

10. fejezet

FÜGGELÉK

10.1. Matematikai statisztika, geostatisztika

10.1.1. A matematikai statisztika szerepe a térinformatikában

A matematikai statisztika a valószínűség-számításnak az a fejezete, amely megfigyelések és mérések eredményeiből (statisztikai adatokból) következtet valószínűségekre vagy eloszlásfüggvények ismeretlen paramétereire (Természettudományi lexikon, 1967).

Matematikai statisztikai eljárásokat a térinformatika számos feladatának megoldásakor felhasználnak. Példaként említjük a következőket:

- vonatkozási rendszerek kapcsolatát jellemző transzformációs egyenletek paramétereinek meghatározása,
- adatgyűjtési eljárások minőségellenőrzése,
- különböző jellegű elemzési, megjelenítési feladatok megoldása.

A matematikai statisztikai módszerek alkalmazásakor általában a következő típusú részfeladatok fordulnak elő:

- mintavételezés,
- paraméterbecslés,
- hipotézisvizsgálat.

A matematikai statisztikai eljárásokat szokás a következőképpen csoportosítani:

- egyváltozós,
- kétváltozós,
- többváltozós

eljárások. A térinformatikában a leggyakrabban az egy-, illetve kétváltozós eljárásokat alkalmazzák.

A függelékben a szokásos matematikai statisztikai eljárások mellett külön foglalkozunk:

- a legkisebb négyzetek módszerével,
- az interpolációs eljárásokkal,
- a szűrési eljárásokkal,
- a geostatistika alapjaival.

10.1.2. A mintavételezés

A mintavételezés célja valamely alapsokaságból egy további feldolgozásra kerülő rész (az ún. minta) kiválasztása. A mintát úgy kell megválasztani, hogy abból az alapsokaság egészére megbízható következtetéseket lehessen levonni. A mintavételezés minősége meghatározza a későbbi tevékenység minőségét is.

A térinformatikai mintavételezés Longley (2011) alapján lehet:

- véletlenszerű,
- szisztematikus, azonos rácsmérettel,
- szisztematikus, változó rácsmérettel,
- klaszteres,
- izovonal menti.

A felsorolt eseteket szemlélteti a 10.1. ábra.

A felsoroltak mellett elsősorban minőségellenőrzéskor alkalmazzák a szekvenciális mintavételezést is. A szekvenciális mintavételezéskor – amelynek elméletét Abraham Wald magyar származású matematikus dolgozta ki – a vizsgálatot kis elemszámú mintával kezdjük, s az annak alapján végzett számítás eredményét felhasználva döntünk újabb mintaelemek kiválasztásáról.

A mintavételezés eredményeként létrejövő minta elemeit a továbbiakban valószínűségi változóknak tekintjük.

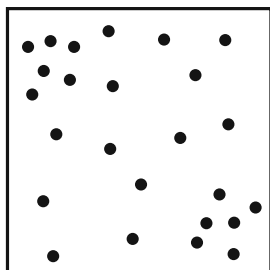
10.1.3. Paraméterbecslés

Egyváltozós eset

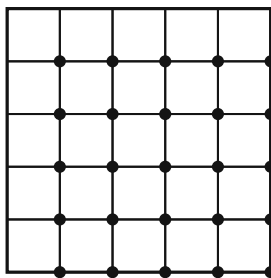
Induljunk ki az egyetlen változóra vett x_1, x_2, \dots, x_n mintából. Egyváltozós mintát alkothatnak valamely objektumosztály objektumainak azonos típusú attribútumai. Például mintaelemnek tekinthetjük egy üzleti célú térinformatikai rendszerben a vizsgált települések népességét.

A minta jellemzésére felhasználható első mennyiség a minta D terjedelme.

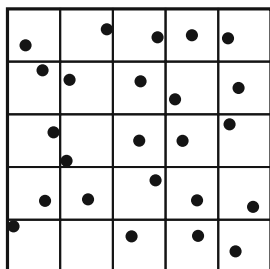
$$D = x_{\max} - x_{\min},$$



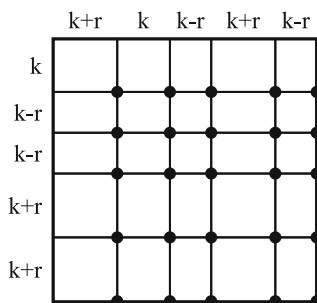
a) véletlenszerű



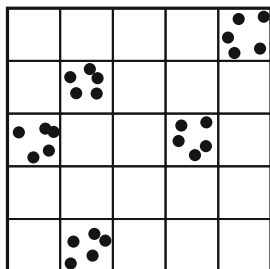
b) szisztematikus, azonos rácsmértű



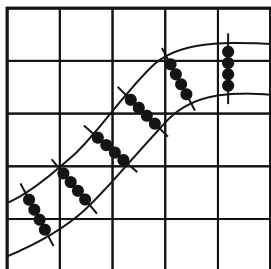
c) szisztematikus, véletlenszerű



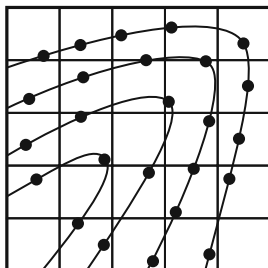
d) szisztematikus, változó rácsmértű



e) klaszteres



f) profilmenti



g) izovonalmenti mintavételezés

10.1. ábra. Mintavételezés esetei (Longley et al. alapján)

ahol $x_{\max} = \max(x_i)$, $x_{\min} = \min(x_i)$.

A paraméterbecslés megkezdése előtt a minta alapján gyakran előállítják az eloszlás tapasztalati sűrűség függvényét, az ún. hisztogramot. A hisztogram – mint azt a 8. fejezetben leírtuk – egyrészt alkalmas a szemléletes megjelenítésre, másrészt egyszerűbbé teheti a későbbi számításokat. A hisztogram előállításakor a minta terjedelmét k osztályra bontják oly módon, hogy a legkisebb mintaelem az első, a legnagyobb mintaelem az utolsó osztályba essék. Az osztályok számának és az osztópontok helyének meghatározására nincsen pontos előírás. Az osztályok k számának meghatározására különböző szabályokat ajánlanak, amelyek közül a következőket mutatjuk be:

$$\begin{aligned} 5 &\leq k \leq 12, \\ k &\leq 5 \log(n), \\ k &\approx n^{1/2} \end{aligned}$$

Az első összefüggés a kisebb elemszámú minták esetén, a két utóbbi pedig a nagyobb elemszámú minták esetén használható. Gyakorlati szempontból előnyös, ha a z_j osztópontok kerek értékűek, és az egyes $z_j - z_{j-1}$ intervallumok (osztályok) azonos méretűek. Az egyes osztályokba eső mintaelemek száma adja meg az osztályhoz tartozó abszolút gyakoriságot. Például a $z_j - z_{j-1}$ intervallum által meghatározott j -edik osztály esetén az abszolút gyakoriság legyen

$$f_j(x).$$

A relatív gyakoriság a következő:

$$h_j(x) = f_j(x)/n$$

A relatív gyakoriságot gyakran százalékosan adják meg:

$$p_j(x) = h_j(x) \cdot 100\%$$

A felsorolt gyakoriságokat grafikusán ábrázolva jutunk a hisztogramhoz. A sűrűségi hisztogram előállításakor az egyes osztályok z_j , z_{j-1} határpontjai fölé $h_j(x)$ magasságú téglalapot rajzolnak. A sűrűségi jelzót azért használtuk, mivel az $f_j(x)$ értékek alapján is lehetséges hisztogramszerkesztés, ez utóbbit gyakorisági hisztogramnak nevezik.

Egyetlen változó esetén a paraméterek részben a változók valamely jellemző értékkel történő helyettesítésére, részben a jellemző értéktől történő eltérés mértékének tükrözésére szolgálnak. Az említett példában ilyen jellemző érték lehet

a települések átlagos népessége. A most leírt jellemző értékek lehetnek a következő paraméterek:

- mintaközép: $\underline{a} = 1/n \sum x_i$
- módusz: a leggyakrabban előforduló érték (a hisztogram maximuma),
- medián \underline{m}_e : az ún. rendezett minta közepe.

A rendezett mintát az eredeti mintából úgy nyerjük, hogy a minta elemeit növekvő sorrendbe rendezzük:

$$x_1^* \leq x_2^* \leq \dots \leq x_n^*$$

Ha $n = 2m + 1$, akkor $m_e = x_{m+1}^*$.

Ha $n = 2m$, akkor $m_e = 1/2(x_m^* + x_{m+1}^*)$.

A három jellemző paraméter közül legtöbbször a mintaközepet használják. Ha feltételezhető, hogy a minta elemeinek valamelyikét jelentős hiba terheli, akkor célszerűbb a medián vagy a módusz használata.

A minta elemeinek a mintaközéptől való eltérésének jellemzésére szolgáló mennyiségek:

- a (tapasztalati) variancia: $\underline{s}^2 = (1/(n-1))\sum(x_i - \underline{a})^2$
- a (tapasztalati) szórás: $\underline{s} = (\underline{s}^2)^{1/2}$

A szórás a varianciából vont négyzetgyök pozitív értékeként értelmezik.

A matematikai statisztikai szakirodalomban a (tapasztalati) variancia számítására használják az

$$\underline{s}^2 = (1/n)\sum(x_i - \underline{a})^2$$

összefüggést is. Az így számított érték az elméleti varianciának ún. torzított becslése. Ha n értéke nagy, a két összefüggés gyakorlatilag egyenértékű.

A levezetett paraméterek az eredeti sokaság ismeretlen paramétereinek becslései. Például a mintaközép a várható érték becslése. A tapasztalati jelző az ismeretlen elméleti értéktől való megkülönböztetést szolgálja.

Kétfváltozós eset

Két változó esetén a minta elemei a következők:

$$x_1, x_2, \dots, x_n,$$

$$y_1, y_2, \dots, y_n$$

Az előbb említett üzleti célú rendszerben kétfváltozós mintát alkothat az egyes települések népessége és az egy főre eső személygépkocsik száma a településeken. A minta elemeihez rendelt osztályok grafikus ábrázolása a hisztogramhoz hasonló módon, de két dimenzióban történhet.

