Moran eta Gearyren autokorrelazio-adierazleen aldaera bana

Yosu Yurramendi Mendizabal

Konputazio Zientziak eta Adimen Artifiziala saila, UPV/EHU,
yosu.yurramendi@ehu.es

Eduardo Concepción Morales

Universidad de Cienfuegos, Facultad de Informática, Cuatro Caminos,
Cienfuegos, Cuba, econcep@ucf.edu.cu

Laburpena: *Moran* (1950) eta *Geary*ren (1954) espazioko autokorrelazio-adierazleak (I eta C, hurrenez hurren) aski ezagunak dira espazioko datu estatistikoen azterketan. Lan honetan espazioko batezbestekoa eta bariantzaren definizio berriak ematen ditugu, eta bi adierazle horien aldaera bana lortzen dira (I^* eta C^*, hurrenez hurren). Aldaerek, adierazle klasikoek ez bezala, desberdintza hauek betetzen dituzte: $-1 \leq I^* \leq +1$ eta $0 \leq C^* \leq 2$. Gainera, elkarrosagarriak dira biak: $I^* + C^* = 1$.

Abstract: *Moran*'s I (1950) and *Geary*'s C (1954) are two spatial autocorrelation indices, well-known in spatial statistical data analysis. In this paper we give new definitions for the spatial mean and variance, and two modified spatial autocorrelation indices are defined: I^* and C^*. The modified indices, but not the classical ones, satisfy the inequalities: $-1 \leq I^* \leq +1$ and $0 \leq C^* \leq 2$. Morever, both are complementary: $I^* + C^* = 1$.

SARRERA

Espazioko datuen analisiak aztergai den objektu-multzoaren ezaugarri geometriko edo topologikoak hartzen ditu kontuan, datu-analisiaren teknika klasikoek ez bezala.

Datu-analisi mota honen teknika berriak XX. mendearen bigarren erdian hasi ziren bilakatzen garatzen, eta egun ere indarrean daude espazioko erlazioek problema berri asko mahaigaineratzen dituztelako, eta beren aplikazioa jakintz-eremu askotara zabaltzen delako.

Objektuak espazio batean daudenean antolatuta, hurbiltasunak, maiz, korrelazio bat eragiten du fenomeno edo aldagai baten balioen artean; espazioko elkarreragina autokorrelazioa izendatzen den kontzeptuaren bitartez uler daiteke.

Autokorrelazioa datu-multzo askoren ezaugarri komun bat da. Adibidez, irudi digital baten pixelen intentsitatearen balioak elkarrekin lotuta daude, orokorrean irudian duten hurbiltasunaren arabera. Zelula batean dauden proteinek betetzen duten funtzioak ba omen du zerikusirik zelulan duten lekuarekin. Lurralde baten espazio-banaketa eta zati bakoitzaren garapen ekonomikoaren maila zer edo zer gehiago da zoriak ematen duena baino. Epidemia baten zabaltze eta gelditze prozesuak lekuen ezaugarriekin du elkarreragina. Giza harremanak grafo baten bitartez adieraz daitezke, korapilo edo nodoak pertsonak edo erakundeak direlarik, eta loturak beren arteko erlazioak direlarik. Web orriak ere espazio birtual batean daude nolabait semantikoki erlazionatuta, aipatzen dituzten gaien arabera.

Objektuak denboran zehar daudenean antolatuta, bestalde, denbora bera dimentsio bakarreko espaziotzat har daiteke, linealtzat hain zuzen, eta objektuen arteko elkarreragina aztertu egin daiteke autokorrelazioaren bitartez.

Lan honetan aurkezten ditugun espazio-autokorrelazioa neurtzeko bi adierazle, joan den mendeko bigarren erdiaren hasieran definitu ziren bi adierazle klasikoen aldaerak dira ([1],[2]). Beren ezaugarri nagusiak aurkezten ditugu, eta bien arteko lotura ere.

ESPAZIOKO AUTOKORRELAZIOA

Izan bedi $\Omega = \{\omega_i / i = 1, \ldots, n\}$ mapa edo espazio batean kokatuta dagoen objektu multzo bat; hau da, espazioan egituratuta dagoen multzo bat. Objektuetan X aldagai bat (edo batzuk) behatzen da, $\{x_i / i = 1, \ldots, n\}$, eta horren arabera antolaketa-ezaugarri bat aztertzen da. Espazio-autokorrelazioaren helburua zehazki antolaketa-ezaugarri hori salatzen saiatzea da. «Antolaketa-ezaugarri» terminoaren ulerkera bat baino gehiago izan daitezke, eta hortik datoz autokorrelazioaren definizio ezberdinak.

