PEC 7 - Function approximation and regression (I)

Ricard Santiago Raigada García

Date: 24/05/2023

Contents

1 Integration

1.1 Generate samples

Use the entire historical series to generate samples of the random variable representing the variation in average temperature between two consecutive days. The values of the samples corresponding to the last twelve days should be displayed.

Solución:

To calculate the variation in average temperature between two consecutive days, you can consider the vector of historical temperatures. Simply subtract the consecutive elements of the temperature vector:

$$\Delta t_i = t_{i+1} - t_i, \quad i = 1, 2, \ldots, n-1$$

Where n is the length of the vector t.

The samples from the last twelve days are:

$$[-1.0, 0.2, 3.0, 0.9, 0.0, -2.7, -1.8, -0.4, -0.5, -1.6, 4.0, 3.4]$$

```r
# Clear the workspace
rm(list = ls())

####### Funciones auxiliares #######
source('Lectura_datos_por_fecha.R')

# Funcion que evalua los polinomios trigonometricos de '-n' hasta 'n' en 'x'
my_fourier_exp = function(x, n){

  expy = matrix(0, 2*n+1, length(x))
  for (k in -n:n) {
    expy[k+n+1, ] = exp(1i*k*x) #TODO expresi n de las funciones base de Fourier (formulacion exponencial)
  }

  return(expy)
}

# Dados los coeficients de Fourier, retorna la funci n aproximada evaluada en 't'
myf_Fourier_exp = function(t, a, b, ck){

  x = 2*(pi*(t - a)/(b-a))-pi #TODO cambio de variable
  n = floor(length(ck)/2)

  expx = my_fourier_exp(x, n)

  return(Re(colSums(ck*expx))) # NO MODIFICAR

```

```
}

####### PEC 7 #######

## Lectura de datos
#TODO cubrir los campos para recuperar las temperaturas
t = myReadData_byDate('0200E_19200101_20181231.csv', '01/01/ 1926', '31/12/ 2018', 'TMEDIA')
m = length(t)

####### Ejercicio 1 #######
## Generar muestras de variacion de temperatura
 #TODO variaci n de temperatura entre dos dias consecutivos
delta_t = t[2:m] -t[1:m-1]

tail(delta_t,12)
```

Listing 1: R code

1.2 Approximating the probability density function

Based on the samples of the random variable obtained in the previous section, approximate the corresponding probability density function using trigonometric polynomials (Fourier series) with $n = 2$. Since the samples from the previous section are not in the applicable domain of these polynomials, i.e., $(-\pi, \pi)$, it will be necessary to perform a change of variables for their use. Display the obtained coefficients.

Solution:

The function approximation is based on a linear combination of basis functions (trigonometric polynomials). The Fourier series is used to approximate the temperature, which can be approximated as follows for $f(x)$:

$$f(x) \approx \hat{f}(x) = \sum_{j=0}^{n} c_j \phi_j(x)$$

where $\phi_j(x)$ are the trigonometric basis functions and c_j are the corresponding coefficients. To obtain the coefficients c_j, we solve the following system of equations:

$$\langle f, \phi_j \rangle = c_0 \langle \phi_j, \phi_0 \rangle + c_1 \langle \phi_j, \phi_1 \rangle + \ldots + c_n \langle \phi_j, \phi_n \rangle$$

The solution for the coefficients c_j is obtained using the formula:

$$c_j = \frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle}$$

Once the coefficients are obtained, the approximation $\hat{f}(x)$ can be used to estimate the distribution of the daily temperature variation.

First, the change of variable needs to be performed.

$$x = \frac{2\pi(\Delta t - a)}{b - a} - \pi$$

Where a and b are the minimum and maximum values of the temperature variations, respectively. In the code, these can be obtained using the functions min() and max().

$$a = -12$$

$$b = 8.1$$

Next, we calculate the derivative of x, dx, which is used in the calculation of the dot product. Adapting the generic description in the statement to the Fourier series, we have:

$$\langle f, \phi_j \rangle = \int_{-\pi}^{\pi} f(x)\bar{\phi}_j(x)dx$$

where $\bar{g}$ denotes the conjugate of g. Additionally, $\langle \phi_j, \phi_j \rangle = 2\pi$.

Considering $\phi_j(x) = e^{ijx}$ and the change of variable:

$$c_j = \frac{1}{2\pi} \int_a^b f\left(\frac{2\pi(\Delta t - a)}{b - a} - \pi\right) e^{-ij\left(\frac{2\pi(\Delta t - a)}{b-a} - \pi\right)} dt$$

Mediante la media aritmética de las muestras:

$$c_j \approx \frac{1}{b-a}\frac{1}{m}\sum_{i=1}^{m} e^{-ij\left(\frac{2\pi(\Delta t_i - a)}{b-a} - \pi\right)}$$

The approximation of the density function is:

$$\hat{f}(x) = \sum_{j=-2}^{2} c_j e^{ij\left(\frac{2\pi(\Delta t - a)}{b-a} - \pi\right)}$$

According to the statement, $n = 2$, and the obtained coefficients are:

$$c_{-2} = 0.00710249 + 0.02735364i, \quad c_{-1} = 0.03473345 + 0.02494534i,$$

$$c_0 = 0.04975124{+}0.0i, \quad c_1 = 0.03473345{-}0.02494534i, \quad c_2 = 0.00710249{-}0.02735364i$$