A kétváltozós mintát jellemző paraméterek:

- az egyes változókra külön-külön az egyváltozós esetről bemutatott módon kiszámított várható értékek ($\underline{a}_x, \underline{a}_y$) és szórások ($\underline{s}_x, \underline{s}_y$),
- a két változó kapcsolatát jellemző (tapasztalati) kovariancia:

$$\underline{c}_{xy} = (1/(n-1)) \sum((x_i - \underline{a}_x)(y_i - \underline{a}_y))$$

A tapasztalati kovariancia tetszőleges értékeket felvehet, ezért nehezen értelmezhető. A könnyebb értelmezhetőséget szolgálja a (tapasztalati) korrelációs együttható:

$$\underline{r}_{xy} = \underline{c}_{xy} / (\underline{s}_x \cdot \underline{s}_y)$$

A korrelációs együtthatóra igaz a következő:

$$-1 \leq \underline{r}_{xy} \leq +1$$

Független valószínűségi változók közötti korrelációs együttható értéke 0. (Ez az állítás azonban nem fordítható meg, azaz ha két valószínűségi változó közötti korrelációs együttható 0, abból még nem következik, hogy a változók függetlenek). A +1 és a -1 érték azt jelzi, hogy a változók között lineáris függvénykapcsolat áll fenn. A + előjel azt mutatja, hogy az egyik változó növekedése a másik változó növekedését vonja maga után. Említett példánkban a korrelációs együttható arról nyújthat felvilágosítást, hogy kapcsolatban van-e az egyes települések népessége az egyes településeken egy főre jutó gépkocsik számával. A szakirodalomban a korrelációs együttható mellett találkozhatunk az r^2 determinisztikus index fogalmával is, amely 0 és 1 közötti értékeket vehet fel. Minél közelebb van a determinisztikus index értéke 1-hez, annál szorosabb a mennyiségek közötti kapcsolat.

A lineáris függvénykapcsolatot leíró egyenes az ún. regressziós egyenes, amelynek együtthatói az eddig levezetett paraméterekből meghatározhatók. Legegyenes az egyenes

$$Y = mX + b$$

alakú, ahol $m = \underline{r}_{xy} \cdot \underline{s}_y$ és $b = \underline{a}_y - m\underline{a}_x$

Regressziós egyenes képét tüntettük fel a 8.2. ábrán. A regressziós egyenes mellett egyéb függvénykapcsolatok is elképzelhetők.

Többváltozós eset

Többváltozós esetben a minta többváltozós. Példaként említhetjük az egyes települések népességét, a településen az egy főre jutó gépkocsik számát, a gépkocsik átlagos értékét stb. A számítások célja különböző lehet, például (Horvai, 2001):

- paraméterbecslés: az egyes változók jellemző paramétereinek meghatározása.
- faktoranalízis: a változóknak a lehetséges legkisebb számú közös faktor segítségével történő kifejezése,
- főkomponens-elemzés: az eredeti – egymással korrelált – változók helyettesítése kevesebb, korrelálatlan változóval,
- regresszióanalízis: többváltozós függvények meghatározása.

A számításba bevonhatók nem valószínűségi változó jellegű ismeretlen mennyiségek is. A felsorolt feladatok különböző módszerekkel oldhatók meg. A lehetséges megoldási módok közül a legkisebb négyzetek módszerét a későbbiekben részletesebben tárgyaljuk. A különböző módszerekről áttekintést nyújt például Detrekői (1991), Horvai (2001), Korn, Korn (1975).

10.1.4. Hipotézisvizsgálatok

A hipotézisvizsgálatok célja egy vagy több valószínűségi változóra vonatkozó feltevések (hipotézisek) ellenőrzése (Farkas 1972). A feltevéseknek különböző fajtái fordulnak elő. Például:

- egy eloszlás valamely paraméterére vonatkozó hipotézis (például a várható érték egyezik-e egy előre adott értékkel),
- két eloszlás paramétereire vonatkozó hipotézis (például azonosnak tekinthető-e két eloszlás várható értéke),
- két vagy több valószínűségi változó függetlenségének vizsgálata (például szignifikánsan eltér-e valamely korrelációs együttható a zérustól),
- két vagy több valószínűségi változó homogenitásának vizsgálata (megegyezik-e az eloszlásuk).

A felsorolt példák közül a függetlenségvizsgálatot a különböző attribútumok kapcsolatának elemzésekor, a homogenitásvizsgálatot az adatok minőségellenőrzésekor alkalmazzák a térinformatikában. Létezik olyan hipotézisvizsgálat, amely a minta véletlen jellegének ellenőrzésére alkalmas (Longley, 2011).

A hipotézisvizsgálatok gyakorlati megvalósítására az ún. statisztikai próbák szolgálnak. Ezek gyakorlati megvalósítását – számpéldákkal együtt – tartalmazza például Detrekői (1991), Longley (2011). A statisztikai próbák végzésekor általában a vizsgált valószínűségi változó valamilyen eloszlását fel kell tételeznünk.

10.1.5. A legkisebb négyzetek módszere

A matematikai statisztika feladatainak megoldásakor gyakran használt eljárás Gauss nevéhez fűződik, aki 1794 óta foglalkozott az eljárással. Az eljárás

használható például vetületi átszámítások transzformációs egyenletei együtthatóinak, vagy elemzéskor mért pontokra illesztendő felület egyenletének meghatározásakor.

Induljunk ki N mennyiségre vett L_1, L_2, \dots, L_N elemű mintából. Tételezzük fel, hogy a minta elemei függetlenek és azonos pontosságúak. Legyen feladatunk a minta elemeivel lineáris kapcsolatban lévő R számú ismeretlen Z_1, Z_2, \dots, Z_R paraméter meghatározása. A megoldás előfeltétele:

$$F = N - R > 0.$$

Az egyenlőtlenség azt fejezi ki, hogy a minta elemeinek száma nagyobb a meghatározandó paraméterek számánál. Az F mennyiséget redundanciának vagy fölös mérésszámnak nevezik.

A minta elemei és a meghatározandó paraméterek kapcsolatát az ún. közvetítő egyenletek írják le. (közvetítő egyenlet lehet például a pontokra illesztendő sík egyenlete). Minden közvetítő egyenletben egyetlen mintaelem U_i ún. kiegyenlített értéke és a meghatározandó paraméterek lineáris függvénye szerepel. A kiegyenlített értéket az eredeti értékből a v_i javítások (rezíduumok) hozzáadásával nyerjük:

$$U_i = L_i + v_i$$

Az i -edik kiegyenlített értékhez tartozó közvetítő egyenlet a következő:

$$U_i = L_i + v_i = a_{0i} + a_{1i}Z_1 + a_{2i}Z_2 + \dots + a_{1R}Z_R$$

Az egyenletben szereplő a_{ji} mennyiségek együtthatók. A közvetítő egyenlet átalakításával jutunk a javítási egyenlethez:

$$v_i = a_{0i} + a_{1i}Z_1 + a_{2i}Z_2 + \dots + a_{1R}Z_R - L_i$$

A javítási egyenleteket vektorok és mátrixok felhasználásával a következő alakban írhatjuk:

$$\mathbf{v} = \mathbf{A} \mathbf{Z} - \mathbf{L}$$

Az egyenletben szereplő, az a_{ji} együtthatókat tartalmazó \mathbf{A} mátrixot a szakirodalom alakmátrixnak, vagy design-mátrixnak nevezi.

Az N számú javítás N számú ismeretlen v_i javítást és R számú ismeretlen Z_j paramétert, tehát összesen $N + R$ ismeretlent tartalmaz. A feladat alulhatározottságát a

$$\mathbf{v}^* \mathbf{v} = \sum v_i^2 = \min$$

feltétel felhasználásával küszöbölik ki. A feltétel a javítások lehetséges rendszerei közül azt rendeli a minta elemeihez, amelyeknek a négyzetösszege a legkisebb. Innen származik a legkisebb négyzetek módszere elnevezés.

A felírt feltétel alapján jutunk a

$$(\mathbf{A}^*\mathbf{A})\mathbf{Z} = \mathbf{A}^*\mathbf{L}$$

normálegyenlethez, amely R egyenletből áll és R ismeretlent tartalmaz. Az egyenletet megoldva megkapjuk az ismeretlen Z_j paramétereket. A megoldást formálisan a következő alakban írhatjuk:

$$\mathbf{Z} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{L}$$

(feltételezve, hogy az $\mathbf{N} = (\mathbf{A}^*\mathbf{A})$ nem szinguláris).

A paraméterek ismeretében a javítási egyenletekből kiszámíthatjuk az egyes v_i javításokat, azok alapján pedig a

$$U_i = L_i + v_i$$

egyenletből a kiegyenlített értékeket.

A legkisebb négyzetek módszere alkalmas a meghatározott mennyiségek varianciáinak, szórásainak a becslésére. A becsléshez elsőként a következő varianciát határozzuk meg:

$$s_0^2 = F^{-1} \sum v_i^2$$

Ez az érték a mintaelemeket együttesen jellemző mennyiség. A mennyiség s_0 négyzetgyökét felhasználva az egyes Z_j paraméterek szórását a következő összefüggésből kapjuk:

$$s_j = s_0(q_{jj})^{-1/2}$$

ahol q_{jj} az $(\mathbf{A}^*\mathbf{A})^{-1}$ mátrix főátlójának j -edik eleme.

Hasonlóképpen kiszámíthatjuk az U_i kiegyenlített értékek szórásait is:

$$s_j = s_0(q_{ii})^{-1/2}$$

ahol q_{ii} az $\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ mátrix főátlójának i -edik eleme.

A legkisebb négyzetek módszerének most ismertetett speciális esete általánosítható a következő módokon:

- az eredeti L_i mennyiségek nem függetlenek és nem azonos pontosságúak,
- a közvetítő egyenletek helyett több kiegyenlített értéket és több paramétert tartalmazó feltételi egyenleteket használunk,
- a kapcsolatokat leíró közvetítő egyenletek nem lineárisak,
- az $(\mathbf{A}^*\mathbf{A})$ mátrix szinguláris.

A felsorolt általánosabb esetek tárgyalása meghaladja jelen könyv kereteit. Részletesebb leírásukat például Detrekői (1992), Horvai (2001) tartalmazza.

Példa. Adott a síkon négy pont. A pontok koordinátái ismertek az x' , y' és az x , y koordináta-rendszerekben. A pontok koordinátáit a következő táblázat tartalmazza:

Pont	$x'(m)$	$y'(m)$	$x(m)$	$y(m)$
1	525,23	111,11	1532,09	300,8
2	718,41	231,84	1692,28	453,4
3	637,12	400,17	1582,93	605,02
4	820,41	280,65	1783,85	519,12

Tételezzük fel, hogy a két rendszer között az 5.6.2. pontban bemutatott hasonlósági transzformációval a következő egyenletek alapján kapcsolat teremthető:

$$\begin{aligned}x &= a_0 + a_1x' - a_2y' \\ y &= b_0 + a_1y' + a_2x'\end{aligned}$$

Tűzzük ki célul a transzformációs egyenletek ismeretlen

$$a_0, b_0, a_1, a_2$$

paramétereinek meghatározását a legkisebb négyzetek módszerével. A számításkor a minta elemeinek az x , y koordinátákat tekintjük, a felsorolt paraméterek felelnek meg az általános tárgyalás Z_j paramétereinek, az együtthatók szerepét az x' , y' mennyiségek töltik be.