Mapa edo espazioaren kokagune jakin batean X aldagaiaren balio garaiak balio garaiz inguratuta badaude, balio txikiak balio txikiz inguratuta, autokorrelazio positiboa dagoela esan daiteke. Balio garaiak balio txikiz inguraturik badaude edo alderantziz gertatzen bada, berriz, autokorrelazio negatiboa dago. Autokorrelazio deskribatzailea esaten zaio aldagai batek kokagune batean hartzen dituen balioak ingurukoen parekoak diren ala ez aztertzen duen ulerkerari. Horixe da lan honetan jorratzen duguna.

Beste ulerkera batzuk probabilitatean oinarritzen dira: kokagune jakin batean behatutako balioa zein izango ote den aztertzen duena, ingurukoena ezagututa. Are gehiago, kokagune batean behatutako balioa inguruko balioekin alderatu eta independenteak ote diren aztertzen duena.

Esan bezala, ulerkera deskribatzailea da lan honen helburua. Deskribapen horretarako bi dira kontuan hartzen diren nozioak, adigaiak:

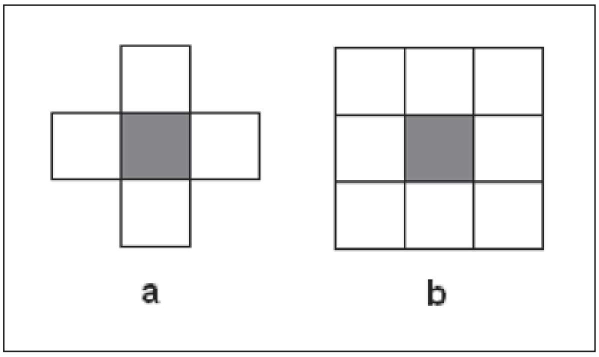
— (ω_i, ω_j) objektu pare bakoitzari autokorrelazioarekin zerikusia duen ρ_{ij} balio bat esleitzen zaio; balio honek X aldagaiak objektuen gainean lortutako balioen arabera erabakitzen da.
— Beste alde batetik, a_{ij} beraien arteko espazio-hurbiltasuna adierazten duen beste balio bat da.

X aldagaiaren autokorrelazioaren adierazle bat balio horien guztien Γ estatistiko bat da. Adibidez:

$$\Gamma = \sum_i \sum_j a_{ij}\, \rho_{\mathrm{ij}}$$

$P = [\rho_{ij}]$ matrizeak behatutako X aldagaiarekiko duten antzekotasuna edo antzekotasun eza adierazten du, eta $A = [a_{ij}]$matrizeak edozein ω_i eta ω_j bi objekturen arteko hurbiltasun edo urruntasuna.

Adibidez, mapa batean bi objekturen arteko hurbiltasuna $a_{ij} = 1$ balioaren bidez adieraz daiteke ondoz ondokoak badira, eta $a_{ij} = 0$ balioaren bidez horrelakoak ez badira. Ondoz ondokoa izatea ere modu bat baino gehiagotan uler daiteke. Esate baterako, espazio-egitura sareta erregular batena bada (irudi digital batena) ondoz ondokoak *4* izan daitezke, *8* (ikus 1. irudia) ala gehiago.



1. irudia. Ondoz ondokotasun bi mota. (a) *4*-ingurua; (b) *8*-ingurua.

Orokorrean, objektu baten ingurukoak aldez aurretik erabakitako r radio edo ohiko distantzia euklidestar batera baino hurbilago daudenak dira. Horrela, *4*-ingurukoak $r = 1$ distantziara edo gutxiagora dauden objektuak dira, eta *8*-ingurukoak $r = \sqrt{2}$ distantziara edo gutxiagora daudenak.

Objektuak espazioan irregularki banatuta dituen eremuak izan badira. Objektu baten ingurukoak izan daitezke berekin mugaren bat dutenak.

Aurreko kasu guztietan ohikoa da $a_{ij} = 1$ ω_i eta ω_j definitzea objektuak ondoz ondokoak direnean, eta $a_{ij} = 0$, bestelakoetan. Hautazkoa izaten da objektu bat norberaren ingurukoa den ala ez erabakitzea, baina ohikoa da $a_{ii} = 0$ erabakitzea. Dena dela, lan honetan hautu honek ez du garrantzirik, erabakia autokorrelazio-adierazlea erabili behar den testuinguruaren barruan dago eta.

