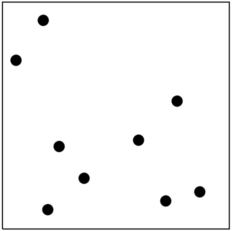
Analisi espaziala

Josemari Sarasola

Gizapedia

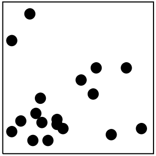
Sarrera

Har dezagun puntuen banaketa bat (minbizia dutenen bizitokiak, salgai dauden etxebizitzak, ...) espazio batean:

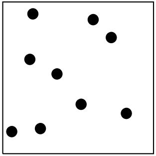


Egitura espazial motak

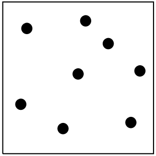
Nolako egitura espazial mota (spatial pattern) ditugu?



Multzokatua

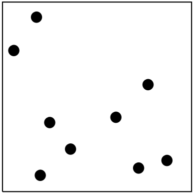


Zorizkoa



Erregularra (overdispersion)

Batzuetan ez dago hain garbi. Nola erabaki zein egiturari dagokio banaketa jakin bat?



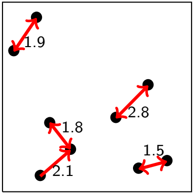
?

Egitura espazialak eta Poisson prozesuak

Problemak badu lotura Poisson prozesuekin: Poisson prozesuak sortzen dira zorizkotasuna erabatekoa denean. Beste alde batetik, Poisson banaketak gertaera puntualen banaketa adierazten du denboran zehar, baina baita ere espazioan zehar (zergatik ez, ba): λ parametroak espazioko unitateko batez besteko puntu dentsitatea adieraziko luke. Horrela, puntuak zoriz banatzen direla erabakitzen bada, Poisson banaketa erabil daiteke puntu kopuru jakinetarako probabilitateak kalkulatzeko.

Zorizkoa? Ondoko gertuenaren metodoa

Puntu bakoitzetik gertuen dagoen punturako distantziak kalkulatzen dira. Ikus ezazu (gezi bikoitzak bi puntuak elkarren gertuenekoak direla adierazten du):



Distantzia horien batezbestekoa kalkulatzen dugu:

$$\bar{d} = \frac{2.1 + 1.8 + 1.8 + 1.9 + 1.9 + 2.8 + 2.8 + 1.5 + 1.5}{9} = 2.01$$

Zorizkoa? Ondoko gertuenaren metodoa

Egitura multzokatuetan batez besteko distantzia txikia izaten da, egitura erregularretan handia, eta zorizkoetan ertaina. Baina noiz da handia, txikia, ertaina?

Froga daiteke zorizko egituretan, batez besteko distantzia λ puntu-dentsitatearen mendean dagoela. λ dentsitatea puntu kopurua eremuko azalerarekin zatituz kalkulatzen da:

$$\lambda = \frac{n}{A}$$

Eta hortik, zorizko egitura batean espero den batez besteko distantzia honela kalkulatzen da:

$$E[d] = \frac{1}{2\sqrt{\lambda}}$$

Zorizkoa? Ondoko gertuenaren metodoa

Adibidean, eremuko azalera $10 \times 10 = 100$ da. 9 puntu daude.
Beraz, dentsitatea 9/100=0.09 da.
Eta itxarondako batez besteko distantzia, zorizko egitura batean:

$$E[d] = \frac{1}{2\sqrt{0.09}} = 1.66$$

Beraz, horrela har liteke erabakia:

- $\overline{d} \approx E(d) \rightarrow$ zorizkotasuna
- $\overline{d} < E(d) \rightarrow$ multzokatzea (clustering)
- $\overline{d} > E(d) \rightarrow$ erregulartasuna

Zorizkoa? Ondoko gertuenaren metodoa

Beraz, 2.01eko batez besteko distantziarekin, puntu egitura zertxobait erregularra dugula esan daiteke, baino zorizko egitura ere adieraz dezake. Zorizkotasunetik aski urrun dagoen eta beraz multzokatzea baiezta daitekeen erabakitzeko, 2.01 (edo gehiago) 1.66tik zenbateraino den "arraroa" neurtu beharko genuke, proba estatistiko baten bitartez.
Beste modu batera, 1.66ko balio ez genuke balio finko eta zehatz gisa hartu behar, eta kasualitatea (lagin errorea) jasotzen duen tarte bat eratu beharko litzateke. Zein da tarte hori?