Valamely i -edik ponthoz tartozó közvetítő egyenlet a következő:

$$\begin{aligned}x_i + v_{xi} &= a_0 + a_1x'_i - a_2y'_i \\ y_i + v_{yi} &= b_0 + a_1y'_i + a_2x'_i\end{aligned}$$

A közvetítőegyenletekből előállíthatók a javítási egyenletek:

$$\begin{aligned}v_{xi} &= a_0 + a_1x'_i - a_2y'_i - x_i \\ v_{yi} &= b_0 + a_1y'_i + a_2x'_i - y_i\end{aligned}$$

A javítási egyenlet alapján az \mathbf{A} alakmátrix és az \mathbf{L} vektor felépítése a következő:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 525,32 & -111,11 \\ 0 & 1 & 111,11 & 525,32 \\ 1 & 0 & 718,41 & -231,84 \\ 0 & 1 & 231,84 & 718,41 \\ 1 & 0 & 637,12 & -400,17 \\ 0 & 1 & 400,17 & 637,12 \\ 1 & 0 & 820,01 & -280,65 \\ 0 & 1 & 280,65 & 820,01 \end{bmatrix} \quad \mathbf{-L} = \begin{bmatrix} 1532,09 \\ 300,80 \\ 1692,28 \\ 453,40 \\ 1582,93 \\ 605,02 \\ 1783,85 \\ 519,12 \end{bmatrix}$$

A normálegyenletet alkotó $\mathbf{N} = (\mathbf{A} * \mathbf{A})$ mátrix és $\mathbf{n} = \mathbf{A} * \mathbf{L}$ így alakul:

$$\mathbf{N} = \begin{bmatrix} 4 & 0 & 2700,86 & -1023,77 \\ 0 & 4 & 1023,77 & 2700,86 \\ 2700,86 & 1023,77 & 2175407,99 & 0 \\ -1023,77 & 2700,86 & 0 & 2175407,99 \end{bmatrix}$$

$$\mathbf{n} = \begin{bmatrix} 6591,15 \\ 1878,34 \\ 5018219,62 \\ -401750,03 \end{bmatrix}$$

A normálegyenlet megoldásából kapjuk az ismeretlenek vektorát

$$\mathbf{Z} = \mathbf{N}^{-1} \mathbf{n} = \begin{bmatrix} 1041,099571 \\ 93,968833 \\ 0,970003 \\ 0,188608 \end{bmatrix}$$

A vektorból adódik:

$$a_0 = 1041,099 \text{ m} \quad b_0 = 93,969 \text{ m}$$

Ezek az értékek a két rendszer közötti eltolást jellemzik.

$$a_1 = 0,970003 \quad a_2 = 0,188608.$$

Ezekben az értékekben együtt jelentkeznek az elforgatás és a méretarány-különbség hatása.

A két értéket felhasználva kapjuk

- a két rendszer közötti elfordulási szöveget: $\alpha = \arctan(a_1/a_2) = 78,996661^\circ$
- a két rendszer közötti méretarány-tényezőt: $k = (a_1^2 + a_2^2)^{1/2} = 0,988170.$

10.1.6. Interpolációs eljárások

Az interpolációs eljárások áttekintése

Az interpoláció célja egy függvény értékének közelítő meghatározása olyan esetekben, amikor a függvény pontos értéke bizonyos pontokban ismert, de más közbülső pontokban is szükséges a függvény közelítő értékének ismerete (Akadémiai kislexikon 1989). A leírt meghatározásból következik, hogy az interpolációs során az ismert értékeket hibátlannak tekintjük.

Az interpolációs eljárásokat a térinformatikában mind az adatgyűjtés, mind az elemzési műveletek, mind a megjelenítés esetén felhasználunk. Az interpoláció célja általában attribútumadatok „sűrítése”, azaz olyan pontokhoz is attribútumértékeket rendelünk, ahol azokkal eredetileg nem rendelkezünk. Gyakorlati szempontból célszerű megkülönböztetni az

- egyváltozós (vonal menti),
- kétváltozós (síkbeli),
- háromváltozós (térbeli)

interpolációs eljárásokat. Mindhárom felsorolt esetben megkülönböztethetünk csak az ismeretlen pontokhoz közeli szomszédos pontokon alapuló lokális eljárásokat, és több pontot is figyelembe vevő – különböző függvényeket felhasználó – eljárásokat.

A továbbiakban néhány gyakran alkalmazott eljárást mutatunk be.

Egyváltozós interpolációk

Egyváltozós esetben ismertek az adott t_1, t_2, \dots, t_N értékekhez tartozó s_1, s_2, \dots, s_N értékek. Keresett valamely nem adott t_P értékhez tartozó s_P érték. (Az s_i jelölés az angol nyelvben elterjedt signal = jel elnevezésre utal).

A legközelebbi szomszéd elve alapján történő (lépcsős) interpoláció esetén az interpolált értéket a hozzá legközelebb eső adott értékkel tekintik azonosnak. Az interpoláció első lépéseként meg kell határozni, hogy a t_P érték mely ismert értékek közé esik. Ennek alapján legyen igaz a következő egyenlőtlenség:

$$t_G \leq t_P < t_{G+1}$$

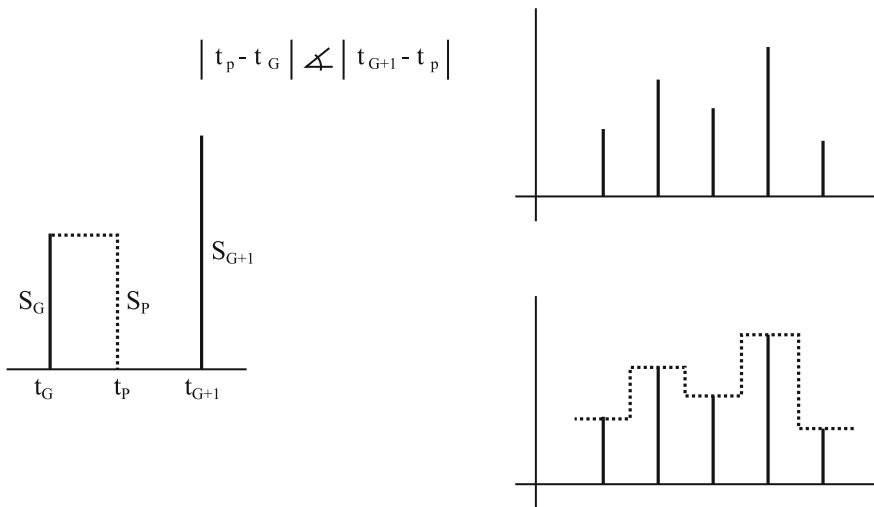
Ekkor

$$s_P = s_G, \text{ ha } |t_P - t_G| < |t_{G+1} - t_P|$$

és

$$s_P = s_{G+1}, \text{ ha } |t_P - t_G| \geq |t_{G+1} - t_P|.$$

A lépcsős interpoláció elvét szemlélteti a 10.2. ábra.



10.2. ábra. Lépcsős interpoláció

A *lineáris interpoláció* esetén az interpolált értéket a két szomszédos értéket összekötő egyenes felhasználásával határozzuk meg. Ennek az eljárásnak is első lépése a t_p értékkel szomszédos értékek meghatározása. Legyenek ezek ismét a következők:

$$t_G \leq t_p < t_{G+1}$$

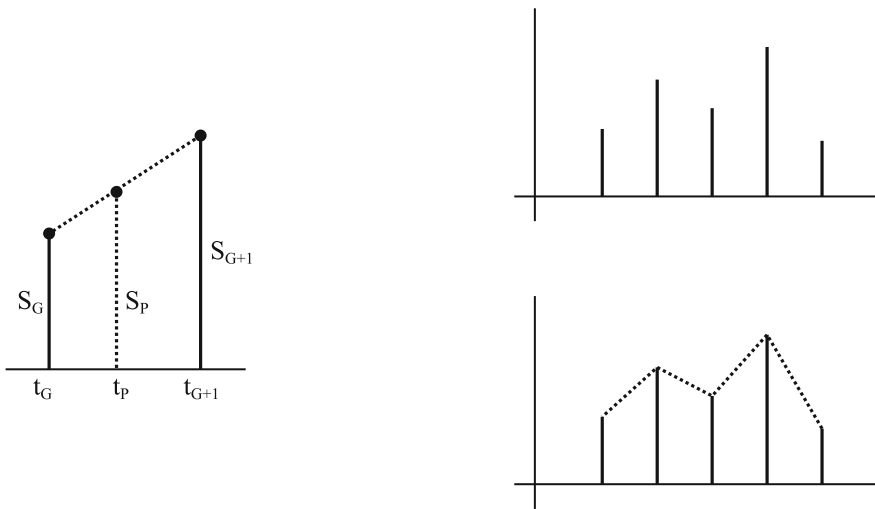
Ekkor

$$s_p = s_G + ((t_p - t_G)(s_{G+1} - s_G)) / (t_{G+1} - t_G).$$

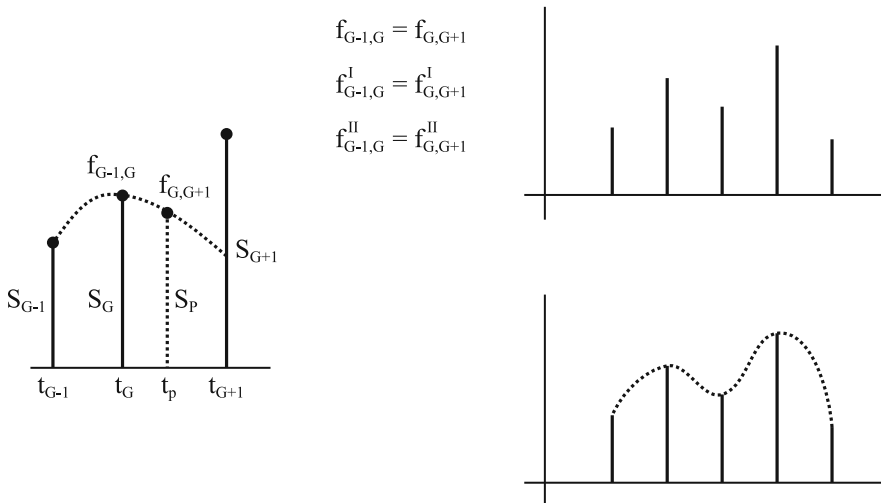
A lineáris interpoláció elvét a 10.3. ábrán tüntettük fel.

