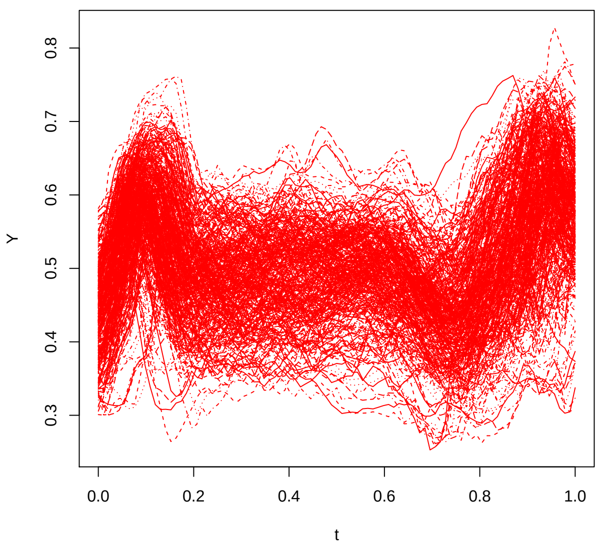
Functional and Shape Data Analysis
Weekly assignments 2 autumn 2019

Solutions for the following questions should be returned to Moodle's quiz platform. Platform will be opened within couple of days. The last return time is 24 January at 12.00.

1. Diffusion tensor imaging, DTI, is a magnetic resonance imaging methodology which is used to measure the diffusion of water in the brain. Water diffuses isotropically (i.e. the same in all directions) in the brain except in white mater where it diffuses anisotropically (i.e. differently in different directions). This allows researchers to utilize DTI to generate images of white matter in the brain. We consider fractional anisotropy tract profiles of the corpus callosum, a portion of the data DTI data set in the refund R package. The corpus collosum, the largest white matter structure in the brain, is a bundle of nerve fibers connecting the two hemispheres of the brain. Fractional anisotropy is a value between 0 and 1 which measures the level of anisotropy, and therefore the quantity of white matter, at a particular location. A total of 376 patients are considered, with each tract measured at 93 equally spaced locations. The arguments t_{ij} thus denote a spatial rather than temporal location. The resulting observed curves are shown in figure below.

```
> library(refund)
> data(DTI)
> Y<-t(DTI$cca[-c(125,126,130,131,319,321),])  # removal of missing values and transpose of the data
> t<-seq(0,1,length=dim(Y)[1])
> Y[1:5,1:10]
        1001_1    1002_1    1003_1    1004_1    1005_1    1006_1    1007_1    1008_1    1009_1    1010_1
cca_1  0.4909345 0.4721627 0.5023738 0.4021894 0.4018747 0.4507296 0.5537167 0.4477326 0.4953400 0.4580893
cca_2  0.5168018 0.4868219 0.5136516 0.4225127 0.4055805 0.4535052 0.5577723 0.4807621 0.5057313 0.4705768
cca_3  0.5356539 0.5022577 0.5392542 0.4398983 0.3985487 0.4575306 0.5604173 0.5024963 0.5176728 0.4796741
cca_4  0.5553587 0.5233635 0.5742101 0.4600235 0.3860009 0.4655026 0.5710832 0.5181574 0.5409608 0.4886211
cca_5  0.5927610 0.5524401 0.6031339 0.4751297 0.4088248 0.4804154 0.5848086 0.5304736 0.5834464 0.5056029
>matplot(t,Y, type="n")
>matlines(t,Y, col="red")
```



(a) Convert the Y dataset to functional objects by using $K = 10$ B-splines of order four (cubic splines) and knots with equal intervals, and plot the 376 smoothed curves on one graph. Use t as the "time" variable, and do the smoothing by using the least squares method to estimate coefficient vectors $\mathbf{c}_i$.

(b) Convert the Y dataset to functional objects by using $K = 100$ B-splines of order four (cubic splines) and knots with equal intervals, and plot the 376 smoothed curves on one graph. Use t as the "time" variable, and do the smoothing by using the penalized least squares method with roughness penalty being the square of the second derivative $[D^2 x_i(t)]^2$ to estimate coefficient vectors $\mathbf{c}_i$. You may use the value $\lambda = 0.01$ as the weight to the roughness penalty.

(c) Find the "best" smoothing parameter value λ when everything else are kept same as in (b). Use the generalized cross validation as the criteria for the "best" λ value, and start your search with λ set as $\lambda = 0.00000001$.

2. Let us continue with the diffusion tensor imaging example given in the dataset Y.

(a) After obtaining smoothed functionals $\hat{x}_i(t)$ in previous question, construct the sample mean function

$$\bar{x}(t) = \frac{\sum_{i=1}^{N} \hat{x}_i(t)}{N},$$

where N is the amount of sampling units in the data. What is the sample mean at $t = 0.75$, i.e., what is $\bar{x}(0.75)$?

(b) Construct also the sample standard deviation function

$$\hat{\sigma}(t) = \sqrt{\frac{\sum_{i=1}^{N} [\hat{x}_i(t) - \bar{x}(t)]^2}{N-1}}.$$

(Note that in lecture slides notations are bit different. Here used notations are more in line with the book of Ramsay & Silverman.) What is the sample standard deviation at $t = 0.5$, i.e., what is $\hat{\sigma}(0.75)$?

(c) Carry out a functional principal component analysis for smoothed functionals $\hat{x}_i(t)$. Find out coefficient estimates $\hat{\mathbf{c}}_{\xi_j}$ for the first three eigenfunctions (harmonics) $\xi_1(t), \xi_2(t), \xi_3(t)$. Particularly, try to reconstruct functionals $\hat{x}_i(t)$ as

$$\hat{x}_i(t) \approx \bar{x}(t) + \sum_{j=1}^{3} f_{ij} \xi_j(t) = \bar{x}(t) + \sum_{j=1}^{3} f_{ij} \boldsymbol{\phi}(t)' \hat{\mathbf{c}}_{\xi_j},$$

where f_{ij} are principal component scores, and $\boldsymbol{\phi}(t)$ is a vector of basis functions. What is the value $\bar{x}(0.75) + \sum_{j=1}^{3} f_{1j} \xi_j(0.75)$ for the sampling unit $i = 1$?

3. Will be published later ...