Zorizkoa? Ondoko gertuenaren metodoa

%90eko konfiantza baterako, zorizkotasuna adierazten duen batez besteko distantziaren tartea hau da, n puntu kopurua izanik eta $\lambda = n/A$ dentsitatea:

$$\frac{1}{2\sqrt{\lambda}} \pm 1.64\sqrt{\frac{4-\pi}{4\pi n\lambda}}$$

Ohartzekoa da puntuak zenbat eta gehiago izan, tartea orduan eta estuagoa izango dela (logikoa guztiz, datu gehiagorekin kasualitatearen efektua txikiagoa delako).
Gure adibidera aplikatuta, hau da tartea:

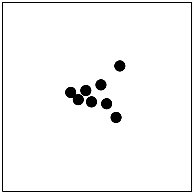
$$\frac{1}{2\sqrt{0.09}} \pm 1.64\sqrt{\frac{4-\pi}{4\pi \times 20 \times 0.09}} : (1.66 \pm 0.32) : (1.34, 1.98)$$

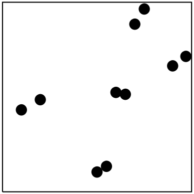
Suertatu den batez besteko distantzia, 2.01 alegia, tartetik dagoenez, eta goitik gainera, puntuen egitura erregularki banatzen dela esan daiteke %90eko konfiantzaz.

Zorizkoa? Ondoko gertuenaren metodoa

Ondoko gertuenaren metodoak baditu akats batzuk:

(1) Adibidez, honako bi egitura hauek desberdinak dira, baino batez besteko distantzia berdintsuak emango lituzkete. Beraz, gertueneko distantziak ez ditu behar bezala bereizten luku edo errazimo gisako multzoak (eskuinean).





(2) perspektibak eragin handia batez besteko distantzian. Ondoko iruduan, gertutik ikusita egitura erregularra genuke, baina urrutitik multzokatzea:

Zorizkoa? K funtzioaren metodoa

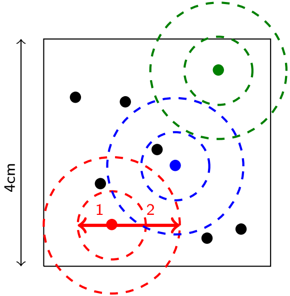
K funtzioaren metodoak ez du gertueneko distantziaren metodoaren lehenengo akatsa. Zertan datza?

1. Puntu bakoitzetik h distantzia desberdinetara, hots h erradioko zirkulu batean batean, dagoen p_i puntu kopurua zenbatzen da. p_i kopuru hauen P batura kalkulatzen da
2. K funtzioa kalkulatzen da, h desberdinetarako:

$$K(h) = \frac{P}{n\lambda}$$

3. $K(h)$ balio horiek puntuen zorizko banaketa batean suertatuko liratekeen batezbestekoekin alderatzen da, multzokatzea, zorizkotasuna edo erregulartasuna atzemateko.

Zorizkoa? K funtzioaren metodoa



$$n = 9$$
$$azalera = 4cm \times 4cm = 16$$
$$\lambda = 9/16 = 0.56$$

Puntua	K(1cm)	K(2cm)
p_1	0	1
p_2	1	1
p_3	0	0
P	1	2
$K(h)$	$\frac{1}{3 \times 0.56} = 0.59$	$\frac{2}{3 \times 0.56} = 1.18$

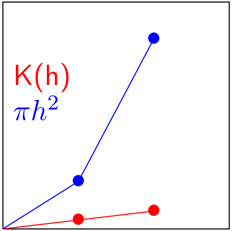
Oharra: Zenbaketa puntu guztietarako egin behar da. Aurrekoa 3 puntuetarako baino ez da egin, adibide gisa.

Interpretazioa

Zorizkotasunean πh^2 da espero den batez besteko K(h) balioa. Beraz, interpretatzeko guri suertatu zaizkigun K(h) balioak harekin alderatu behar dira:

- $K(h) \sim \pi h^2 \rightarrow$ zorizkoa
- $K(h) > \pi h^2 \rightarrow$ multzokatzea
- $K(h) < \pi h^2 \rightarrow$ erregulartasuna

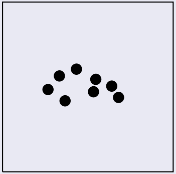
h	$K(h)$	πh^2
1	0.59	3.14
2	1.18	12.56

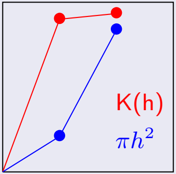


Gure kasuan: $K(h) < \pi h^2$. Beraz, erregulartasuna dagoela ondoriozta daiteke.