A most bemutatott interpolációk viszonylag kevés számítást igényelnek, ugyanakkor – az ábrák tanulsága szerint is – csak meglehetősen durván közelítik a valóságos folyamatokat. Az eddigieknél finomabb közelítést tesz lehetővé a *spline-interpoláció*. Ennél az eljárásnál az interpolációt az ismert pontokra folyamatosan illesztett – általában harmadfokú – parabolák segítségével végzik. Minden $(t_2 - t_1), \dots, (t_G - t_{G-1}), (t_{G+1} - t_G), \dots, (t_N - t_{N-1})$ szakaszra külön parabolát illesztenek oly módon, hogy a szomszédos szakaszokhoz tartozó parabolák

- függvényértékei,
- első deriváltjai (azaz érintői),
- második deriváltjai (azaz görbületei)



10.3. ábra. Lineáris interpoláció



10.4. ábra. Spline-interpoláció

a közös pontban azonosak legyenek. A spline-interpolációt az 10.4. ábrán szemléltetjük. Az eljárás számításigényesebb a korábbiaknál, ugyanakkor jobban közelíti a valóságot. A számításhoz szükséges összefüggések megtalálhatók például Detrekői (1991) munkájában.

Több szomszédos pont együttes hatása figyelembe vehető úgy is, ha az ismeretlen jelet több ismert ponthoz tartozó jel súlyozott számtani közepeként határozzák meg:

$$s_P = (\sum(w_G \cdot s_G)) / \sum(w_G)$$

A súlyok megválasztása különböző módon történhet. A súlyozáskor általában figyelembe veszik az ismert pontok és az ismeretlen pont távolságait:

$$d_G = |t_G - t_P|$$

Súlyozott számtani középként képzett interpolációnak tekinthetjük az itt részletebben nem tárgyalt legkisebb négyzetek módszerén alapuló interpolációt is (Detrekői 1991).

Kétváltozós interpolációk

Kétváltozós (síkbeli) interpolációk esetén ismertek a különböző $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ ponthelyekhez tartozó s_1, s_2, \dots, s_N értékek. Keresett valamely nem adott (x_P, y_P) helyhez tartozó s_P érték. A kétváltozós interpoláció módszerei hasonlítanak az egyváltozós interpoláció módszereihez.

A *legközelebbi szomszéd elve* alapján történő interpoláció esetén az interpolált értéket a hozzá legközelebb eső adott értékkel tekintik azonosnak. A döntés érdekében kiszámítják az adott pontok d_G távolságát az ismeretlen ponttól. Az interpolált értéknek azt az

$$s_P = s_G$$

értéket tekintik, amelyre

$$d_G = \min.$$

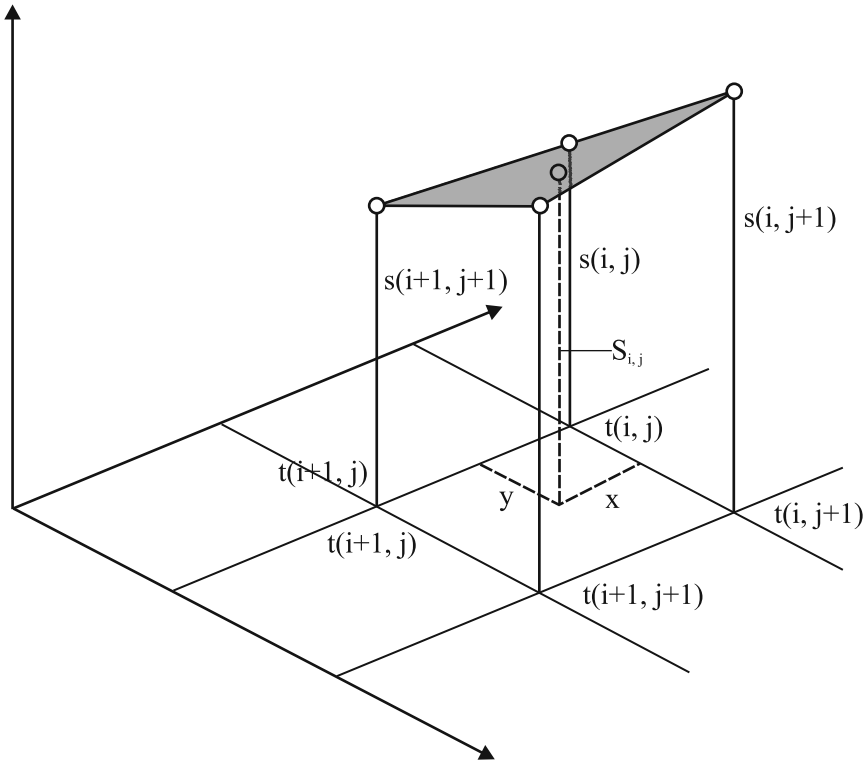
A lineáris interpoláció kétváltozós megfelelőjének tekinthető a *bilineráris interpoláció*. Ekkor az ismeretlen pontot körülvevő négy szomszédos pont alapján végzik az interpolációt oly módon, hogy két-két szomszédos pont jeleire illesztett egyenesekből indulnak ki. Az eljárás viszonylag kevés számítást igényel, ha az adott pontok egy négyzetháló csúcspontjait alkotják (10.5. ábra)

Ha feltételezzük, hogy az adott pontok egységnyi távolságra vannak egymástól, akkor Schenk (1999) alapján az interpolált érték a következő összefüggésekből számítható:

$$R_1 = x \cdot s_{i,j} + (1-x)s_{i+1,j}$$

$$R_2 = x \cdot s_{i,j+1} + (1-x)s_{i+1,j+1}$$

$$s_P = yR_1 + (1-y)R_2$$



10.5. ábra. Bilineáris interpoláció

A most bemutatott összefüggést gyakran alkalmazzák a digitális képfeldolgozásban.

A most leírthoz hasonló módon lehetséges lineáris jellegű interpoláció valamely háromszög belsejében is. A számításhoz szükséges összefüggések megtalálhatók Bill (1999) munkájában.

Több szomszédos pont együttes hatása vehető figyelembe, ha az ismeretlen jelet az ismert pontokhoz tartozó jelek *súlyozott számtani közepeként* határozzák meg:

$$s_P = (\sum(w_G \cdot s_G)) / \sum(w_G)$$

A súlyok megválasztása különböző módon történhet. A súlyozáskor általában figyelembe veszik az ismert pontok és az ismeretlen pont d_G távolságait:

$$d_G = ((x_G - x_P)^2 + (y_G - y_N)^2)^{1/2}$$

Igen gyakran alkalmazzák a

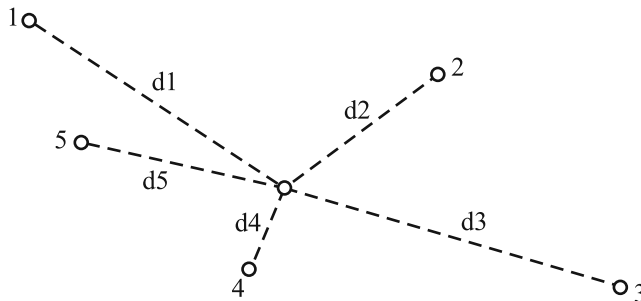
$$w_G = 1/d_G$$

súlyokat. Ilyen eljárásnak tekinthetők a geostatisztika később ismertetésre kerülő módszerei is.

Példa. A 10.6. ábrán feltüntetettünk 5 ismert és egy interpolálandó ismeretlen pontot. Az ismert pontok számát, a pontokban adott jelértékeket és az egyes pontok távolságát az ismeretlen ponttól a táblázat tartalmazza. Határozzuk az ismeretlen ponthoz tartozó jelértéket a

- legközelebbi szomszéd elve alapján,
- a távolságokkal fordítottan arányos súlyokat felhasználva súlyozott számtani középként (A w_G súlyok és a $w_G \cdot s_G$ szorzatok szintén szerepelnek a táblázatban).

G	s	d	w	w_s
1	3	4,30	0,23	0,70
2	4	2,90	0,34	1,38
3	2	5,50	0,18	0,36
4	4	1,00	1,00	4,00



10.6. ábra. A pontok elrendezése

A táblázat alapján a minimális d távolság a 4. ponthoz tartozik. Ennek megfelelően a legközelebbi szomszéd elve alapján interpolált érték:

$$s_P = s_4 = 4.$$

A súlyozott számtani közép alapján az interpolált érték:

$$s_P = (\sum(w_G \cdot s_G)) / \sum(w_G) = 9,44 / 2,25 = 4,19.$$

Háromváltozós (térbeli) interpolációk

Háromváltozós (térbeli) interpolációk esetén ismertek a különböző $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_N, y_N, z_N)$ ponthelyekhez tartozó s_1, s_2, \dots, s_N értékek. Keresett valamely nem adott (x_P, y_P, z_P) helyhez tartozó s_P érték. A háromváltozós interpoláció esetén az eddig bemutatott módszerek közül felhasználhatjuk

- a legközelebbi szomszéd elve alapján történő és
- a súlyozott számtani közepet felhasználó

interpolációs eljárásokat. A feladat megoldásához használhatók a külön pontban tárgyalt geostatistikai módszerek is.

10.1.7. Szűrések

Szűréseknek nevezik általában mindazokat az eljárásokat, amelyekkel az eredetileg nyert értékekről az esetleg fellelhető (többnyire zavaró jellegű) hatásokat leválasztják. Az interpolációhoz hasonlóan itt is kiindulhatunk abból, hogy valamely függvény értéke bizonyos pontokban ismert. Az eltérés az interpolációtól abban jelentkezik, hogy a függvényértékeket nem tekinthetjük hibátlannak. A szűrések feladata az ismert pontokban a zavaró jellegű hatások leválasztása (esetleg kiemelése). A szűrési eljárások egy része alkalmas a nem mért pontokban az ismeretlen értékek meghatározására is. A szűréseket a térinformatikában az adatgyűjtéskor, az elemzéskor és esetenként a megjelenítéskor is széles körben alkalmazzák.

