Lecture 6. Inference from a Random Sample

Functional Data Analysis

Jun Song

Department of Statistics
Korea University

Table of contents

Sampling Distribution (functional data)

Setting

We assume that we have an iid sample of random functions, X_n, from $\mathcal{H} = L^2[0,1]$, with $\mathrm{E}\,\|X_n\|^2 < \infty$. We will denote the common mean as μ and covariance as C.

Sample versions

The mean and covariance have sample versions:

$$\hat{\mu}(t) = \frac{1}{N}\sum_{n=1}^{N} X_n(t) \qquad \hat{c}(t,s) = \frac{1}{N}\sum_{n=1}^{N}(X_n(t) - \hat{\mu}(t))(X_n(s) - \hat{\mu}(s)).$$

Note that we can also exress the sample covariance operator as

$$\hat{C}(x) = \frac{1}{N}\sum_{n=1}^{N}\langle X_n - \hat{\mu}, x\rangle (X_n - \hat{\mu}) \qquad \text{or}$$

$$\hat{C} = \frac{1}{N}\sum_{n=1}^{N}(X_n - \hat{\mu}) \otimes (X_n - \hat{\mu}).$$

Simple lemma

If $X, Y \in \mathcal{H}$ are independent then

$$\mathrm{E}\langle X, Y\rangle = \langle \mu_x, \mu_y\rangle.$$

How can you prove this?

Convergence rates

We can then show that

$$\mathrm{E}\,\|\hat{\mu} - \mu\|^2 = O(N^{-1}).$$

What kinds of convergence does this imply?

Proof

$$
\begin{aligned}
\mathrm{E}\,\|\hat{\mu} - \mu\|^2 &= N^{-2} \sum_n \sum_m \mathrm{E}\langle X_n - \mu, X_m - \mu\rangle \\
&= N^{-1}\,\mathrm{E}\,\|X_1 - \mu\|^2 = O(N^{-1}).
\end{aligned}
$$

See if you can show that

$$
\mathrm{E}\,\|X_1 - \mu\|^2 = \sum_{i=1}^{\infty} \lambda_i .
$$

Sample Covariance

If $\mathrm{E}\,\|X_n\|^4 < \infty$, then we have

$$\mathrm{E}\,\|\widehat{C}\|_{\mathcal{S}}^2 \leq \mathrm{E}\,\|X_1\|^4$$

and

$$\mathrm{E}\,\|\widehat{C} - C\|_{\mathcal{S}}^2 \leq \frac{1}{N}\,\mathrm{E}\,\|X_n\|^4.$$

Proof 1

The first result follows from the CS inequality:

$$
\begin{aligned}
\mathrm{E}\,\|\widehat{C}\|_{\mathcal{S}}^2 &= N^{-2}\sum_n\sum_m \mathrm{E}\langle X_n \otimes X_n, X_m \otimes X_m\rangle_{\mathcal{S}} \\
&= N^{-2}\sum_n\sum_m \mathrm{E}\langle X_n, X_m\rangle_{\mathcal{S}}^2 \\
&\le N^{-2}\sum_n\sum_m \mathrm{E}[\|X_n\|^2\|X_m\|^2] \\
&\le N^{-2}\sum_n\sum_m (\mathrm{E}\,\|X_n\|^4)^{1/2}(\mathrm{E}\,\|X_m\|^4)^{1/2} \\
&= \mathrm{E}\,\|X_1\|^4.
\end{aligned}
$$

Proof 2

The second result follows from a direct calculation

$$\begin{aligned}
\mathrm{E}\,\|\widehat{C} - C\|_{\mathcal{S}}^2 &= \frac{1}{N^2}\sum_n \sum_m \mathrm{E}\langle X_n \otimes X_n - C, X_m \otimes X_m - C\rangle_{\mathcal{S}} \\
&= \frac{1}{N}\,\mathrm{E}\langle X_1 \otimes X_1 - C, X_1 \otimes X_1 - C\rangle_{\mathcal{S}} \\
&= \frac{1}{N}[\mathrm{E}\,\|X_1\|^4 - 2\,\mathrm{E}\langle X_1 \otimes X_1, C\rangle_{\mathcal{S}} + \|C\|_{\mathcal{S}}^2] \\
&= \frac{1}{N}[\mathrm{E}\,\|X_1\|^4 - \|C\|_{\mathcal{S}}^2] \leq \frac{1}{N}\,\mathrm{E}\,\|X_1\|^4.
\end{aligned}$$

