

NDËRTIMI I GRAFIKËVE 3-DIMENSIONALË NË MATLAB PËR SHPËRNDARJEN E DENSITETEVE PROBABILITARË

*SAATÇIU D.¹, DIBRA S.²

¹Universiteti i Tiranës, Fakulteti i Shkencave të Natyrës, Departamenti i
Informatikës

²ikubINFO

e-mail: denis.saatciu@fshn.edu.al

Përmbledhje

Në këtë punim trajtohet procesi i algoritimit dhe i kodimit në gjuhën matlab për ndërtimin grafikëve 3-dimENSIONALË. Paketa përdor algoritmet optimalë për vrojtimin e marrëdhënieve kovaracionale dhe sjelljeve të ndërsjella të vektorëve të rastit shumë dimensionalë. Vrojtimi i këtyre marrëdhënieve realizohet nëpërmjet tenzorëve të vlerave të përlogaritura të densiteteve të vlerësuar dhe të vizualizimit nëpërmjet një aparati grafik të ndërtuar për këtë rast. Paketa lexon të dhënat nga një skedar i jashtëm. Programi pranon një numër çfarëdo kolonash (variablash) dhe për secilën prej tyre përlogarit vlerat minimale dhe maksimale për përdorim të mëpasshëm. Matrica e formuar nga vlerat e përlogaritura të densiteteve të pozuar përdoret për të llogaritur tenzorin $n + 1$ dimensional të variablove dhe vlerave të funksionit të vlerësuar. Realizimi i prerjeve $n - 2$ dimensionale të dhëna më parë, të çfarëdoshme, është bërë i mundur për të dhënë paraqitje grafike të shpërndarjeve të të dhënave të pozuar. Paketa realizon ndërtim grafikësh 3 dimensionalë dhe rrotullimin e tyre horizontal e vertikal për lehtësi vrojtimi.

Abstract

In this article we will present algorithmic and coding process in MATLAB to build 3-dimensional graphs. The package use improved algorithms to observe covariational relationships and mutual behaviors of multidimensional arrays. Observing these relationships will be possible through tensors of estimated densities calculated values and their visualization on a graphical ambient build for this scope. The software can read data from a file. It accepts a variable number of columns(variables) and for each of those it calculate minimum and maximum values for further use. The formed matrix from calculated values of given densities is used to produce the $n+1$ dimensional tensor of the variables and the produced function values. Realization of any $n-2$ dimensional cuts is used to give a graphical presentation for distribution of given data. The package produce three dimensional graphics with the possibility for horizontal and vertical rotation.

Fjalëkyçe: vlerësim i densiteteve, bërthame, shpërndarja probabilitare Epanechnikov, shpërndarja probabilitare normale, metoda grafike, matlab.

Hyrje

Në kohën kur të dhënat, sa po vijnë e po rriten në madhësi dhe kompleksitet, janë të nevojshme teknikat e vizualizimit dhe analizës së të dhënave, që jo vetëm të adresojnë problemet e një sasive të madhe të të dhënave, por

gjithashtu i lejon shkencëtarëve të kuptojnë më mirë proceset që prodhojnë këto të dhëna dhe tendencat e bashkësisë së rezultateve.

Vizualizim është teknik a e shprehjes së mesazheve përmes krijimit të imazheve, diagramave, apo animimeve.

Përdorimi i vizualizimit të të dhënave përforcon njohjen, ndërtimin dhe arsyetimin e hipotezave. Pra, vizualizimi është transformimi i të dhënave nga simulimet apo eksperimentet në një strukturë gjeometrike, për të eksploruar, analizuar dhe kuptuar të dhënat.

Vizualizimi shkencor fokuson dhe vë në dukje paraqitjen e të dhënave duke përdorur teknika grafike dhe të animuara.

Sot janë zhvilluar shumë platforma për vizualizimin e këtyre të dhënave, si SPSS, EViews, STATA etj, të cilat janë të specializuara në nxjerrjen e analizave statistikore të të dhënave hyrëse. Këto janë platforma të mbyllyra dhe përdoruesit mund të përdorin vetëm analizat e ofruara.