A ω_i eta ω_j objektuen arteko d_{ij} espazio-distantziaren bitartekoa da hurbiltasun-matrizea definitzeko beste bide bat. Hau lineala edo euklidestarra izan daiteke, ohi bezala, ala sare baten zeharkakoa. d_{ij} distantzia definituz gero, $a_{ij} = 1/d_{ij}$ edo $a_{ij} = 1/d_{ij}^2$ har daitezke, edo beren aldaera esponentzialak diren $a_{ij} = e^{-d_{ij}}$ edo $a_{ij} = e^{-d_{ij}^2}$, esate baterako. A matrizearen aldaera hauetan guztietan kontuan hartzekoa da $d_{ij} = 0$ gerta daitekeela, hau da, bi objektu espazioko leku berean egon daitezkeela, edo areago, $d_{ii} = 0$ saihestu beharreko kasua ere.

Edozein dela A matrizea, ezinbestekoa da autokorrelazio-adierazlearen balio ezberdinak aukera horren arabera daudela gogoratzea.

Ondoren *Moran* eta *Gearyk* proposatutako autokorrelazio-adierazle klasikoen *aldaera* bana aurkeztuko ditugu ([1], [2]). Aldaera hauen bitartez bi ikuspegiak elkartzen ditugu, eta beraz, biak modu bakarrean aztertzeko bidea ematen digu; gainera, bidearen zenbait ezaugarri interesgarri erakusten zaizkigu bertan. Izan badira beste zenbait autokorrelazio-adierazle, baina testuinguru honetatik at geratzen dira.

ALDEZ AURRETIKO ZENBAIT DEFINIZIO ETA EZAUGARRI ESTATISTIKO

Atal honetan, lehendabizi, proposatzen ditugun autokorrelazio-adierazleen oinarrian dauden X aldagaiaren batezbesteko balio eta bariantzaren definizioak aurkezten ditugu. Gero, objektuen inguruetan oinarritutako beste bi aldagai definitzen ditugu, eta beren bi ezaugarri nagusiak ematen ditugu. Autokorrelazio-adierazleak ulertze aldera, X aldagaiarekin duten kobariantza ikusiko dugu.

Objektuen gaineko bi aldagairen datu gordinak taula batean errepresentatzen diren modura, A hurbiltasun-matrizearen lerro eta zutabeak lerroka azaltzean datza (ikus 2.irudia) X aldagaiaren eta A matrizearen balioen taula-irudikapen bateratu bat.

x_1	x_2	x_n
y_1	y_2	y_n

(a)

K	a_{i1}	a_{i2}	a_{in}
K	x_i	x_i	x_i
K	x_1	x_2	x_n

(b)

2. irudia. Taula-errepresentazioaren analogia: (a) bi aldagairen datu gordinak; (b) espazioko datu gordinak.

Nabarmentzekoa da grafo baten taula-errepresentazio hau, garatuko dugun guztiaren oinarrian baitago.

Ohiko estatistikaren ikuspegitik Ω multzoa izaten da unibertso edo populazioa, eta ω bere elementuak izaten dira objektuak, erregistroak, indibiduoak edo unitate estatistikoak. Ohiko estatistikak unitate hauetan behatzen diren balioen gainean egiten dira. Ω multzoa egituratua dagoenean, aldiz, beharrezkoa da egiturak ematen duen informazioa estatistikoen kalkuluetan sartzea.

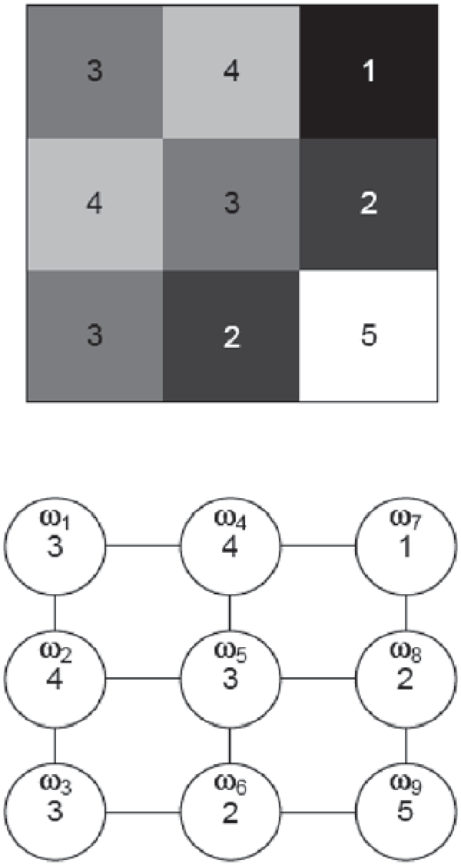
Taula-irudikapen berrian, unitate estatistiko bat ez da objektu soil bat, ondoz ondo dauden bi objekturen balioei dagokien ondoz ondokotasun elementua baizik. Ikuspegi honetan, objektu bat bere balioaren eta bere ondoz ondoko objektuen balioen bitartez adierazten da, eta Ω unibertsoa objektuen multzoa denez gero, bere objektuen balioen eta bere gainean definitutako A matrizearen bitartez ere.