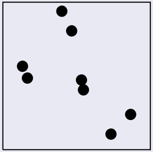
Zorizkoa? K funtzioaren metodoa

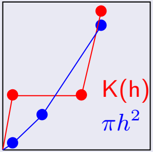
Ohiko egitura espazialak: Multzokatua (cluster bakarra)





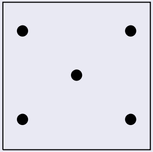
Ohiko egitura espazialak: Multzokatua (cluster anitz)

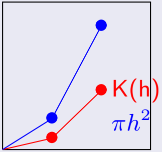




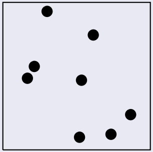
Zorizkoa? K funtzioaren metodoa

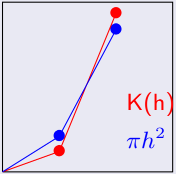
Ohiko egitura espazialak: Erregularra *(overdispersion)*





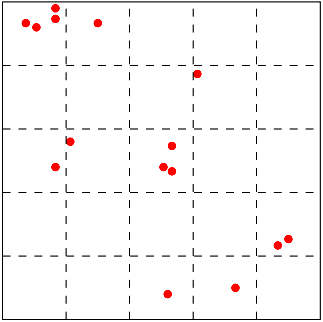
Ohiko egitura espazialak: Zorizkoa





Zorizkoa? Karratuen metodoa

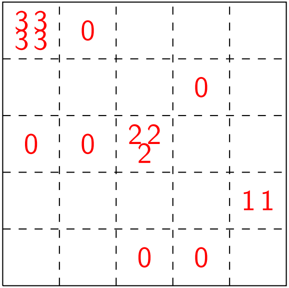
Aztertu beharreko azalera karratu (ingelesez, *quadrat*) direlakoetan banatzen da (laukizuzenak ere izan daitezke, gero karratu bakoitzean dagoen puntu kopurua zenbatzeko.



4	1	0	0	0
0	0	0	1	0
1	1	3	0	0
0	0	0	0	2
0	0	1	1	0

Zorizkoa? Karratuen metodoa: Lloyd-en mordokako batezbestekoa

Lloyden metodoa garatzeko lehen pausoa *crowding mean* edo mordokako batezbestekoa kalkulatzea da, pilaketaren neurri bat. Horretarako, karratuko puntuetako bakoitza karratu horretan zenbat "lagun" dituen zenbatzen da, eta berdin karratu guztietarako eta gero balio horien batezbestekoa kalkulatzen da:



Zorizkoa? Karratuen metodoa: Lloyd-en mordokako batezbestekoa

Mordokako batezbestekoa "lagun" kopuru horien batezbestekoa besterik ez da:

$$m^* = \frac{3+3+3+3+0+0+0+0+2+2+2+1+1+0+0}{15 \text{ puntu}} = 1.33$$

Orain, karratuko batez besteko kopurua kalkulatzen da:

$$m = \frac{4+1+0+\cdots+1+0}{25 \text{ karratu}} = 0.66$$

Zorizkoa? Karratuen metodoa: Lloyd-en *patchiness* indizea

Orain irregulartasun-indizea kalkulatzen da (ingelesez, *patchiness index*):

Patchiness indizea

$$\frac{m^*}{m}$$

Indize hau zenbat eta handiagoa izan, multzokatzea (mordoka pilatzeko joera) hainbat eta handiagoa izango da:

- $\dfrac{m*}{m} \approx 1 \rightarrow$ banaketa zorizkoa (eta beraz, Poisson gisakoa)
- $\dfrac{m*}{m} > 1 \rightarrow$ banaketa pilatua (clustering)
- $\dfrac{m*}{m} < 1 \rightarrow$ banaketa uniforme edo erregularra (puntu kopuruak berdintsuak izateko joera)

Adibidean, $\dfrac{m*}{m} = 2$. Beraz, banaketa pilatua da.