Krahas këtyre, janë zhvilluar edhe platforma të tjera, si MATLAB, ku secili prej kërkuesve mund të krijojë paketa të shumta për vizualizimin e këtyre të dhënave. MATLAB ofron një procesim shumë të shpejtë të tyre dhe gjithashtu ofron një mjedis pune për elementë statistikorë.

Materiali dhe metodat

Për zhvillimin e paketës grafike është zgjedhur platforma MATLAB.

MATLAB është ndër platformat më të përdorura prej kërkuesve ekonomikë për prodhimin sa më të thjeshtë dhe të shpejtë të rezultateve që ata dëshirojnë të nxjerrin. Pikërisht për këtë dhe për mundësinë e madhe që të ofron për gjenerimin e grafikëve është zgjedhur si platformë zhvilluese e paketës grafike 3-Dimensionale.

Kjo paketë do të ofrojë grafikë 3-Dimensionalë për vizualizimin e të dhënave që do të gjenerohen nga funksionet probabilitare dhe në veçanti nga shpërndarjet probabilitare. Pas studimeve të bëra në paketat e programet e ndryshëm statistikorë, është vënë re një mungesë në paraqitjen grafike të shpërndarjes së densiteteve probabilitare të variablove të ndryshëm.

Paketa do të ofrojë ndërtimin e grafikëve 3-Dimensionalë për disa nga funksionet bazë të shpërndarjeve probabilitare, dhe në vazhdim do të pasurohet akoma më shumë me këto lloj funksionesh. Për ilustrim, do të marrim një shembull, ku do të paraqesim vizualizimin në një grafik 3-dimensional të një funksioni shpeshësie të elementëve, brenda një hapësire të caktuar të dhënash.

Më poshtë janë listuar algoritmet e koduar.

Bërthama normale

Algoritmi 1: Vlerësimi i densiteteve, rasti d dimensional x_1, x_2, \dots, x_n duke përdorur bërthamë normale

Hapi 1: Fikso parametrin $h > 0$.

Hapi 2: $\forall \underline{t} \in R^d$, llogarit:

$$\hat{f}(\underline{t}) = \frac{1}{nh^d(2\pi)^{d/2}} \cdot \sum_{i=1}^n \exp \left\{ -\frac{1}{2h^2} \left[(t_1 - x_i^1)^2 + (t_2 - x_i^2)^2 + \dots + (t_d - x_i^d)^2 \right] \right\}$$

.Hapi 3: Fund.

Algoritmi 2: Vlerësimi i densiteteve, rasti d dimensional $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ duke përdorur bërthamë normale, me kusht m dimensional

Hapi 1: Fikso m parametra $k+1, k+2, \dots, k+m \in \{1, \dots, d\}$ (fikso m variabla ekonomikë), ku $1 \leq m \leq d-1$.

Hapi 2: Fikso një vlerë për cdo $t_{k+1}, t_{k+2}, \dots, t_{k+m}$. (fikso vlerat e variablave ekonomikë)

Hapi 3: Fikso një parametër $h > 0$.

Hapi 4: $\forall \underline{t} = (t_1, \dots, t_{k+1}, \dots, t_{k+m}, \dots, t_d) \in R^d$ (ku $t_{k+1}, t_{k+2}, \dots, t_{k+m}$ është e fiksuar) përllogarit:

$$\begin{aligned} \hat{f}_{\{1, \dots, k, k+m+1, \dots, d\}}(t_1, \dots, t_k, t_{k+m+1}, \dots, t_d) &= \\ &= \frac{1}{h^{d-m}(2\pi)^{\frac{d-m}{2}}} \\ & \sum_{i=1}^n \exp \left\{ -\frac{1}{2h^2} \left[\overbrace{(t_1 - x_i^1)^2 + \dots + (t_{k+1} - x_i^{k+1})^2 + \dots + (t_{k+m} - x_i^{k+m})^2}_{m \text{ copë}} \dots + (t_d - x_i^d)^2 \right] \right\} \\ & \cdot \frac{1}{\sum_{i=1}^n \exp \left\{ -\frac{1}{2h^2} \left[\overbrace{(t_{k+1} - x_i^{k+1})^2 + \dots + (t_{k+m} - x_i^{k+m})^2}_{m \text{ copë}} \right] \right\}} \end{aligned}$$

Hapi 5: Fund.