Adibidez, irudi digital bat sareta erregulartzat har daiteke, eta pixel batean gertatu den intentsitatea behatutako aldagaitzat. A hurbiltasun-matrizea *4*-inguruaren arabera defini daiteke: ω_i pixel bakoitzaren ondoz ondoko pixelak dira bere gainean, azpian, eskuinean eta ezkerraldean dauden pixelak. Hauentzat $a_{ij} = 1$, eta gainontzekoentzat $a_{ij} = 0$. 3. irudiko *4*-inguruari dagokion A matrizea hauxe da:

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Ohartzekoa da adibidean $a_{ii} = 0$ dela. 3. irudian azaltzen dira irudia bera, dagokion grafoa eta bere taula-irudikapena ($a_{ij} = 0$ duten bikoteak kenduta).

Objektuen *4*-inguruei erreparatuz gero, hau da, ω_i pixel bakoitzaren balioa eta ondoz ondokoena kontuan hartuta, atzematen da ω_1-k inguru guztiz homogeneo bat duela, bere baliotik ez oso desberdina; gauza bera esan daiteke ω_2 eta ω_9 pixelei buruz. ω_8 pixelak, aldiz, oso inguru heterogeneo du. ω_4 eta ω_6 pixelek ez dute hain inguru heterogeneoa, baina alde handia dago beren balioen eta inguruaren artean. Heterogeneoagoa da ω_5, baina bere ba-



Ω	Ω	a_{ij}	X	X'
ω_1	ω_2	1	3	4
ω_1	ω_4	1	3	4
ω_2	ω_1	1	4	3
ω_2	ω_3	1	4	3
ω_2	ω_5	1	4	3
ω_3	ω_2	1	3	4
ω_3	ω_6	1	3	2
ω_4	ω_1	1	4	3
ω_4	ω_5	1	4	3
ω_4	ω_7	1	4	1
ω_5	ω_2	1	3	4
ω_5	ω_4	1	3	4
ω_5	ω_6	1	3	2
ω_5	ω_8	1	3	2
ω_6	ω_3	1	2	3
ω_6	ω_5	1	2	3
ω_6	ω_9	1	2	5
ω_7	ω_4	1	1	4
ω_7	ω_8	1	1	2
ω_8	ω_5	1	2	1
ω_8	ω_7	1	2	3
ω_8	ω_9	1	2	5
ω_9	ω_6	1	5	2
ω_9	ω_8	1	5	2

3. irudia. Irudi digital bat, bere grafoa eta *4*-inguruaren taula-irudikapena.

lioa inguruko balioen tartean dago. ω_3 pixelaren ingurua ez da hain heterogeneoa eta bere balioa ingurukoen antzekoa da. Ez da horrelakoa ω_7 pixelaren kasua: ω_3-ren inguru bera du, baina bere balioa txikiagoa da.

X' aldagaia *X* bera da baina berrordenatua, eta horrela objektuen arteko espazioko ondoz ondokotasuna adierazten da.

Datozen definizio eta propietateek adibidean bertan azaldutako oharrak zenbakiz adierazteko bidea egitea dute helburu.

ω_i objektu bakoitzaren espazio-pisua honela definitzen da:

$$A_i = \sum_{j=1}^{n} a_{ij}$$

Guztirako espazio-pisua horien guztien batura da:

$$A_{\text{Total}} = \sum_{i=1}^{n} A_i = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$$

Ikusi beharrekoa da A matrizea simetrikoa bada ($a_{ij} = a_{ji}$), espazio-pisuak bi aldiz batzen direlarik (beti ere $a_{ii} = 0$ bada).

Objektuen espazio-pisuak kontuan hartuta, X aldagaiaren balioen batezbesteko balioa honela definitzen da:

$$\bar{x}_A = \frac{1}{A_{\text{Total}}} \sum_{i=1}^{n} A_i x_i$$

Hortaz, espazio-pisuaren araberakoa da objektu bakoitzak batezbestekoaren balioari ematen diona. Irudi digital edo sareta erregular baten kasuan, ertzeko pixelek barneko pixelek baino gutxiago eragiten diote batezbestekoari, eta izkineko pixelek askoz ere gutxiago barnekoek baino.

