PAC 7 - Aproximació de funcions i regressió (I)

Data de lliurament: 24/05/2023

Universitat Oberta de Catalunya

Descripció del problema

El clima i la meteorologia són objectes d'estudi molt importants ja que afecten directament el desenvolupament de les activitats d'una societat. El seu tractament es basa a analitzar els diferents fenòmens meteorològics que són observables i mesurables. És evident que aquests fenòmens estan fortament relacionats entre ells, comptant al seu torn amb un gran component estacional. Aquestes característiques impliquen que el seu tractament matemàtic sigui altament complex.

Les agències de meteorologia han estat recaptant dades de diversos fenòmens en els últims temps. Per tant, una gran quantitat d'informació heterogènia està disponible públicament per al seu tractament. Basada en ella es poden desenvolupar models estadístics que expliquin el comportament històric d'aquests fenòmens, així com simular i predir possibles escenaris futurs.

Disposeu de dades històriques de mesuraments de diversos fenòmens des del 01/01/1926 fins al 31/12/2018. En particular, per a aquesta activitat s'utilitzen les dades recollides en l'estació meteorològica de *Barcelona, Fabra*. Se us proporciona el fitxer *0200E-19200101-20181231.csv*. Els dies (amb dades disponibles) es numeraran de manera absoluta, ascendent i consecutiva, sent el dia 01/01/1926 el dia 1. Per a importar les dades se us proporciona un *script* de R, que facilita la correcta lectura (per data) del fitxer anterior.

Variació diària de temperatura

Una de les prediccions meteorològiques més utilitzades en la pràctica és la variació de temperatura que es produeix entre dos dies consecutius. En aquesta situació, l'efecte de l'estacionalitat es dilueix ja que no es consideren les temperatures absolutes sinó la diferència entre elles. En aquest exercici es pretén obtenir les probabilitats que la variació tèrmica entre dos dies es trobi en determinades situacions d'interès. Per a això serà molt útil realitzar una aproximació o estimació (basada en les dades històriques) de la funció de densitat de la variable aleatòria de variació de temperatures. Com s'ha vist en la guia, una funció qualsevol es pot aproximar per

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

on els coeficients, c_j, s'obtenen resolent el sistema d'equacions normals,

$$\begin{aligned}
\langle \phi_0, f(x) \rangle &= c_0 \langle \phi_0, \phi_0 \rangle + c_1 \langle \phi_0, \phi_1 \rangle + ... + c_n \langle \phi_0, \phi_n \rangle \\
\langle \phi_1, f(x) \rangle &= c_0 \langle \phi_1, \phi_0 \rangle + c_1 \langle \phi_1, \phi_1 \rangle + ... + c_n \langle \phi_1, \phi_n \rangle \\
&\vdots \\
\langle \phi_n, f(x) \rangle &= c_0 \langle \phi_n, \phi^0 \rangle + c_1 \langle \phi_n, \phi_1 \rangle + \cdots + c_n \langle \phi_n, \phi_n \rangle.
\end{aligned}$$

Quan les funcions base $\phi_j(x)$ són ortogonals, la solució del sistema anterior se simplifica i el càlcul dels coeficients, c_j, es redueix a

$$c_j = \frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle}, \quad j = 0, 1, 2, \ldots, n.$$

El càlcul de $\langle \phi_j, \phi_j \rangle$, ve determinat per les bases seleccionades (veure guia de l'activitat). El terme, $\langle f, \phi_j \rangle$, es pot obtenir aplicant la definició de producte escalar,

$$\langle f, \phi_j \rangle = \int_{x_1}^{x_2} f(x) \phi_j(x) w(x) \mathrm{d}x,$$

on, com és aquest cas, f és una **funció de densitat**, es té la igualtat

$$\langle f, \phi_j \rangle = \int_{x_1}^{x_2} f(x) \phi_j(x) w(x) \mathrm{d}x = \mathbb{E} \left[\phi_j(X) w(X) \right],$$

sent X una variable aleatòria. Donats els elements d'una mostra d'una variable aleatòria, $X_i, i = 1, 2, \ldots, m$, l'esperança matemàtica es pot aproximar mitjançant la mitjana aritmètica

$$\mathbb{E} \left[\phi_j(X) w(X) \right] \approx \frac{1}{m} \sum_{i=1}^{m} \phi_j(X_i) w(X_i).$$

La funció de distribució es pot calcular (aproximar) a partir de la funció de densitat, aplicant la seva definició en forma integral,

$$F(t) := \int_{-\infty}^{t} f(x) \mathrm{d}x \approx \int_{t_{\min}}^{t} \hat{f}(x) \mathrm{d}x,$$

on $t_{\min}$ és el valor mínim dels elements de la mostra donats. En el segon pas s'ha utilitzat l'aproximació de la funció de densitat obtinguda anteriorment.