Bërthama Epanechnikov

Algoritmi 3: Vlerësimi i densiteteve, rasti d dimensional $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ duke përdorur bërthamë Epanechnikov

Hapi 1: Fikso një parametër $h > 0$.

$$\text{Hapi 2: Llogarit: } C_d = \begin{cases} \frac{\pi^{d/2}}{\left(\frac{d}{2}\right)!}, & d - \text{cift} \\ \frac{2^{\frac{d+1}{2}} \cdot \pi^{\frac{d-1}{2}}}{d!!}, & d - \text{tek} \end{cases}$$

Hapi 3: $\forall \underline{t} \in R^d$, llogarit:

$$a_i = \begin{cases} \left\{ 1 - \frac{1}{h^2} \left[(t_1 - x_i^1)^2 + (t_2 - x_i^2)^2 + \dots + (t_d - x_i^d)^2 \right] \right\}, & nqs \{ \dots \} > 0 \\ 0, & nqs \{ \dots \} \leq 0 \end{cases}$$

$$\text{Hapi 4: Llogarit: } \hat{f}(\underline{t}) = \frac{1}{nh^{\frac{d+2}{2}}} \cdot \frac{d+2}{2 \cdot C_d} \cdot \sum_{i=1}^n a_i.$$

Ne llogarisim:

$$\hat{f}(\underline{t}) = \frac{1}{nh^{\frac{d+2}{2}}} \cdot \frac{d+2}{2 \cdot C_d} \cdot \sum_{i=1}^n \left\{ \frac{\overbrace{h^2 - [(t_1 - x_i^1)^2 + (t_2 - x_i^2)^2 + \dots + (t_d - x_i^d)^2]}^{(\dots) > 0; pm=0}}{\underbrace{\hspace{10em}}}\right\}$$

Hapi 5: Fund

Algoritmi 4: Vlerësimi i densiteteve, rasti d dimensional $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ duke përdorur bërthamë Epanechnikov, me kusht m dimensional

Hapi 1: Fikso m parametra $k+1, k+2, \dots, k+m \in \{1, \dots, d\}$ (fikso m variabla ekonomikë), ku $1 \leq m \leq d-1$

Hapi 2: Fikso një vlerë për secilin $t_{k+1}, t_{k+2}, \dots, t_{k+m}$. (fikso vlerat e variablave ekonomikë)

Hapi 3: Fikso një parametër $h > 0$.

Hapi 4: $\forall \underline{t} = (t_1, \dots, t_{k+1}, \dots, t_{k+m}, \dots, t_d) \in R^d$ (ku $t_{k+1}, t_{k+2}, \dots, t_{k+m}$ janë të fiksuara) llogarit:

$$\begin{aligned} & \hat{f}_{\{1, \dots, k, k+m+1, \dots, d\}}(t_1, \dots, t_k, t_{k+m+1}, \dots, t_d) = \\ & = \frac{\Gamma\left(1 + \frac{d-m}{2}\right)}{h^{d-m} \cdot \pi^{\frac{d-m}{2}}} \cdot \\ & \sum_{i=1}^n \left\{ \frac{\overbrace{h^2 - \left[(t_1 - x_i^1)^2 + \dots + \underbrace{(t_{k+1} - x_i^{k+1})^2 + \dots + (t_{k+m} - x_i^{k+m})^2}_{m \text{ copë}} \dots + (t_d - x_i^d)^2 \right]}^{(\dots) > 0; tetjërati 0}}{\underbrace{\hspace{10em}}}\right\} \\ & \cdot \frac{2n h^2}{d-m+2} - \sum_{i=1}^n \left[\frac{(t_{k+1} - x_i^{k+1})^2 + \dots + (t_{k+m} - x_i^{k+m})^2}{m \text{ copë}} \right] \end{aligned}$$

Hapi 5: Fund.