Ohiko batezbestekoan, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, ez dira espazio-pisuak kontuan hartzen (edo, beste modu batez esanda, denek dute pisu berdina):

X aldagaiaren bariantza, beraz, honela definitzen da:

$$s_A^2 = \frac{1}{A_{\text{Total}}} \sum_{i=1}^{n} A_i \left(x_i - \bar{x}_A\right)^2$$

ohiko bariantza, $s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ formularen bitartez kalkulatzen den bitartean.

Ohikoen ondoan estatistiko berri hauei, *batezbesteko balio aldatua* eta *bariantza aldatua* esaten diegu, elkarrengandik bereizte aldera.

Dena dela, esan behar da sareta nahiko zabala bada bi batezbesteko balioak eta bi bariantzak oso antzekoak izango direla, ertzek eta izkinek proportzioan garrantzi gutxiago izango dutelako.

Bada ohiko bariantzaren propietate bat, bariantzaren ulerkera errazten duena, eta bariantza aldatuak ere baduena:

$$\frac{1}{A_{\text{Total}}^2} \sum_{i=1}^{n}\sum_{j=1}^{n} A_i A_j \left(x_i - x_j\right)^2 = 2 s_A^2$$

Berdintzaren ezkerreko aldeak X aldagaiaren homogeneotasuna adierazten du, eta bariantzaren bikoitza da. Hau ulertzeko kontuan hartzekoa da berdintzaren ezker aldeko batugai karratu bakoitza bi aldiz batzen dela.

Gure lanaren emaitzak ez dira beraz $A_i\, A_j \,/\, A^2_{\text{Total}}$ pisu-sisteman, $a_{ij} \,/\, A_{\text{Total}}$ pisu-sistemaren erabileran baizik:

$$\frac{1}{A^2_{\text{Total}}}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(x_i - x_j\right)^2 = 2s_A^2 C^*$$

C^* horren balioa eta esanahia aurrerago zehaztuko ditugu, baina jakina, zerikusia du espazio-autokorrelazioarekin.

X aldagaia, bestalde, ohi bezala estandariza daiteke: $z_i = \dfrac{x_i - \bar{x}_A}{s_A}$

Hortaz, Z aldagaiaren A_i, $i = 1, 2, \ldots, n$ pisu-sistemarekiko batezbesteko balioa *0* da, eta bariantza aldatua *1*.

X aldagaiak bere buruarekin duen A hurbiltasun-matrizearekiko *kobariantza aldatua* definitzen da:

$$\operatorname{cov}\left(X, X'\right) = \frac{1}{A_{\text{Total}}}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(x_i - \bar{x}_A\right)\left(x_j - \bar{x}_A\right)$$

X' aldagaia jatorrizko X aldagaiaren berrordenazio bat da, adibidearen taula-irudikapenean azaldu dugun legez; horrela, objektuen inguruak behar bezala adierazten dira. Jakina, bere batezbesteko balio aldatua eta bariantza aldatua *X*ren berdinak dira.

MORANEN AUTOKORRELAZIO-ADIERAZLE ALDATUA

Adierazle hau lehen aipatutako Γ formula orokorraren kasu berezi bat da: a_{ij} balioak dira A matrizeko elementuak, baina guztirako A_{Total} pisuarekiko erlatibizatuta $\left(\dfrac{a_{ij}}{A_{\text{Total}}}\right)$, eta ρ_{ij} balioak ω_i eta ω_j objektuen gainean X aldagaiak hartzen dituen z_i eta z_j balio estandarren biderkadura gurutzatua ($z_i = z_i \cdot z_j$):

— ρ_{ij} positiboa da, baldin eta x_i eta x_j bi balioak X aldagaiaren $\bar{x}_A$ batezbestekoa baino handiagoak badira ($z_i > 0$ eta $z_j > 0$), edo txikiagoak biak ($z_i < 0$ eta $z_j < 0$).

— ρ_{ij} negatiboa da, baldin eta $z_i > 0$ eta $z_j > 0$ badira, edo $z_i < 0$ eta $z_j < 0$).