Central Limit Theorem

We have already showed that both $\hat{\mu}$ and $\widehat{C}$ satisfy the CLT. In particular,

$$\sqrt{N}(\hat{\mu} - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}_{L^2[0,1]}(0, C).$$

If in addition $\mathrm{E}\,\|X_n\|^4 < \infty$ then

$$\sqrt{N}(\widehat{C} - C) \xrightarrow{\mathcal{D}} \mathcal{N}_{\mathcal{S}}(0, \Gamma).$$

What is Γ?

Estimated FPCA

As we have seen, the theoretical FPCs are the eigenfunctions of C. Therefore, the estimated FPCs, $\hat{v}_j$, are the eigenfunctions of $\widehat{C}$ with corresponding eigenvalues $\hat{\lambda}_j$. There is, however, a small technical issue. It is unclear how to determine the sign of $\hat{v}_j$, in particular, one usually assumes that

$$\hat{v}_j = \hat{c}_j \hat{v}_j \qquad \hat{c}_j = \operatorname{sign}\langle \hat{v}_j, v_j \rangle.$$

The signs $\hat{c}_j$ cannot be computed, thus any inference we make must be invariant to the choice of sign.

Convergence rates

Convergence rates established using the following operator bounds:

$$\|\hat{v}_j - v_j\|^2 \leq \frac{8\|\hat{C} - C\|_{\mathcal{L}}^2}{\alpha_j^2} \qquad |\hat{\lambda}_j - \lambda_j|^2 \leq \|\hat{C} - C\|_{\mathcal{L}}^2,$$

where $\alpha_j = \min\{\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j\}$.

Asymptotic normality

(Kokoszka and Reimherr, 2013)

The eigenelements are also asymptotically normal. In particular, one can show that

$$\sqrt{N}(\hat{\lambda}_j - \lambda_j) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2\lambda_j^2)$$
$$\sqrt{N}(\hat{v}_j - v_j) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C_j)$$
$$C_j = \sum_{k \neq j} \frac{\lambda_k \lambda_j}{(\lambda_j - \lambda_k)^2} (v_k \otimes v_k)$$

Hypothesis Testing

Hypothesis Testing

Our goal here is to evaluate hypotheses of the type

$$H_0 : \mathrm{E}[X_n] = \mu = \mu_0 \qquad \text{vs} \qquad \mu \neq \mu_0.$$

Unlike in the scalar case, there is not a "best" way to do this (or at least the community hasn't settled on one yet).

Multivariate setting

Suppose that $X_1, \ldots, X_n$ were iid random vectors in $\mathbb{R}^p$ with mean μ and covariance matrix $\mathbf{\Sigma}$. The classic way to test if $\mu = \mu_0$ is to use Hotelling's T^2 test (equivalent to the likelihood ratio):

$$T^2 = N(\bar{X} - \mu_0)^\top \mathbf{\Sigma}^{-1}(\bar{X} - \mu_0).$$

Under H_0, $T^2 \sim \chi^2_p$, while under H_A, $T^2 \xrightarrow{\mathcal{P}} \infty$. Why won't this generalize to a functional setting?

Functional problems

- Defining inverse of C is challenging ($p = \infty$ in Hilbert space)

We have to try something else. There are two common approaches to doing this, though this is still a fairly open area.

General Strategy

Under $H_0 : \mu = \mu_0$, we know from the KL expansion that

$$\langle \sqrt{N}(\hat{\mu} - \mu), v_i \rangle \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda_i),$$

and that they are asymptotically independent (since they are uncorrelated). So our strategies can all be thought of as some weighted combination of these projections.

