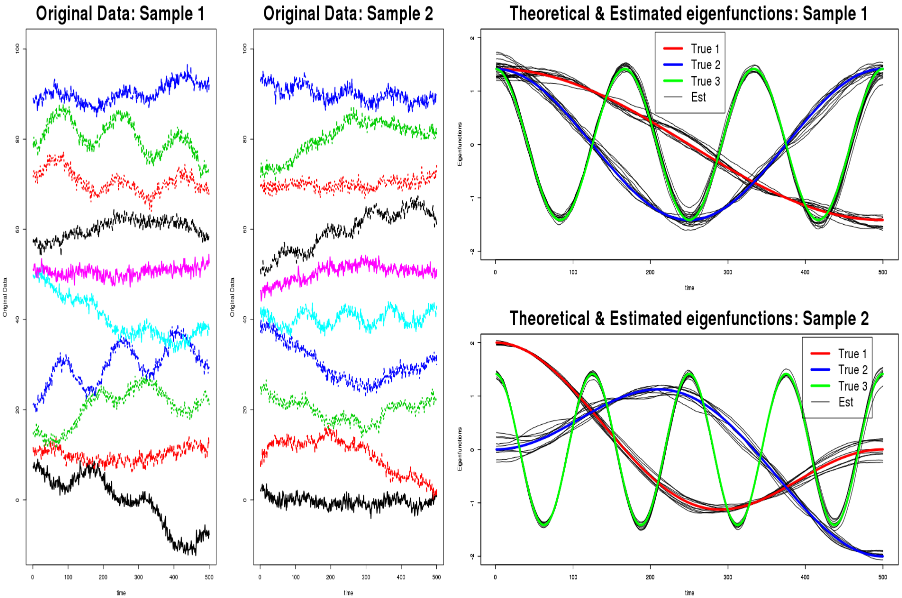
$$K_\alpha = \left\{ Y_{ij}^{(k)}(k = 1, 2) : \; U_{n,N;l} > q_{\alpha,m;l} \text{ for at least one } l \in \{1, ..., m\} \right\}.$$

Remark 6

- *Fremdt, Steinebach, Horváth, Kokoszka (2011) and Boente, Rodriguez, Sued (2011): covariance operator & without errors.*
- *Benko, Härdle and Kneip (2009): eigenvalues, eigenfunctions and eigenspaces & without errors.*
- *We focus on the equality of eigenspaces of two samples with dependent errors.*

	Sample 1	Sample 2
Theoretical Model		
Y_{ij}	$\xi_{i1}\phi_1(t_j) + \xi_{i2}\phi_2(t_j) + \xi_{i3}\phi_3(t_j) + \epsilon_i(j)$	$\xi_{i1}\phi_1(t_j) + \xi_{i2}\phi_2(t_j) + \xi_{i3}\phi_3(t_j) + \epsilon_i(j)$
n	416	416
N	500	500
ϵ_i	farima$(0, 0.3, 0)$	farima$(0, 0.3, 0)$
p	3	3
ϕ	$\left(\sqrt{2}\cos\pi t, \sqrt{2}\cos 2\pi t, \sqrt{2}\cos 6\pi t\right)$	$\left(\cos\pi t + \cos 2\pi t, \cos\pi t - \cos 2\pi t, \sqrt{2}\cos 8\pi t\right)$
m	2	2
	$\mathcal{U} = span\left(\sqrt{2}\cos\pi t, \sqrt{2}\cos 2\pi t\right)$	$\mathcal{V} = span\left(\cos\pi t + \cos 2\pi t, \cos\pi t - \cos 2\pi t\right)$
λ	$(6, 4, 2)$	$(6, 3, 1)$
Simulated Results		
Nsim	400	400
$K_1(t)$	$\frac{1}{2}\mathbf{1}_{\{-1<t<1\}}$	$\frac{1}{2}\mathbf{1}_{\{-1<t<1\}}$
b	0.0185	0.0185
ave$\hat{\lambda}$	$(6.099, 3.993, 1.98)$	$(6.015, 3.034, 0.926)$

11 Simulation



11 Simulation

Rejection Probability		
α	0.05	0.01
α_{U_1}	0.0425	0.005
α_{U_2}	0.045	0.005
α_{Bonf}	0.0275	0.005

Bibliography

Benko, M., Härdle, W. and Kneip, A. (2009). Common functional principal components. *The Annals of Statistics*, **37(1)**, 1-34.

Beran, J. and Liu, H. (2014a). On estimation of mean and covariance functions in repeated time series with long-memory errors. *Lithuanian Mathematical Journal*, **54(1)**, 8-34.

Beran, J. and Liu, H. (2014b). Estimation of eigenvalues, eigenvectors and scores in FDA models with strongly dependent errors. *Journal of Multivariate Analysis*, (submitted).

Beran, J., Liu, H. and Telkmann, K. (2015). On two sample inference for eigenspaces in FDA with dependent errors. *Journal of Statistical Planning and Inference*, (submitted).

Boente, G., Rodriguez, D. and Sued, M. (2011). Testing the equality of covariance operators. *Recent advances in functional data analysis and related topics*, Physica-Verlag HD, 49-53.

Fremdt, S., Steinebach, J.G., Horváth, L. and Kokoszka, P. (2013). Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, **40(1)**, 138-152.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis (Second Edition)*. Springer.

Yao, F., Müller, H.G. and Wang, J.L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100(470)**, 577-590.

Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *Journal of Multivariate Analysis*, **98(1)**, 40-56.