*Moran*en *autokorrelazio-adierazle aldatua* honela definitzen da, beraz:

$$I^* = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{a_{ij}}{A_{\text{Total}}}\right)\left(z_i z_j\right) \equiv \frac{\operatorname{cov}\left(X, X'\right)}{s_A^2} \equiv \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(x_i - \bar{x}_A\right)\left(x_j - \bar{x}_A\right)}{\sum_{i=1}^{n} A_i\left(x_i - \bar{x}_A\right)^2}$$

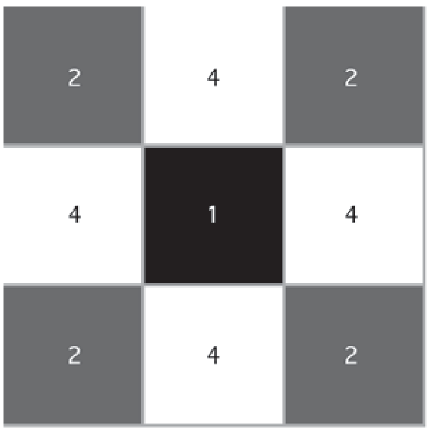
Adierazle honetan ohiko batezbestekoa eta ohiko bariantza erabiliz gero $\left(\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ eta } s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right.$, hurrenez hurren), *Moran*en *I* autokorrelazio-adierazle klasiko bera lortuko da ([1]).

*Moran*en adierazlea, bere aldetik, bi aldagairen arteko korrelazio lineala neurtzeko erabiltzen den *Pearson*en *r* koefizientearen kasuan oinarritzen da:

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{s_x s_y} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

s_x eta s_y *X* eta *Y* aldagaien desbideratze estandarrak izanda, hurrenez hurren.

Hain zuzen, *Moran*en autokorrelazio-adierazlea *X* aldagaiak bere buruarekin (*X*') duen *A* hurbiltasun-matrizearekiko korrelazio lineal aldatua da.



$I = -1.0577$, $I^* = -0.9245$

4. irudia. Moranen *I* adierazle klasikoa [–1, +1] tartetik at; *I** ez bezala.

I adierazle klasikoaren kasuan ez da zehazki *Pearson*en r koefizientea. Izan ere, $-1 \leq r \leq +1$ da, eta adierazle klasikoa muga horietatik at har ditzake bere balioak (ikus 4. irudia). *Moran*en adierazle klasikoa eta aldatuaren arteko jokabide ezberdinak argitzeko, kontuan hartu behar da nola egin diren kalkuluak: bietan erabili dira ondoz ondoko espazio-erlazioak kobariantzaren balioa kalkulatzean, baina klasikoarenean ez dira erabili batezbestekoa eta bariantzarenak kalkulatzean, eta aldatuarenean bai.

*Moran*en autokorrelazio-adierazle aldatuaren kasuan, unitate estatistikoak ondoz ondokotasun elementuak dira, eta beren pisuak kalkulu guztietan erabiltzen dira; beraz I^* *Pearson*en r koefizientearen kasu berezi bat dugu eta ondorioz,

$$-1 \leq I^* \leq +1$$

$I^* = +1$ balioak adierazten du erabateko autokorrelazio lineal positiboa, eta $I^* = -1$ balioak autokorrelazio lineal negatiboa. 3. irudiko adibidean, $I^* = -0.2857$ da. Honek autokorrelazio negatibo samarra adierazten du, eta hori bat dator irudiari egindako oharrekin; alegia, pixel baten balioa garaia denean ingurukoek balio txikiagoa izateko joera azaltzen dute zenbaitetan, eta alderantziz besteetan.

Espazio-egituraren ikuspegitik, objektu guztiak guztiekin erlazionatuta daude ($a_{ij} = 1$, baita $a_{ii} = 1$ ere) egoera klasikoan:

$$a_{ij} = 1 \qquad A_i = \sum_{j=1}^{n} a_{ij} = n \qquad A_{\text{Total}} = \sum_{i=1}^{n} A_i = n^2 \qquad \frac{a_{ij}}{A_{\text{Total}}} = \frac{A_i}{A_{\text{Total}}} \frac{A_j}{A_{\text{Total}}}$$

$$\bar{x}_A = \frac{1}{A_{\text{Total}}} \sum_{i=1}^{n} A_i x_i = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

$$s_A^2 = \frac{1}{A_{\text{Total}}} \sum_{i=1}^{n} A_i \left(x_i - \bar{x}_A\right)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = s^2$$