A szűrési eljárások elméleti alapjainak tárgyalása meghaladja jelen függelék célját. Az elméleti alapok bemutatása nélkül megemlíjtük, hogy a szűrések jellegük szerint lehetnek

- alul-áteresztők (céljuk a zavaró jellegű hatások kiküszöbölése),
- felül-áteresztők (céljuk bizonyos hatások kiemelése).

A szűrési eljárások jellegének szemléltetésre egyváltozós alul-áteresztő szűrőt mutatunk be.

Legyenek ismertek az adott t_1, t_2, \dots, t_N pontokhoz tartozó s_1, s_2, \dots, s_N értékek. Az értékeket az interpolációnál felhasznált s_G helyett \underline{s}_G jelöléssel láttuk el. Az aláhúzás az értékek hibával terhelt jellegére utal. A \underline{s}_G mennyiséget terhelő hibát a szakirodalomban zajnak nevezik, s az angol noise szóból kiindulva n_G betűvel jelölik. Az eredeti (hibával terhelt) \underline{s}_G jelek, a szűrt (hibátlannak tekinthető) s_G jelek és a n_G zaj (hiba) között a következő összefüggés áll fenn:

$$\underline{s}_G = s_G + n_G,$$

A szűrési eljárás során az s_G szűrt értéket a vizsgált jelet határoló, a jelhez képest szimmetrikusan elhelyezkedő $2M + 1$ jel súlyozott számtani közepeként állítjuk elő:

$$s_G = \sum (w_J \cdot s_{G+J}),$$

ahol J a $-M, +M$ közötti egész értékeket veszi fel.

A most leírt – adott szimmetrikus intervallumra vonatkozó – súlyozott számtaniközép-képzést a szakirodalom konvolúciónak nevezi.

A szűrési eljárás jellegét a w_J súlyok felvételének módja szabja meg. Például az ún. futó átlagoláskor

$$w_J = 1/(2M + 1).$$

Lineárisan csökkenő súlyokkal történő ún. simítás esetén

$$w_{-J} = w_{+J} = ((M + 1 - |J|)/(M + 1))^2$$

A bemutatott két súlyfajta értékei $M = 2$ esetére a következők:

- futó átlagolás: 0,2; 0,2; 0,2; 0,2; 0,2.
- simítás lineárisan csökkenő súlyokkal: 0,111; 0,222; 0,333; 0,222; 0,111.

A most bemutatott szűrési eljárás alkalmas kiugró jelértékek kimutatására is. Ebben az esetben az eredeti s_G jeleket azok szűrt s_G értékével vetik össze. Amennyiben az eredeti és a szűrt jel különbsége egy bizonyos \underline{e} határértéket meghalad, azaz

$$|s_G - s_G| > e,$$

akkor a jel – nagy valószínűséggel – kiugrónak tekinthető. Ilyenkor a továbbiakban az eredeti jel helyett a szűrt értéket indokolt felhasználni.

10.1.8. A geostatisztika alapjai

A geostatisztika gyűjtőneve a felhasznált értékek térbeli és esetleg időbeli kapcsolatát felhasználó – mindenekelőtt a földtudományok területén alkalmazott – felületek meghatározására és interpolációs jellegű feladatok megoldására szolgáló eljárásoknak. Az eljárások elméleti alapjait a francia Matheron fektette le. Gyakorlati megvalósítási módjainak kidolgozása a dél-afrikai Krige nevéhez fűződik. (Az eljárásokat ezért a szakirodalomban „krigelés”-nek nevezzük).

A geostatisztika tárgyalásakor induljunk ki abból, hogy bizonyos mennyiségek értéke (például a lehullott csapadék mennyisége) bizonyos pontokban ismert, s ezek alapján akarjuk meghatározni a csapadék lefolyási felületét, s további pontokhoz tartozó – nem mért – értékeket. A diszkrét pontokhoz tartozó

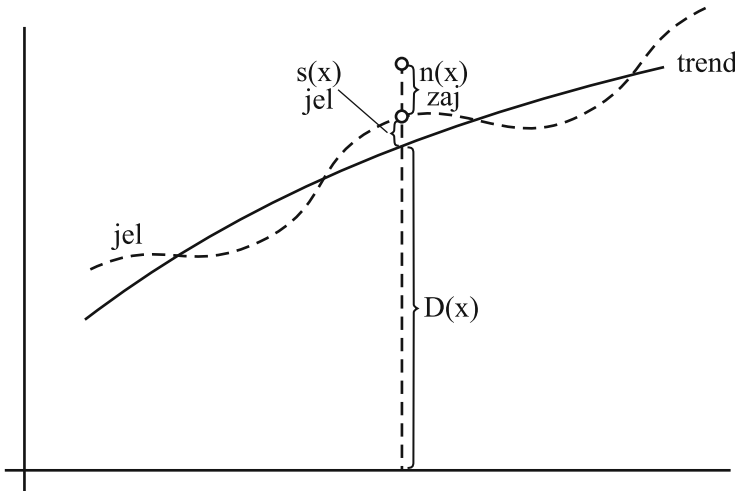
$Z(x)$ értékek (ahol x lehet egy-, két- vagy háromváltozós) a következő három részből tevődnek össze:

- a vizsgált terület egészét jellemző (determinisztikus) függvényből, a trendből: $D(x)$,
- a vizsgált jelenségnek az adott pontban a trendtől való eltérését jellemző jelből: $s(x)$,
- a mért értékeket terhelő – általában a mérésből adódó – hibából, a zajból: $n(x)$.

Ennek megfelelően

$$Z(x) = D(x) + s(x) + n(x).$$

A felbontás elvét a 10.7. ábra mutatja.



10.7. ábra. Trend, jel, zaj

A felbontásban szereplő mennyiségek közül a $D(x)$ trendfüggvény különböző jellegű függvény lehet. Legegyszerűbb esetben $D(x) = 0$. Ha például x kétváltozós, a trendfüggvény lehet

- vízszintes sík: $D(x, y) = k$,
- ferde sík: $D(x, y) = l + mx + ny$.

Természetesen léteznek a bemutatottaknál bonyolultabb trendfüggvények is.

A jelfüggvények vizsgálatakor abból indokolt kiindulni, hogy a szomszédos pontok egymástól nem függetlenek. Elméleti feltevések alapján a jelek ún. sta-

cionárius sztochasztikus folyamatokkal jellemezhető. A jelek kapcsolata lehet:

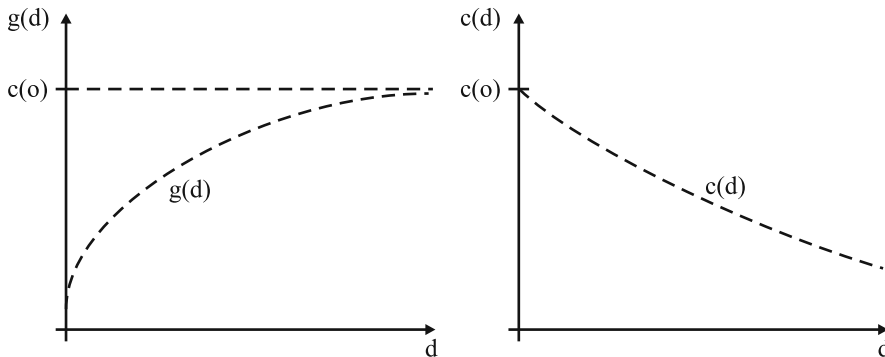
- irányfüggetlen (izotróp),
- irányfüggő (anizotróp).

A jelek függése mértékének jellemzésére a kovarianciafüggvényt, vagy – krigeléskor – a szemivariancia-függvényt használják. Az említett függvények izotróp esetben csupán a pontok d távolságától függenek. Legyen a kovariancia-függvény $c(d)$. A $g(d)$ szemivariancia-függvény ez esetben a következő módon határozható meg:

$$g(d) = c(0) - c(d),$$

ahol $c(0)$ a zérus távolsághoz tartozó variancia.

Egy lehetséges kovariancia- és a szemivarianciafüggvény-párt a 10.8. ábrán tüntetünk fel.



10.8. ábra. Kovariancia- és szemivariancia-függvény

Gyakorlati számításokhoz különböző típusú szemivariancia-függvényeket használnak fel. Például Chang (2004) többek között megkülönböztet Gauss-féle, lineáris, exponenciális szemivariancia-függvényt. Mind a kovariancia-, mind a szemivariancia-függvények alkalmasak a térbeli autokorreláció jellemzésére.

A geostatistika gyakorlati alkalmazásakor különböző „krigelési” módszereket használnak fel. Ezek két legelterjedtebb módja

- a szokásos (ordinary) krigelés,
- az általános (universal) krigelés.

A *szokásos krigelés* esetén a számítás célja ismeretlen pontokban a jelek meghatározása. A számításkor a trendfüggvényt zérusnak tekintjük. Valamely is-

meretlen ponthoz tartozó jelet a szomszédos pontok jeleinek súlyozott számtani közepeként határozzák meg. A súlyokat a pontok közötti tapasztalati szemivariancia-értékek felhasználásával feltételes szélsőértékként számítják. (Feltétel, hogy a súlyok összege 1 legyen).

Az *általános krigelésnél* a számítás célja kettős:

- egyrészt a trendfüggvény ismeretlen együtthatóinak meghatározása,
- másrészt a jelek becslése az ismeretlen pontokban.

A trendfüggvény paramétereit az ismert pontok mért értékeit felhasználva – lényegében a legkisebb négyzetek módszerének alkalmazásával – határozzák meg. Az ismeretlen pontok jeleinek meghatározásakor a szokásos krigelésnél bemutatott elvet követik.

Mind a szokásos, mind az általános krigelés alkalmas a becsült értékek variánciáinak meghatározására. A konkrét számítási összefüggések megtalálhatók például Van Dijk (1999) és Chang (2004) munkáiban.

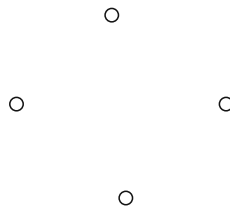
10.2. Gráfelméleti alapok

10.2.1. Alapfogalmak

A gráfelmélet a matematika egyik ága, a hálózatok matematikai elmélete (Britannica Hungarica (1997)). A gráfelmélet eredményeit a térinformatikában széles körűen alkalmazzák

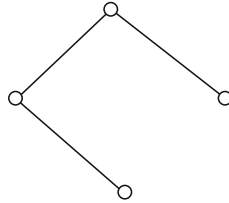
- az adatmodellezéskor (topológiai modellek létrehozása),
- az adatnyerési eljárások minőségellenőrzésekor (konzisztenciavizsgálatok),
- az adatok keresésekor,
- bizonyos elemzési feladatoknál (például a legkedvezőbb útvonal megválasztásakor).