FPCA approach

The first approach is to only test the first few principal component directions:

$$T^2_{PC} = \sum_{i=1}^{p} \frac{N\langle \hat{\mu} - \mu_0, \hat{v}_i \rangle^2}{\hat{\lambda}_i}$$

where $i = 1, \ldots, p$. One can show that, under H_0 we have $T^2_{PC} \xrightarrow{\mathcal{D}} \chi^2(p)$, while under H_A one has that $T^2_{PC} \xrightarrow{\mathcal{P}} \infty$ as long as $\langle \mu - \mu_0, v_i \rangle \neq 0$ for some $i \leq p$.

This strategy works, but it can be very sensitive to the choice of p.

Norm approach

The second approach is to test all of the FPC directions, but not to normalize by the eigenvalues:

$$T^2_{norm} = N\|\hat{\mu} - \mu_0\|^2 = \sum_{i=1}^{\infty} N\langle \hat{\mu} - \mu_0, \hat{v}_i\rangle^2$$

Note that the sum will always terminate at some finite value since the $\hat{\lambda}_i$ will be zero at some point

One can show that, under H_0 we have $T^2_{norm} \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \lambda_i \chi^2_i(1)$, while under H_A one has that $T^2_{norm} \xrightarrow{\mathcal{P}} \infty$.

Norm approach

- This strategy works and has the advantage of not having to choose p.
- Sometimes it is more powerful than the PC approach, sometimes not.
- Note that to compute quantiles from a weighted sum of chi-squares is not very standard, but one can use the `CompQuadForm` package in `R` to do the needed calculations.

Choi Test

(Choi and Reimherr, JRSSB, 2018)

This is not in the book, but there is a middle ground between the two approaches. In particular, the idea is to use

$$T^2_{Choi} = \sum_{i=1}^{p} \frac{N\langle \hat{\mu} - \mu_0, \hat{v}_i \rangle^2}{\hat{\lambda}_i^{1/2}}.$$

Under H_0 we have that $T^2_{Choi} \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \lambda_i^{1/2} \chi_i^2(1)$ while under H_A $T^2_{Choi} \xrightarrow{\mathcal{P}} \infty$. This compromises between the two tests, but requires that $\sum \lambda_i^{1/2}$ be finite (which is usually true).

Proof for T_{PC}

- For the T_{PC} test we require that $\lambda_1 > \cdots > \lambda_p > \lambda_{p+1}$, i.e. that the first p eigenvalues be distinct.
- Recall that by the CLT $\sqrt{N/\lambda_i}\langle\hat{\mu} - \mu_0, v_i\rangle \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, then convergence under H_0 follows from a continuous mapping theorem.

Proof for T_{PC}

Under H_A we have that

$$\hat{\mu} - \mu_0 = \hat{\mu} - \mu + (\mu - \mu_0) = O_P(N^{-1/2}) + \Delta.$$

So we have that

$$T_{PC} = \sum_{i=1}^{p} \frac{N\langle \Delta + O_P(N^{1/2}), \hat{v}_i \rangle^2}{\hat{\lambda}_i},$$

$$= \sum_{i=1}^{p} \frac{N\langle \Delta, \hat{v}_i \rangle^2}{\hat{\lambda}_i} + O_P(N^{1/2})$$

which tends to infinity (again by Slutky's) as long as $\langle \Delta, v_i \rangle \neq 0$ for some i.

Proof for T_{norm}

Under H_0, by the continuous mapping theorem we have

$$N\|\hat{\mu} - \mu_0\|^2 \xrightarrow{\mathcal{D}} \|Z\|^2 \qquad \text{where } Z \sim \mathcal{N}(0, C).$$

Applying the KL expansion + Parcevals to Z we get that

$$\|Z\|^2 = \sum \langle Z, v_i \rangle^2 \sim \sum \lambda_i \chi_i^2(1).$$

Confidence Bands/Regions

Confidence Bands

- Lastly, before jumping into examples, we will present a strategy for constructing confidence bands. This means that we want to construct a function $r(t)$ such that

$$P(r(t) \leq \hat{\mu}(t) - \mu(t) \leq r(t)) \geq 1 - \alpha,$$

for some significance level α.

- Notice that inside the probability we can only require this to hold for almost all t since we are in $L^2[0,1]$.
- While we prefer an exact equality, we often have to settle for bands which are a bit conservative (wide).