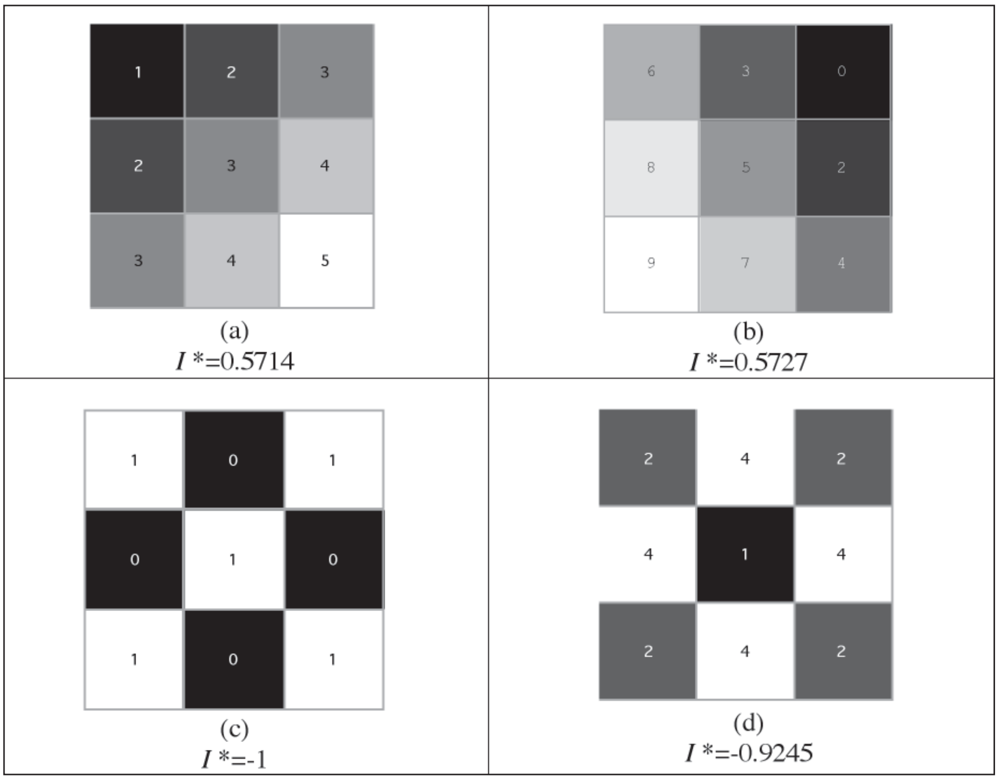
$$\frac{1}{A_{\text{Total}}} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \left(x_i - \bar{x}_A\right)\left(x_j - \bar{x}_A\right) = \sum_{i=1}^{n} \frac{A_i}{A_{\text{Total}}} \left(x_i - \bar{x}_A\right) \sum_{j=1}^{n} \frac{A_j}{A_{\text{Total}}} \left(x_j - \bar{x}_A\right) = 0$$

Ondorioz, ohiko egoera hau espazio-autokorrelaziorik ezaren kasu bat dela esan dezazkegu: $I^* = 0$. Pentsa daiteke ere ohiko egoeran $a_{ij} = 1$ izanda $a_{ii} = 1$ dela; orduan, froga daiteke $I^* = \dfrac{-1}{n-1}$ dela.

Egiturarik gabeko egoera denean ($a_{ij} = 1$ izanda, $a_{ii} = 0$) $I^* = 1$ gertatzen da, eta hau ez dator bat egiturarik gabeko ustearekin.

Oro har, adierazle aldatuaren mugak *+1* eta *–1* izateak ez du esan nahi edozein egituratan iritsi daitezkeenik. Izan ere, frogatuta dago bi mugak, altuena eta baxuena, zehatzago lotuta daudela *A* matrizearen autobalioekin [3].

5. irudian *4*-inguruko sareta erregularren zenbait adibide ematen dira.



6. irudia. 3 × 3 sareta erregularrak, I^*-ren balioak *4*-ingurukoa erabiliz.

I^*-ren balioan, esan bezala, inguruaren helmenak (*r* radioak, eta, ondorioz, *A* matrizeak) badu zeresanik. Gerta daiteke ingurua estua denean oso autokorrelazionatuta egotea, eta ingurua zabala denean autokorrelazionatuta ez egotea.

Esan beharra dago halaber korrelazio koefizienteen antzera, horrelako adierazle bat tresna egokia dela espazioko datu-analisirako erabiltzen diren zenbait teknikatan, hala faktore-analisia, nola erregresio-analisian.