1 Generar la mostra

Utilitzar tota la sèrie històrica per a generar la mostra de la variable aleatòria de variació de temperatura mitjana entre dos dies consecutius. S'han de mostrar els elements de la mostra corresponents als deu últims dies.

2 Aproximar la funció de densitat

Basant-se en la mostra de la variable aleatòria obtinguda en l'apartat anterior, aproximar la funció de densitat corresponent emprant polinomis trigonomètrics (sèries de Fourier) amb $n = 2$. Com la mostra de l'apartat anterior no estarà en el domini d'aplicació d'aquests polinomis, és a dir $(-\pi, \pi)$, serà necessari realitzar un canvi de variables per a la seva utilització. Mostrar els coeficients obtinguts.

3 Representar gràficament

Augmentar progressivament el nombre de coeficients en els polinomis (fins a $n = 12$). Representar tres de les solucions obtingudes per a $n = 1$, $n = 4$ i $n = 12$, sobre un histograma basat en les dades, per a observar la millora en l'aproximació en augmentar el nombre de termes en l'expansió.

4 Aproximar la funció de distribució i calcular probabilitats

A partir de l'aproximació de la funció de densitat (amb $n = 12$), utilitzar un mètode d'integració numèrica (PAC 6) per a resoldre la integral involucrada en l'expressió de la funció de distribució i obtenir una aproximació per a aquesta. Amb ella, calcular les següents probabilitats:

- Probabilitat que la temperatura disminueixi almenys 2 graus en un dia.
- Probabilitat que la temperatura varïi menys de 3 graus d'un dia al següent.
- Probabilitat que la temperatura de demà sigui almenys 3 graus major que la d'avui.

Notes

- En la guia s'explica com fer un canvi de variable per a transformar valors en un interval (a, b) a un interval $(-1, 1)$. Una vegada fet aquest canvi, per a treballar en $(-\pi, \pi)$, bastaria amb multiplicar per π.
- Per a fer la representació de l'histograma, es pot utilitzar el comando *hist* de R.
- Donada una variable aleatòria X, se sap que la probabilitat $P(X \leq x) = F(x)$, on F és la seva funció de distribució. Així mateix, també es compleix que $P(X > x) = 1 - F(x)$ i $P(a < X \leq b) = F(b) - F(a)$.
- En l'exercici 4, es pot utilitzar el comando de R *integrate* per a comprovar el correcte funcionament del codi d'integració numèrica emprat.

Criteris de correcció i puntuació dels apartats

Aquesta PAC tindrà un valor de **10 punts** repartits com segueix:

- Exercici 1: La correcta generació de la variable de variació de temperatura entre dies consecutius es valorarà amb 1 punt. Total: 1 punt.
- Exercici 2: La implementació de l'aproximació d'una funció de densitat mitjançant sèries de Fourier es valorarà amb 2 punts. El canvi de variable es valorarà amb 0.5 punts. Cada coeficient calculat correctament es valorarà amb 0.5 punts. Total: 5 punts.
- Exercici 3: Les representacions gràfiques de les densitats aproximades es qualificaran amb 1 punt. Total 1 punt.
- Exercici 4: La utilització adequada de la integració numèrica per a l'avaluació de la funció de distribució es qualificarà amb 1.5 punts. El càlcul correcte de les probabilitats demanades es valorarà amb 0.5 punts cadascuna. Total 3 punts.

Referències

[1] Howard, J. P. (2017). Computational methods for numerical analysis with R. Nueva York: Chapman & Hall/CRC.

[2] Leitao Rodríguez, A., Salvador Mancho, B., Sancho Vinuesa, T. (2022). Aproximació de funcions i regressió.