Bootstrap approach

Recall that in using a Bootstrap, one simulates new samples from the estimated distribution of the data. Using these new samples, one can obtain an estimated distribution for your desired test statistic.

- Based on the empirical distribution: non parametric bootstrap
- Based on some probability model (e.g. Gaussian): parametric bootstrap

Bootstrap approach

In our case, we aim to generate, say X_{nb} for $n = 1, \ldots, N$ and $b = 1, \ldots, B$. Each X_{nb} is generated as

$$X_{nb}(t) = \hat{\mu}(t) + \varepsilon_{nb}(t).$$

In nonparametric bootstrap one would sample ε_{nb} randomly from the original residuals $\hat{\varepsilon}_n = X_n - \hat{\mu}(t)$. In a parametric bootstrap, one would generate them from a $\mathcal{N}(0, \widehat{C})$.

Bootstrap approach

So how do we use these ideas for a confidence band? Well, we will form a band as

$$\hat{\mu}(t) \pm \sqrt{\widehat{C}(t,t)}c_{1-\alpha}$$

and we will use the bootstrap samples to approximate the value of $c_{1-\alpha}$. Note that one cannot compute it in a closed form, simulation has to be used. We use

$$D_b = \sup \frac{|\hat{\mu}(t) - \hat{\mu}_b(t)|}{\sqrt{\widehat{C}(t,t)}},$$

we then take $c_{1-\alpha}$ as the $1-\alpha$ quantile of the D_b.

Choi's Band

It turns out that Choi's test also leads to a band. In particular we can take

$$r(t) = N^{-1/2}\sqrt{c_N(\alpha)\sum \lambda_i^{1/2} v_i(t)^2},$$

where $c_N(\alpha)$ is the upper α quantile of a weighted sum of chi-squares $\sum \lambda_i^{1/2}\chi_i^2(1)$.

Choi's Band

This follows from the Cauchy-Schwarz inequality and the KL expansion:

$$N(\hat{\mu}(t) - \mu(t))^2 = \left(\sum Z_i \lambda_i^{1/2} v_j(t)\right)^2 \leq \sum Z_i^2 \lambda_i^{1/2} \sum \lambda_i^{1/2} v_j(t)^2.$$

Simulation

Simulation

Here we will simulate data and estimate the various parameters. We will illustrate how to carry out hypothesis testing and confidence bands.

Setting

We will simulate data according to

$$X_n(t) = \mu(t) + \varepsilon_n(t),$$

where $\varepsilon_n(t)$ are iid Gaussian processes with either one of two covariances:

$$\exp\{10|t-s|\} \qquad \text{or} \qquad \exp\{10|t-s|^2\}.$$

We will simulate the mean according to

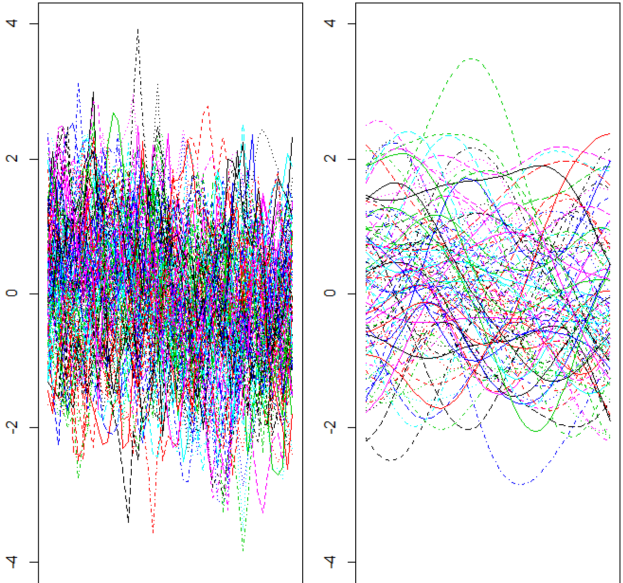
$$\mu(t) = a\cos(\pi t),$$

where we will take a view values of a. For testing, the null will be $H_0 : a = 0$.