GEARYREN AUTOKORRELAZIO-ADIERAZLE ALDATUA

Espazio-autokorrelazioa neurtzeko formula orokor horretan, *A* matrizea *Moran*en adierazlearen kasuan bezala defini daiteke, eta horrela beste zenbait adierazle lor daitezke, $P = [\rho_{ij}]$ ω_i eta ω_j objektuen arteko antzekotasun ezako matrize gisa. Adibidez,

$$\rho_{ij} = \frac{1}{2}\left(z_i - z_j\right)^2$$

Horrelaxe definitzen da *Geary*ren *autokorrelazio-adierazle aldatua*:

$$C^* = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{a_{ij}}{A_{\mathrm{Total}}} \frac{1}{2}\left(z_i - z_j\right)^2 \equiv \frac{1}{2} \frac{\frac{1}{A_{\mathrm{Total}}}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(x_i - x_j\right)^2}{\sum_{i=1}^{n} A_i\left(x_i - \bar{x}_A\right)^2}$$

Pentsa daiteke 1/2 faktorea erabiltzen dela objektuen balioen arteko aldeak, hau da batugaiak, *2* aldiz zenbatzen direlako *A* simetrikoa denean.

Definizio honek argitu egiten du lehen azaldutako formula baten ulerkera:

$$\frac{1}{A_{\mathrm{Total}}}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(x_i - x_j\right)^2 = 2 s_A^2 C^*$$

Horrela, aberastuta geratzen da *X* aldagairen Ω objektu multzoegituratuarekiko batezbesteko homegeneotasunaren bitarteko bariantza klasikoaren ulerkera, espazio-egituraren informazioa azterketan txertatzen denean. Izan ere, *X*-ren bariantza eta *Geary*ren autokorrelazio-adierazlearen barruan dago.

*C** adierazle hau *Geary*ren *C* autokorrelazio-adierazle klasikoaren ([2]) aldaera bat da. Hartan ere ohiko batezbestekoa eta bariantza erabiltzen dira, baina ez *A* matrizearen bitartez kalkulatutakoak, eta, gainera, $1/n$ faktorearen ordez, estimazio edo zenbatespen alboragabetuaren ikuspegiak ematen duen $1/(n - 1)$ faktorea erabiltzen da bariantzaren kalkulua egiteko.

Definizio horren pean *Geary*ren autokorrelazio-adierazle aldatua estuki lotuta dago *Moran*en autokorrelazio-adierazle aldatuarekin; *1*-ekiko elkarrosagarriak dira hain justu:

$$I^* + C^* = 1$$

Propietate hau da egin dugun aurkikuntzarik garrantzitsuena, bi adierazle klasikoak ez baitira horrela lotzen:

$$C = \frac{n-1}{n}\left[\frac{n}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\right)}\frac{\sum_{i=1}^{n}A_i\left(x_i-\bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}-I\right]$$

Bi adierazle aldatuen arteko elkarrosagarritasuna frogatzeko nahikoa da *Geary*ren autokorrelazioko zenbakitzailea, *Moran*ena lortzearren, behar bezala garatzea:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\left(x_i-x_j\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\left(\left(x_i-\bar{x}_A\right)\left(x_j-\bar{x}_A\right)\right)^2 =$$
$$= 2\sum_{i=1}^{n}\left(x_i-\bar{x}_A\right)^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\left(x_i-\bar{x}_A\right)\left(x_j-\bar{x}_A\right)$$

Elkarrosagarritasunaren ondorioz, espazio-autokorrelazioa neurtzeko nahikoa da bietako bat erabiltzea.

*Geary*ren adierazle klasikoaren balioak *0* eta *2* balioen artean daude gehienetan, baina, *Moran*en adierazle klasikoaren kasuan bezala, muga hauetatik at ere koka daiteke. *Geary*ren adierazle aldatuaren kasuan, berriz, bi muga horiek finkoak dira eta ez dira inoiz gainditzen.

$$0 \leq C^* \leq 2$$

0 balioak adierazten du erabateko autokorrelazio positiboa (objektu guztien balioak bere ingurukoak beste dira, $x_i = x_j$), *2* balioak erabateko autokorrelazio negatiboa, eta erdibideko *1* balioak espazio-autokorrelazio eza. *Moran*en adierazlearen kasuan bezala, bi muga horietaraino iritsi ahal izatea *A* matrizearen egituraren barruan dago.

*Geary*ren adierazlearen definizioa ondoz ondokoak diren objektuen arteko antzekotasun eza edo distantzian oinarritzen da. Agerian geratzen da, beraz, adierazle hau tresna baliotsua dela espazioko datu-analisiko teknika miatzailea den sailkapen automatikoan erabiltzeko.

ERREFERENTZIAK

[1] MORAN, P.A.P. (1950): «Notes on Continuous Stochastic Phenomena», *Biometrika 37*, pp. 17-33.

[2] GEARY, R.C. (1954): «The Contiguity Ratio and Statistical Mapping», *5. The Incorporated Statistician*. pp. 115-145.
[3] LEBART L., MORINEAU A. PIRON M.L. (1998): *Statistique Exploratoire Multidimensionnelle*. Dunod, Paris.