A *gráf* csúcsok és élek halmaza. A gráf *csúcsainak* valamilyen ponthalmaz elemeit nevezzük. A csúcsokat szokás pontnak, szögpontra, vagy csomópontnak is nevezni. A 10.9. ábrán egy lehetséges ponthalmazt tüntettünk fel.



10.9. ábra. Gráf csúcsai

A gráf két csúcsát összekötő vonaldarabot a gráf *élének* nevezzük. A gráfok élei lehetnek egyenes szakaszok vagy ívek (10.10. ábra). A gráf akkor adott, ha ismerjük csúcsait és tudjuk, hogy az élek mely csúcsokat kötnék össze.



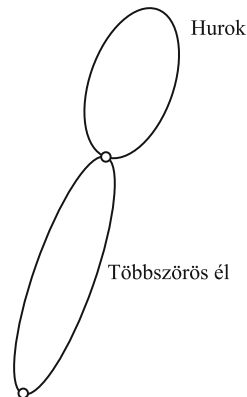
10.10. ábra. Gráf élei

Katona Gy., Recski A., Szabó Cs. (2002) a gráfok következő definícióját adja: Egy *gráf* egy rendezett pár, $G = (V, E)$, ahol V egy nem üres halmaz, E pedig ebből a halmazból képezhető párok halmaza, V elemeit *pontoknak* vagy *csúcsoknak*, E elemeit *éleknek* nevezzük. A pontok és élek mellett a továbbiakban felhasználjuk a *lapok* (*poligonok*) fogalmát is.

A csúcsok és az élek speciális fajtái a következők (10.11. ábra):

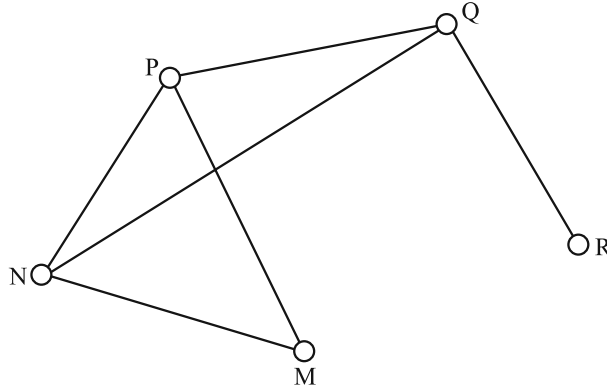
- *Izolált csúcs*: nem indul ki belőle él.
- *Többszörös (párhuzamos) él*: olyan élpár, amely ugyanazon csúcsokat köt össze.
- *Hurok*: olyan él, amelynek kezdő és végpontja egybeesik.

Izolált csúcs



10.11. ábra. Izolált csúcs, többszörös él, hurok

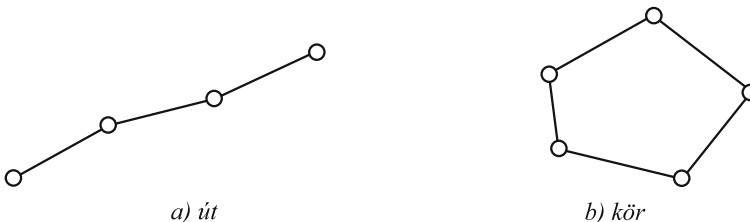
Egyszerű gráf az olyan hurok nélküli gráf, amely nem tartalmaz párhuzamos éleket (10.12. ábra).



10.12. ábra. Egyszerű gráf

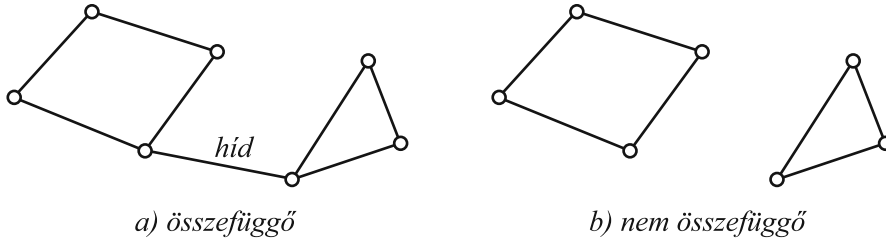
A gráf *rendje* a gráf csúcsainak száma. A gráf rendjét a továbbiakban c betűvel jelöljük. A gráf egy csúcsának a *foka* a csúcsban végződő élek száma. A 10.12. ábrán feltüntetett gráf rendje $c = 5$, az ábrán jelölt B csúcs foka $f = 3$. Egy G gráf *részgráfja* minden olyan gráf, amelyet úgy kapunk, hogy csúcsai és élei közül bizonyosakat elhagyunk.

Vonalnak (*pályának*) nevezünk egy részgráfot, ha egymáshoz csúcsban csatlakozó élek sorozatából áll, és nincs olyan él, amely az élsorozatban kétszer fordul elő. A vonal *hossza* a benne előforduló élek száma. A vonal *nyílt*, ha kezdő és végpontja különböző (különben a vonal *zárt*). Az *út* olyan nyílt vonal, amely nem megy át kétszer egyetlen csúcson sem. A *kör* olyan zárt vonal, amely nem megy át kétszer egyetlen csúcson sem. A 10.13. ábrán utat és kört ábrázolunk.



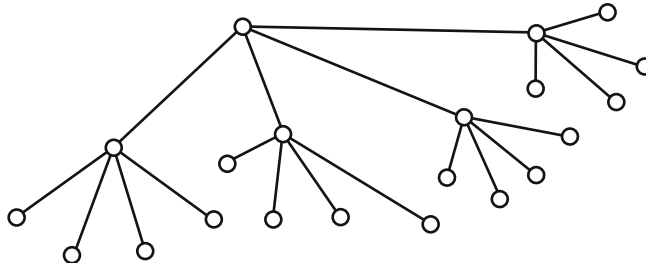
10.13. ábra. Út és kör

A gráf *összefüggő*, ha bármely két csúcsát tartalmazza a gráf egy éle. A nem összefüggő gráf olyan részekből, *komponensekből* áll, amelyeket nem kapcsol össze él. A 10.14. ábra összefüggő és nem összefüggő gráfokat tartalmaz. Ha a nem összefüggő gráf két komponensét összekötik, az összekötéshez felhasznált élt *hídnak* nevezik.



10.14. ábra. Összefüggő és nem összefüggő gráf

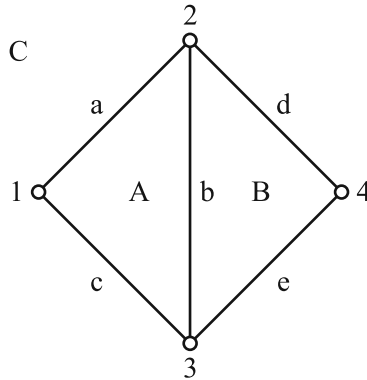
A körmentes összefüggő gráfokat *fának* nevezzük. *Erdőnek (ligetnek)* mondjuk a gráfot, ha komponensei fák. Minden fában van legalább két elsőfokú csúcs. Az n csúcsú fának $n - 1$ éle van. A *négyes fa (quad tree)* olyan fa, amelynek minden nem elsőfokú csúcsából négy él indul ki. Négyes fát tüntetünk fel a 10.15. ábrán.



10.15. ábra. Négyes fa (quad tree)

Az *irányított gráfok* olyan gráfok, amelyekben megjelölik az élek irányát, azaz megadják, melyik csúcsból melyik csúcsba megy az él. Az irányított gráfok megadásához szükséges a kezdőpont és a végpont ismerete.

Síkgráfnak (topológikus gráfnak) nevezzük az olyan gráfot, amely a síkon úgy ábrázolható, hogy élei folytonos görbe ívek legyenek, és két élének csak a végpontjaiban lehessen közös pontja. A *sokszöggráf* olyan összefüggő síkgráf, amely nem tartalmaz hidat. A 10.16. ábrán egy sokszöggráfot tüntetünk fel.



10.16. ábra. Sokszöggráf

A sokszöggráf olyan körei, amelyek belsejében nincs csúcs, *lapokat (poligonokat)* határolnak.

A sokszöggráf csúcsainak c , éleinek e , és lapjainak p száma között az Euler-formula állít fel kapcsolatot (Reimann, 1986):

$$c + p - e = 2$$

(Ha a végtelen tartományt nem számítjuk a lapok közé, akkor

$$c + p - e = 1)$$

A 10.16. ábrán feltüntetett sokszöggráf esetén $c = 4$, $e = 5$, $p = 3$, ebből adódik:

$$4 + 3 - 5 = 2.$$

10.2.2. Illeszkedés és szomszédság

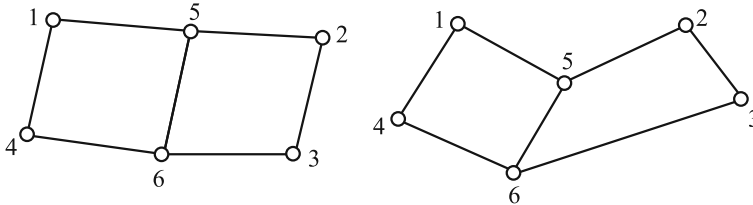
A gráfok jellemzésére használt két alapfogalom

- az illeszkedés és
- a szomszédság.

Az *illeszkedés* a gráfok alkotóelemeire (csúcsok, élek) vonatkozó sajátosság, amely szerint valamely él illeszkedik a kezdő- és végpontjára, illetve fordítva, valamely csúcsból kiinduló élek illeszkednek a csúcsra. Az illeszkedés valamely gráf különböző jellegű elemei közötti kapcsolatot írja le.

Szomszédság áll fenn, ha két csúcsot összeköt egy él, ugyanígy szomszédosak az ugyanazon végpontba futó élek. A szomszédság valamely gráf azonos jellegű elemeire vonatkozik.

Az illeszkedés és a szomszédság a síkgráfok különböző leképezéseire invariánsak. A leképezések során izomorf gráfok jönnek létre. Két gráf *izomorf*, ha csúcsaik között olyan, kölcsönösen egyértelmű megfeleltetés létesíthető, hogy két csúcs az egyik gráfban akkor és csakis akkor van összekötve, ha megfelelőik a másik gráfban is össze vannak kötve. A 10.17. ábrán izomorf gráfokat tüntetünk fel.



10.17. ábra. Izomorf gráfok

A továbbiakban áttekintjük a csomópontok (csúcsok), élek, lapok (poligonok) illeszkedésének és szomszédságának legfontosabb eseteit. Az eseteket a 10.18. ábrán feltüntetett gráf egyes részeinek példáján szemléltetjük.