R Code

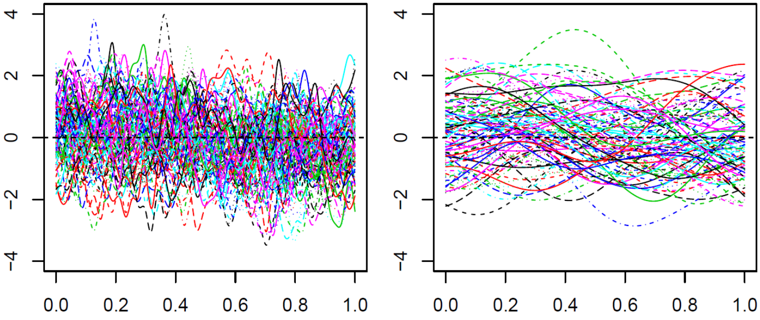
```
N<-100; M = 50; a = 1/4
pts<-seq(from=0,to=1,length=M)
mu<-function(t){a*cos(pi*t)}
mu.vec<-mu(pts)

d<-outer(pts,pts,FUN="-")
Sig1<-exp(-abs(d)*10)
Sig2<-exp(-d^2*10)

X1<-mvrnorm(N,mu.vec,Sig1)
X2<-mvrnorm(N,mu.vec,Sig2)
```

Plots



Convert to FD

```
mybasis<-create.bspline.basis(c(0,1),nbasis=50)
X1.f<-Data2fd(pts,t(X1),mybasis)
X2.f<-Data2fd(pts,t(X2),mybasis)
```



Norm Test

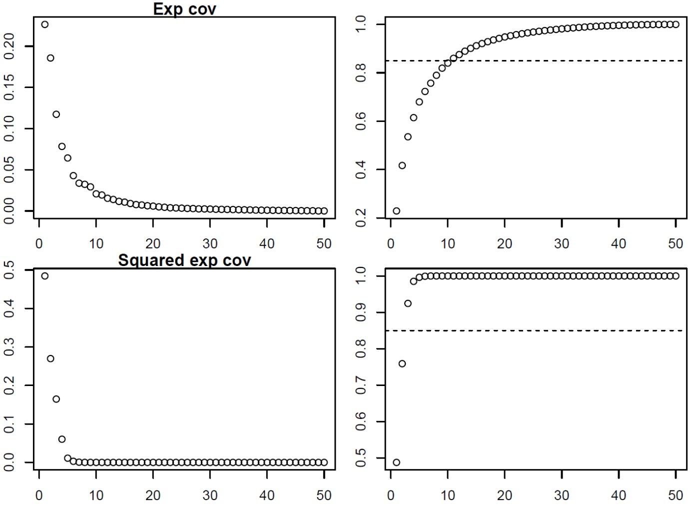
```
mu_f1<-mean.fd(X1.f)
mu_T_f<-Data2fd(pts,rep(0,length=M),mybasis)
Delta1 = mu_f1 - mu_T_f
T_norm1 = N*inprod(Delta1,Delta1)

mu_f2<-mean.fd(X2.f)
Delta2 = mu_f2 - mu_T_f
T_norm2 = N*inprod(Delta2,Delta2)
```

Norm P-value - Imhof

```r
library(CompQuadForm)
pca_fd1<-pca.fd(X1.f,nharm=10)
lambdas1 <- pca_fd1$values
imhof(T_norm1,lambdas1,epsabs=1e-10)$Qq

## [1] 3.634495e-06

pca_fd2<-pca.fd(X2.f,nharm=3)
lambdas2 <- pca_fd2$values
imhof(T_norm2,lambdas2)$Qq

## [1] 0.03405214

# There are other options in the package
# if you are interesed.
```

PCA Method

First we have to decide how many FPCs to use.



PCA Method

So, for the exponential covariance, we need about 10 FPCs, while for the squared exponential we only need 3.

```
T_PC_1<-N*sum(inprod(Delta1,
  pca_fd1$harmonics[1:10])^2/lambdas1[1:10])
pchisq(T_PC_1,10,lower.tail=FALSE)

## [1] 1.620017e-05

T_PC_2<-N*sum(inprod(Delta2,
  pca_fd2$harmonics[1:3])^2/lambdas2[1:3])
pchisq(T_PC_2,3,lower.tail=FALSE)

## [1] 0.02564701
```

Simulation - Type 1 Error (reject true H_0)

We repeat the discussed procedures with a zero mean and compare the different tests. Repeat 1000 times.