Është realizuar është një aplikim, i cili merr hapësirën e të dhënave nga një skedar i jashtëm me të dhëna të strukturuar. Çdo strukturë e tij do të ketë një tip të dhëne, e cila do mbajë vlerat e të dhënës, që në këtë rast është hyrëse. Këto tipe të dhënash do të pasqyrohen si variabla për t'u marrë parasysh apo jo në gjenerimin e grafikut 3-dimensional.

Element tjetër i rëndësishëm janë vlerat që kanë të dhënat hyrëse, të cilat mbahen në skedar. Ato janë praktikisht matje të regjistruara. Fillimisht, intervali i vlerave të tyre në aplikacion do jetë minimumi dhe maksimumi i vlerave të skedarit.

Përdoruesi do të këtë të drejtën e:

- zgjedhjes për të dhënat hyrëse, se cilat prej tyre do të marrë parasysh në ndërtimin e grafikut
- zgjedhjes nëse të dhënat do kenë vlera fikse apo vlera e tyre do të variojë në vlerat e ruajtura në skedar
- hapi, i cili do të përcaktojë largësinë e pikave të grafikut lidhur me boshtet e tij.

Për ndërtimin e një pakete të tillë, fillimisht është ndërtuar ndërfaqja e saj grafike për përdoruesin. Kjo është bërë e mundur falë GUIDE, që lejon ndërtimin e ndërfaqeve grafike për përdoruesin.

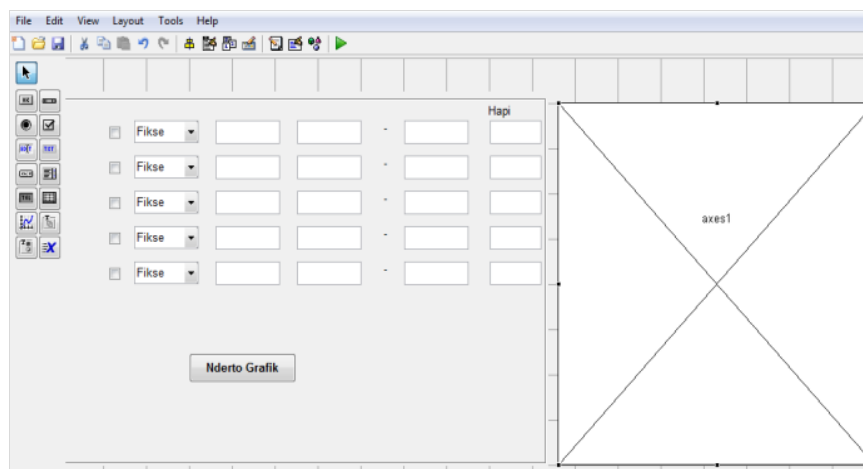


Figura1. Ndërtimi i ndërfaqes grafike të aplikacionit ne GUIDE.

Nëse e vini re, fillimisht nuk ka asnjë emërtim për rreshtat që janë mbushur me elemente. Kjo, sepse të dhënat që do të gjenerojnë grafikun merren nga një skedar Excel. Ndërkohë që, nëse do ta ekzekutonim ndërfaqen grafike, do shfaqej në këtë formë:

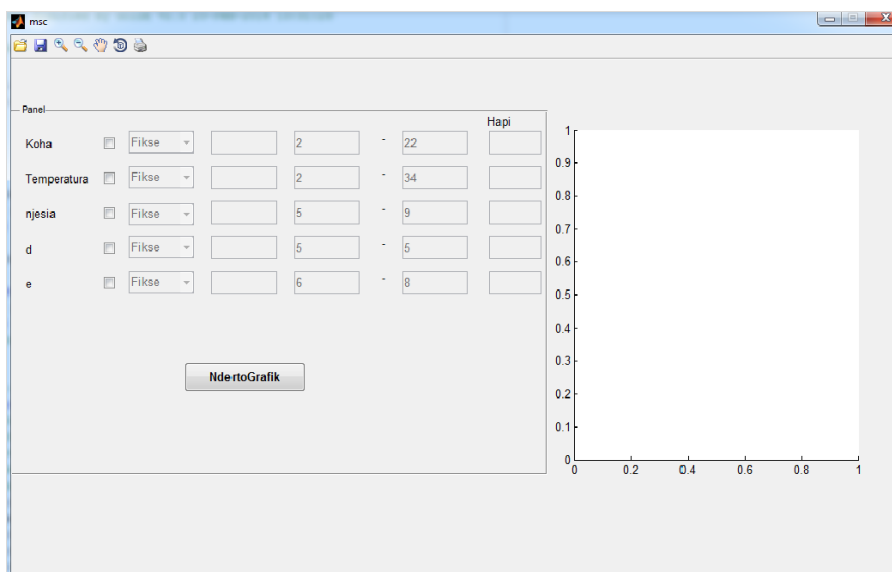


Figura 2. Ekzekutimi i ndërfaqes grafike te aplikacionit.

Pra, fillimisht aplikimi lexon skedarin Excel. Nga skedari lexon emrat e variablove dhe ia vendos ato si emërtime çdo rreshti elementesh të ndërfaqes grafike. Në fillim, të gjitha kontrollet e ndërfaqes grafike janë të paeditueshme.

Përdoruesi fillimisht zgjedh nëse dëshiron ta marrë parasysh në grafikun e tij një të dhënë. Nëse ai dëshiron ta marrë parasysh, ai zgjedh kutinë e kontrollit në kraht të te variablit. Pas këtij veprimi, shfaqet aktive lista e zgjedhjes për mënyrën e vlerëdhënies së kësaj të dhëne.

Elementi tjetër i rreshtit është një listë me dy vlera: Fikse dhe Interval.

- Nëse përdoruesi dëshiron që e dhëna e zgjedhur të ketë një vlerë fikse, atëherë ai zgjedh opsionin ‘Fikse’.
- Nëse përdoruesi dëshiron që e dhëna e zgjedhur të ketë vlerën në një interval vlerash, atëherë ai zgjedh opsionin ‘Interval’.

Në varësi të zgjedhjes që bën përdoruesi, aktivizohet fusha e parë, kur zgjidhet opsioni ‘Fikse’ dhe dy fushat e plotësuara me vlera, kur zgjidhet opsioni ‘Interval’.

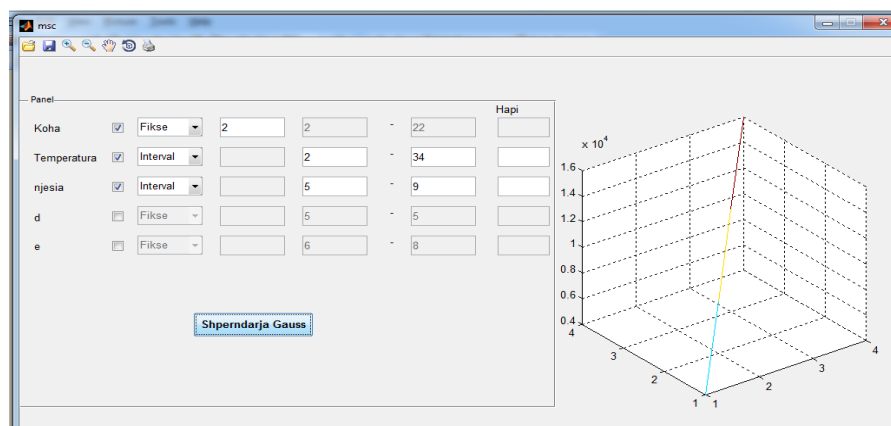
Për opsionin ‘Fikse’, përdoruesi duhet të vendosë një vlerë tek fusha, që aktivizohet.