Illeszkedések

- *csomópont – él* illeszkedés: az a, b, d élek illeszkednek a 2 csomópontra.
- *lap – él* illeszkedés: $a: A-C, b: A-B$.
- *csomópont – lap* illeszkedés:

csomópontra: $(2 - A, B, C)$

lapra $(A - 1, 2, 3)$.

Szomszédságok

- *csomópont-szomszédság*: 1 és 2 csomópont szomszédos az a élen keresztül.
- *élszomszédság*:

csomóponton keresztül: $(3 - c, b, e)$,

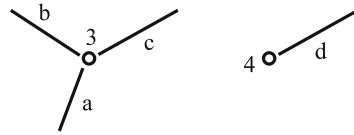
lapon keresztül: $(A - a - b - c)$.

- *lap- (poligon-) szomszédság*: $(A - a - C), (A - b - B)$.

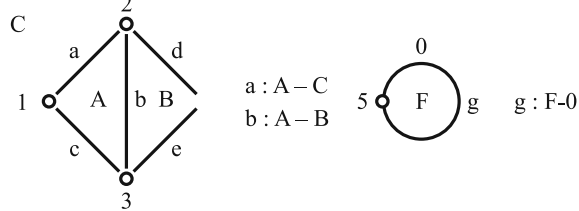
A gráfok illeszkedésének és szomszédságának meghatározásához felhasználhatók a gráfok mátrixrepresentációi. A lehetséges mátrixrepresentációk közül a következőket tárgyaljuk:

- illeszkedési mátrix,
- szomszédsági mátrix.

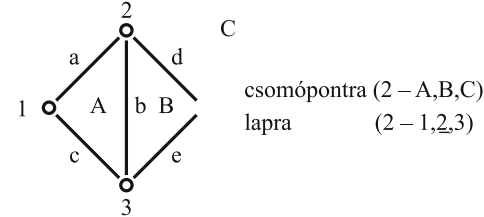
A) csomópont – él illeszkedés



B) lap – él illeszkedés



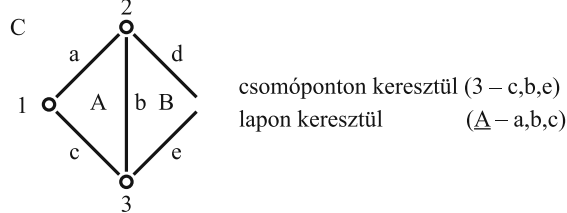
C) csomópont – lap illeszkedés



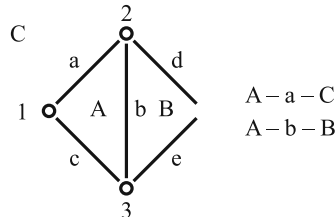
D) csomópont szomszédság



E) él szomszédság



F) lap (poligon) szomszédság



10.18. ábra. Csomópont, él és lap illeszkedése

Legyen egy c csomópontú G gráfnak e éle. A $c \times e$ méretű $\mathbf{B} = (b_{ij})$ illeszkedési mátrix elemeit a következő képen definiálhatjuk:

$$b_{ij} = \begin{cases} 0, & \text{ha a } j\text{-edik él nem illeszkedik az } i\text{-edik pontra,} \\ 1, & \text{ha a } j\text{-edik élnek az } i\text{-edik pont a kezdőpontja,} \\ -1, & \text{ha a } j\text{-edik élnek az } i\text{-edik pont a végpontja.} \end{cases}$$

Ugyanilyen c csomópontú G gráf esetén a $c \times c$ méretű $\mathbf{A} = (a_{ij})$ szomszédsági mátrix elemei a következők:

$$a_{ij} = \begin{cases} i = j & \text{esetben az } i\text{-edik csomópontban található élek száma,} \\ -1, & \text{ha az } i\text{-edik pontra és a } j\text{-edik pontra azonos él illeszkedik,} \\ 0, & \text{ha az } i\text{-edik pont és a } j\text{-edik pont nem szomszédos.} \end{cases}$$

Az illeszkedési mátrix és a szomszédsági mátrix között fennáll a következő összefüggés:

$$\mathbf{A} = \mathbf{B} \cdot \mathbf{B}$$

Példa. A 10.18. ábrán feltüntetett gráf esetében írjuk fel a topológiai leírás-hoz szükséges táblázatokat, s azok alapján határozzuk meg az illeszkedési és a szomszédsági mátrixot!

Poligonok táblázata

Poligon	Él
A	a, b, c
B	b, d, e
C	a, c, d, e

Csomópontok táblázata

Csomópont	Él
1	a, c
2	a, b, d
3	c, b, e
4	d, e

Élek táblázata

Él	Kezdőpont	Végpont	Bal poligon	Jobb poligon
a	1	2	C	A
b	2	3	B	A
c	1	3	A	C
d	2	4	C	B
e	3	4	B	C

Koordináták táblázata

Csomópont	X	Y
1	x1	y1
2	x2	y2
3	x3	y3
4	x4	y4

Az illeszkedési mátrix a következő:

$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

A szomszédsági mátrix pedig a következő:

$$\mathbf{A} = \mathbf{B} * \mathbf{B} = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}$$

Az illeszkedési és szomszédsági mátrixok mellett a gráfok egyéb reprezentációi is léteznek (Katona, Recski, Szabó, (2002)).

A gráfok felhasználásával történő adatmodellezés alkalmas a topológiai kapcsolatok leírására. A gráfok alkalmazásából bizonyos konzisztenciafeltételek következnek, amelyek jól használhatók az elkészült állományok ellenőrzésekor. A konzisztenciafeltételeket részletesen tárgyalja Bill, Fritsch (1991) munkája. Példaként a következőket említjük:

- Általános – hurkot nem tartalmazó – hálózatban négy konzisztenciaössze-függés adható meg:
 - valamely csomópontnak illeszkednie kell az összes benne végződő élre,
 - valamely élnek illeszkednie kell a végpontjaira,
 - valamely csomópontnak szomszédosnak kell lennie a vele éllel összekötött csomópontokkal,
 - valamely élnek szomszédosnak kell lennie az ugyanazon csomópontban végződő élekkel.
- Konzisztenciafeltételnek tekinthető a korábban tárgyalt Euler-formula is.

10.2.3. Legkedvezőbb utat kereső algoritmusok

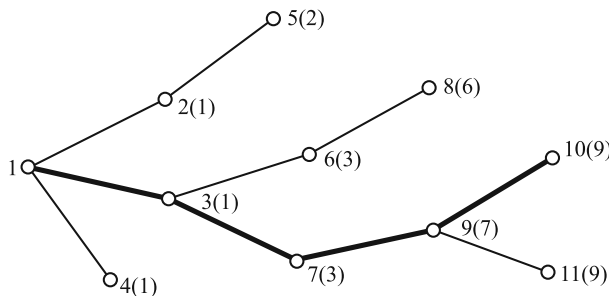
A legkedvezőbb út keresése a térinformatikai elemzések klasszikus feladata. A feladat különböző feltételezések esetén egyaránt megoldható. Ilyen feltételezések lehetnek:

- a gráfok nem irányítottak, az élek azonos hosszúságúak,
- a gráfok nem irányítottak, az élek különböző hosszúságúak,
- a gráfok irányítottak,
- az útvonalon akadályok jelentkeznek,
- figyelembe vesszük a forgalom alakulását.

A felsorolt feladatok közül az elsőnek a megoldására ismertetünk vázlatosan Katona, Recski, Szabó (2002) alapján egy algoritmust. Adott egy összefüggő $G = (V, E)$ gráf és annak egy kitüntetett $s \in V$ pontja. Keressük a legrövidebb utat s -ből egy másik kiválasztott t pontba (vagy akár minden más pontba).

Azonos hosszúságú élek esetén az út hossza a szükséges élek száma. Ekkor s összes szomszédja egységnyi távolságra van s -től, ezen szomszédok minden olyan szomszédja, ami maga nem szomszédja s -nek, kétegységnyi távolságra van s -től, és így tovább.

A feladat megoldásához az ún. szélességi keresést használjuk fel. A szélességi keresés elvét Katona, Recski, Szabó (2002) alapján egy a 10.19. ábrán feltüntetett példán szemléltetjük. Induljunk ki a bejárás során az $s = 1$ *jelű* pontból. Először járjuk be összes szomszédját (2,3,4). Utána járjuk be az első szomszédjának (2) összes olyan szomszédját, ahol még nem jártunk (5), majd második szomszédjának (3) összes szomszédját (6,7) és így tovább. Ha már bejártuk az 1 összes szomszédjának összes szomszédját, akkor menjünk vissza abba a pontba, ahol legrégebben jártunk azok közül, amelyeknek még nem néztük végig a szomszédjait (6). Ezt az eljárást addig folytatjuk, amíg tudjuk.



10.19. ábra. Legrövidebb út meghatározása szélességi kereséssel

Amikor sorban bejárjuk a gráf pontjait, akkor tulajdonképpen sorszámot adunk a pontoknak abban a sorrendben, ahogy bejárjuk azokat. Legyen az i -edik pont jelölése v_i . Az i értékek $i = 1$ -gyel kezdődnek, majd minden további pontnál 1-gyel növekednek. Azért, hogy a bejárás végén felrajzolhassuk a bejárt fát, meg kell jegyeznünk, hogy v_x -re melyik pontból léptünk. Ezt a pontot jelölje $q(v_x)$. A bemutatott példán a 10.21. ábrán a $q(v_x)$ értékeket is feltüntettük. Például a 7(3) érték azt mutatja, hogy a 7 jelű pontba a 3 jelű pontból jutottunk el. A gráf pontszáma legyen v .

A legrövidebb útvonal keresésekor a szélességi keresést az $a = s$ választással végezzük el. Amikor az algoritmus véget ér, a kiválasztott t ponthoz határozzuk meg a $q(t) = u_1$ pontot. Ha $u_1 \neq s$, akkor a $q(u_1) = u_2$ pontot, ha $u_2 \neq s$, akkor a $q(u_2) = u_3$ pontot stb. Előbb-utóbb eljutunk egy olyan k indexhez, amelyre $u_k = s$. Belátható, hogy az

$$(s = u_k, u_{k-1}, u_{k-2}, \dots, u_1, t)$$

a legrövidebb út s -ből t -be. A bemutatott példán a legrövidebb út a következő lesz:

$$(1, 3, 7, 9, 10)$$

A legrövidebb utat a 10.19. ábrán vastag vonallal tüntettük fel.