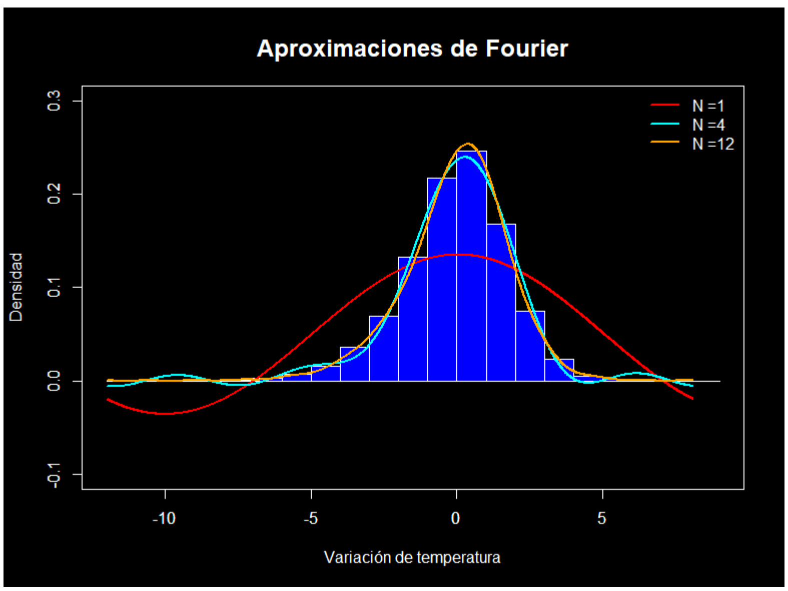
```r
####### Ejercicio 2 #######
## Aproximaci n densidad mediante series de Fourier (formulacion con exponenciales)
# Truncamiento a 2n+1 t rminos
n = 2

# Cambio de variable
#TODO implementar el cambio de variable
a = min(delta_t)
b = max(delta_t)
x = 2*(pi*(delta_t -a)/(b-a))-pi

# Calculo de coeficientes (caso continuo)
expx = my_fourier_exp(-x, n) #TODO completar el primer argumento (fijarse en la expresion (26) de la guia)
ck = (1/(b - a))*rowMeans(expx) # NO MODIFICAR

print(ck)
```

Listing 2: R code

1.3 Graphical representation

Progressively increase the number of coefficients in the polynomials (up to $n = 12$). Plot three of the obtained solutions for $n = 1$, $n = 4$, and $n = 12$, on a histogram based on the data, to observe the improvement in the approximation as the number of terms in the expansion increases.

Solution:



It can be observed that the higher the value of N, the better the fit. However, there is a point where the improvement becomes limited.

```r
####### Ejercicio 3 #######

#TODO implementaci n de la representacion grafica de varias aproximaciones. Para ello, se puede emplear la funcion 'myf_Fourier_exp'
N <- c(1, 4, 12)
colors <- c("#FF0000", "#00FFFF", "#FFA500")
par(bg = "black", col.axis = "white")
hist(delta_t, freq = FALSE, ylim = c(-0.1, 0.3), col = "blue", border = "white", xlab = "", ylab = "")
hh <- 0.01
tt <- seq(a, b, hh)

for (i in 1:length(N)) {
  n <- N[i]
  expx <- my_fourier_exp(-x, n)
  ck <- (1 / (b - a)) * rowMeans(expx)
```

```
  myf_F_exp <- myf_Fourier_exp(tt, a, b, ck)

  lines(tt, myf_F_exp, col = colors[i], lwd = 2.5)
}

legend("topright", legend = paste0("N =", N), col = colors, lwd = 2.5, text.col = "white")
box(col = "white")
axis(1, col = "white", lwd = 1.5)
axis(2, col = "white", lwd = 1.5)

title(xlab = "Variaci n de temperatura", ylab = "Densidad", col.lab = "white")
main_title <- "Aproximaciones de Fourier"
title(main = main_title, col.main = "white", cex.main = 1.5)
```

Listing 3: R code

1.4 Approximate the distribution function and calculate probabilities

Using the approximation of the density function (with $n = 12$), use a numerical integration method (PEC 6) to solve the integral involved in the expression of the distribution function and obtain an approximation for it. With this, calculate the following probabilities:

- Probability that the temperature decreases by at least 2 degrees in one day.
- Probability that the temperature varies by less than 3 degrees from one day to the next.
- Probability that tomorrow's temperature is at least 3 degrees higher than today's temperature.

Solution:

- The probability that the temperature decreases by at least 2 degrees in one day is 0.1260297.
- The probability that the temperature varies by less than 3 degrees from one day to the next is 0.9438068.
- The probability that tomorrow's temperature is at least 3 degrees higher than today's temperature is 0.03224461.

```
####### Ejercicio 4 #######
#NOTA La funcion 'myf_Fourier_exp' puede ser de ayuda

#TODO Probabilidad de que la temperatura disminuya al menos 2 grados en un d a.
print(integrate(myf_Fourier_exp, a, -2.0, a, b, ck)$value)

#TODO Probabilidad de que temperatura var e menos de 3 grados.
print(integrate(myf_Fourier_exp, a, 3.0, a, b, ck)$value - integrate (myf_Fourier_exp, a, -4.0, a, b, ck)$ value)

#TODO Probabilidad de que la temperatura de ma aana sea al menos 3 grados mayor que la de hoy.
print (1 - integrate(myf_Fourier_exp, a, 3.0, a, b, ck)$value)
```

Listing 4: R code

2 References

Here are the references consulted to solve the statement:

Theoretical material from UOC. Book: "Computational Methods for Numerical Analysis with R."