Për opsionin ‘Interval’, përdoruesit i aktivizohen dy fushat me vlera, përkatësisht vlera minimale dhe maksimale që ka e dhëna në skedarin Excel. Përdoruesi mund t’i ndryshojë ose jo këto vlera. Në fund të tyre, ndodhet edhe një fushë, i cila përcakton hapin, pra distancën mes pikave që do gjenerojë grafiku.

KUJDES!

Nuk lejohen më shumë se dy të dhëna, vlerat e të cilave janë interval. Kjo për arsye sepse ne mund të kemi maksimalisht dy variabla që marrim vlera në një segment për të prodhuar rezultat grafik. Këto dy variabla do jenë dy boshte prej boshteve të sistemit, ndërsa boshti i tretë do të jetë vlera e funksionit në pikat përkatëse. Çdo variabël tjetër mund të ketë vetëm një vlerë prerëse.

Pasi përcaktohen variablat dhe vlerat e tyre, klikohet butoni, dhe paketa na paraqet grafikun e krijuar.



Paketa e krijuar ka disa menu për:

- Ruajtjen e grafikut
- Rrotullimin e tij
- Zmadhimin/zvogëlimin
- Printimin

Kjo pakete mund të shndërrohet në një App të MATLAB, të instalohet në platformë, dhe nëse dëshirohet mund të shpërndahet online edhe me përdoruesit e tjerë të MATLAB.

Përfundime

Paketa e ndërtuar në MATLAB ofron një përpunim më të shpejtë të dhënave se platforma e ndërtuar në php.

Paketa e ndërtuar në MATLAB, duke qenë një aplikim desktop, kërkon parametra në kompjuterin e përdoruesit dhe jo në server.

Paketa e ndërtuar në MATLAB është më lehtë e përdorshme nga kërkuesit, pasi funksionalitet e saj përdorin menu të njëjta me shumicën e paketave ekzistuese të MATLAB.

Literatura

- Ceca K. (2011): Vlerësimi i densiteteve probabilitarë (kernel estimation), Seminar teknik, Banka e Shqipërisë, Tiranë
- Tanku A. & Ceca K (2012): Vlerësimi i densiteteve per variablat ekonomike. Konferenca e X ndërkombëtare e Bankës së Shqipërisë “Politika monetare nga e shkuara në të ardhmen”. Tiranë, Albania
- Tanku A. & Ceca K.(2013): Density estimation for economic variables – a genuine application. Seria e materialeve te diskutimit te BSH. Tiranë, Albania
- Tanku A. & Ceca K.(2014): Conditional density estimate for economic variables. Pranuar per botim në Seria e materialeve te diskutimit te BSH. Tiranë, Albania
- Saatçiu D. (2012): Ngritja e disa algoritmeve optimale për përafrimin me anë të metodës së vlerësimit të densiteteve të variablave ekonomike të çfarëdoshëm. Konferenca Fakulteti i Shkencave Natyrore në 100 vjetorin e pavarësisë, Tiranë, Albania
- Saatçiu D. (2013): Computer models in probability density estimation. Applications in economy, in the conference ;Economic research in south east Europe, Bank of Albania, Tirana, Albania, December 2013
- Hendry F. D. (2004): Causality and Exogeneity in Non-stationary Economic Time-Series.” Centre for Philosophy of Natural and Social Science Causality: Metaphysics and Methods. Technical Report 18/04
- Hendry F. D. (2011): Empirical economic model discovery and theory evaluation. University of Oxford. Department of Economics. Discussion paper series. ISSN 1471-04098
- Juselius K. (2005): The Cointegrated VAR Model: Methodology and Applications