	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.01$
Norm - Exp	0.10	0.06	0.01
Norm - SqExp	0.12	0.06	0.01
PC - Exp	0.10	0.05	0.01
PC - SqExp	0.11	0.07	0.01

Simulation - Power (reject false H_0 correctly)

We repeat the discussed procedures with $a = 1/4$ and compare the different tests. Repeat 100 times.

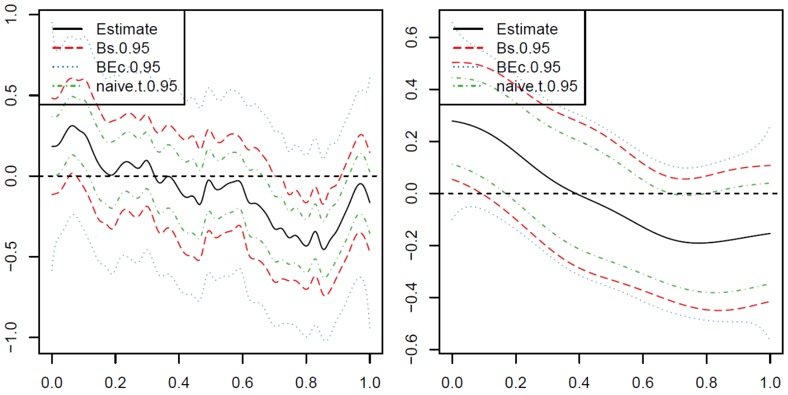
	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.01$
Norm - Exp	0.98	0.97	0.88
Norm - SqExp	0.81	0.72	0.43
PC - Exp	0.96	0.91	0.82
PC - SqExp	0.85	0.74	0.54

Simulation - SCB

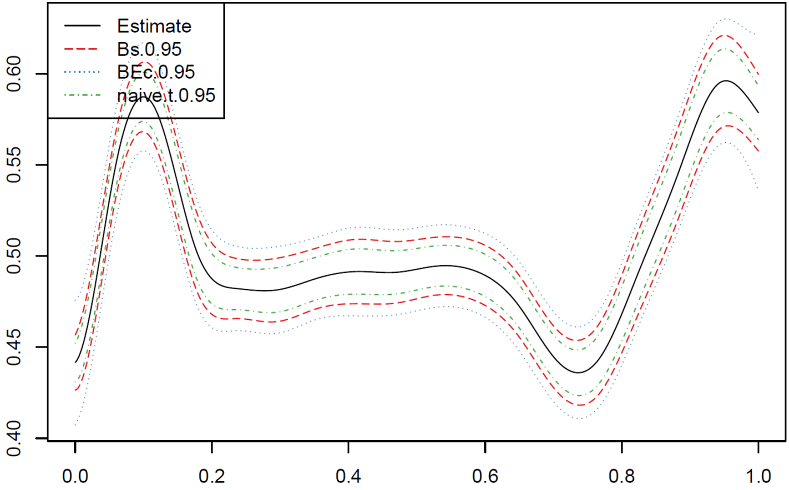
We can fit the simultaneous confidence bands using Choi's package from Github.

```
# require(devtools); install_github("hpchoi/fregion")

library(fregion)
mu_f1<-mean.fd(X1.f)
mu_f2<-mean.fd(X2.f)
C_1<-var.fd(X1.f)
C_2<-var.fd(X2.f)
ty<-c("BEc","Bs","naive.t")
conf_band1<-fregion.band(mu_f1,C_1,N=N,type=ty)
conf_band2<-fregion.band(mu_f2,C_2,N=N,type=ty)
```

Simulation - SCB



DTI


Estimate
Bs.0.95
BEc.0.95
naive.t.0.95
0.40
0.45
0.50
0.55
0.60
0.0
0.2
0.4
0.6
0.8
1.0

Final thoughts on bands

Bands are very useful visualization techniques. This is an ongoing area of research, and it seems impossible to find a method that is both exact (i.e. not conservative) and fast. However, for smoother estimates, the Choi bands are very close to the bootstrap ones.