Ha a gráf irányított lett volna, akkor ugyanez az algoritmus az s -ből t -be vezető irányított utat határozta volna meg.

A pont elején felsorolt feladatok közül a különböző élhosszúságú gráfokra ad algoritmust Dijkstra (1959). Egy hálózaton belül az összes pont pár távolságának meghatározására szolgál Floyd algoritmus. Az említett algoritmusok leírása megtalálható Rónyai et al. (1999) és Katona Gy., Recski A., Szabó Cs. (2002) munkájában. Ha a hálózatokon belül akadályok (például útlezárások) fordulnak elő, akkor speciális algoritmusok felhasználása szükséges, ilyenekre találunk példát Schmed (2001) munkájában.

10.3. Logikai alapok

10.3.1. A matematikai logika fogalma, logikai műveletek

A logika a gondolkodás formáinak és törvényeinek, a valóság elkülöníthető létezői közötti viszony, illetve mozgás általános törvényeinek tudománya. A modern vagy szimbolikus logika a matematikai módszereket és az azokkal nyert eredményeket alkalmazza (Akadémiai kislexikon, 1999). A logikai műveletek olyan műveletek, amelyek logikai ítéleteken vannak értelmezve, s valamely kapott logikai ítélet értéke csak azon ítéletek logikai értékétől függ, amelyekre a műveletet alkalmaztuk (Farkas, 1972).

A logikai műveletek három egymástól jól elkülöníthető csoportjával foglalkozunk:

- Egyenlőségek, egyenlőtlenségek,
- Műveletek élesen elhatárolt (crisp) halmazokkal,
- Műveletek nem élesen elhatárolt (fuzzy) halmazokkal.

A logikai műveleteket a térinformatikai elemzések különböző feladatainál – a lekérdezésektől a modellezésig – széles körűen alkalmazzák. Egy részük alkalmas bizonytalanságok kezelésére is.

10.3.2. Egyenlőségek, egyenlőtlenségek

Az egyenlőségeket és egyenlőtlenségeket elsősorban a térbeli elemzések rendezési feladatainak megoldásakor alkalmazzuk. Ilyen rendezés lehet például egy adatsorban a

- szélsőérték-keresés (például valamely utcában melyik a legdrágább telek),
- bizonyos értéknél nagyobb (vagy kisebb) érték keresése (például valamely utcában mely telkek ára kisebb 10 millió forintnál).

A rendezési feladatok egy része az SQL kérdezőnyelv alkalmazásával közvetlenül megoldható. Más esetben speciális algoritmusok alkalmazása is indokolt lehet. Speciális algoritmusokon alapuló eljárást mutatunk be Rónyai et al. (1999) alapján.

Legyen Z az egész számok halmaza. Tűzzük ki feladatul a számok nagyság szerinti rendezését. A feladatot úgy is megfogalmazhatjuk, hogy a

$$z_1, z_2, \dots, z_n$$

sorozatot a nem csökkenő

$$(v_1, v_2, \dots, v_n)$$

$$v_1 \leq v_2 \leq \dots \leq v_n$$

sorozatba rendezzük át. A feladat különböző algoritmusokkal oldható meg. Induljunk ki a Z halmaz

$$S = \{s_1 < s_2 < \dots \leq s_n\}$$

részalmazából. Legyen $s \in Z$ azaz s a Z egy tetszőleges eleme.

Kérdés $s \in S$? És hol a helye?

A feladatot megoldhatjuk *lineáris* kereséssel. Ekkor s -t először s_1 -gyel hasonlítjuk össze.

Három eset lehetséges:

- Ha $s < s_1$, akkor végeztünk, mivel s nincs S -ben.
- Ha $s = s_1$, akkor sikerrel jártunk, mivel $s \in S$ és a helyét is ismerjük.
- Ha $s > s_1$, akkor a keresést folytatnunk kell.

A *bináris* keresés esetén s_1 értékét először az S középső elemével hasonlítjuk össze, s ennek alapján döntünk arról, hogy melyik részhalmazban lehet s_1 , és ennek megfelelően folytatjuk a keresést.

10.3.3. Műveletek élesen határolt (crisp) halmazokkal

A tárgyalás során a halmazt alapfogalomnak tekintjük. A halmaz bizonyos tulajdonságokkal rendelkező egyedek (objektumok) sokasága. Ha az a objektum eleme az A halmaznak, ezt a tényt az

$$a \in A$$

szimbólummal írhatjuk le.

Példának tekinthetjük egy utcához tartozó telkek vagy egy város lakóinak halmazát. A felsorolt példák élesen elhatárolt (crisp) halmazokra vonatkoznak, hiszen egy telekről egyértelműen eldönthető, hogy valamely utcához tartozik vagy sem, s ugyanez igaz valamely város lakói esetében is.

A halmazt alkotó egyedek a halmaz elemei. A halmazokkal kapcsolatos speciális fogalmak a következők:

- üres halmaz: olyan halmaz, amelynek nincs egyetlen eleme sem.
- részhalmaz: ha az A halmaz minden eleme a B halmaznak is eleme, akkor A a B részhalmaza.

A halmazokat megadhatjuk

- elemeikkel, például: $A = \{0, 2, 7\}$.
- tulajdonságaikkal. Ha $P(x)$ olyan állítás, amely x értékétől függően lehet igaz vagy hamis, akkor $\{x|P(x)\}$ az a halmaz, amelynek elemei azok az x -ek, amelyekre $P(x)$ igaz.

Az első megadási módra példa lehet, ha felsoroljuk egy utcához tartozó telkek ún. helyrajzi számait. A második megadási móddal azon telkek halmazát definiálhatjuk, amelyek nagyobbak mint 800 m^2 .

A halmazokkal különböző műveletek végezhetőek. Ezek közül a következőket tárgyaljuk:

- metszet,
- egyesítés,

- különbség,
- diszkrepancia.

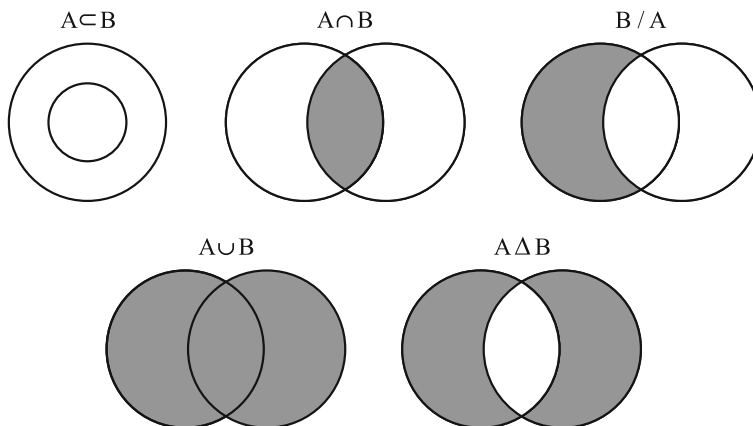
Az A és B halmazok *metszete* az a C halmaz, amelynek elemei mind A -hoz, mind B -hez hozzátartoznak. Jelölése $A \cap B$. Ha például egy utcában az A halmaz elemei azok a telkek, amelyek területe nagyobb mint 800 m^2 , a B halmaz elemei pedig azok a telkek, amelyek ára 10 millió forintnál kisebb, akkor a metszet jelöli ki a 800 m^2 -nél nagyobb, de 10 millió forintnál olcsóbb telkek halmazát.

Az A és B halmazok *egyesítése (uniója)* az a C halmaz, amelynek elemei legalább A -hoz vagy B -hez hozzátartoznak. Jelölése: $A \cup B$. Az előző példát folytatva az egyesítés segítségével találhatjuk meg azokat a telkeket, amelyek vagy 800 m^2 -nél nagyobbak vagy 10 millió forintnál olcsóbbak.

Az A és B halmazok *különbsége* az a C halmaz, amely tartalmazza A mindazon elemeit, amelyek nem elemei B -nek. Jelölése: $A \setminus B$. A példa esetén a különbség szolgáltatja azon 800 m^2 -nél nagyobb telkeket, amelyek 10 millió forintnál nem olcsóbbak.

Az A és B halmazok *diszkrepanciája* az a C halmaz, amely tartalmazza A és B valamennyi elemét metszetükön kívül. Jelölése: $A \Delta B$. Példánkban a diszkrepancia műveletével kizárhatjuk azon 200 m^2 -nél nagyobb telkeket, amelyek 10 millió forintnál nem olcsóbbak, továbbá azon 10 millió forintnál olcsóbb telkeket, amelyek 800 m^2 -nél nem nagyobbak.

A felsorolt műveleteket szokás az ún. Venn-diagramokkal szemléltetni (10.20. ábra). A diagramokon a bevonalkázott rész mutatja azt a területet, amelyre az állítás igaz. A műveletekre igazak a Boole-algebra axiómái.



10.20. ábra. Halmazműveletek szemléltetése Venn-diagrammal

A bemutatott logikai műveletek különösen jól használhatók komplex kapcsolatok esetén. Sokszor a bemutatott példánál jóval több témára használják párhuzamosan a különböző műveleteket.

10.3.4. Műveletek nem élesen határolt (fuzzy) halmazokkal

Az eddigiekben feltételeztük, hogy a felhasznált halmazok határvonala éles, azaz egyértelműen eldönthető, hogy valamely elem a halmazhoz tartozik vagy sem. A valóságban gyakran fordulnak elő olyan halmazok, amelyek határvonala nem tekinthető élesnek.

Példaként említhetjük egy erdő és egy rét közötti átmenetet, vagy előző példánkban a 10 millió forintos határt. Az erdő és a rét közötti határvonal csak bizonytalanul jelölhető ki. A telket vásárolni szándékozó személy gondolataiban sem feltétlenül éles a 10 millió forintos határ, hanem feltehetően 8-10 millió (esetleg 9-11 millió) forintra gondol.

A nem élesen elhatárolt halmazokat fuzzy-halmazoknak nevezzük (például Borgulya 1999, Kóczy, Tikk 2000). A nem élesen elhatárolt halmazokat az elemeik mellett az elemekhez tartozó tagsági értékkel jellemzik. A tagsági érték 0 és 1 között mozoghat, s különböző tagsági függvények segítségével adható meg. (Az élesen határolt halmazok esetén a tagsági érték 0 vagy 1 lehet, a 0 azt jelenti, hogy valamely elem biztosan nem tartozik a halmazba, az 1 pedig, hogy biztosan beletartozik).

Egy nem éles határú A fuzzy-halmaz megadásához felhasználjuk a d elemeket, s az azokhoz tartozó $h(d)$ tagsági értékeket:

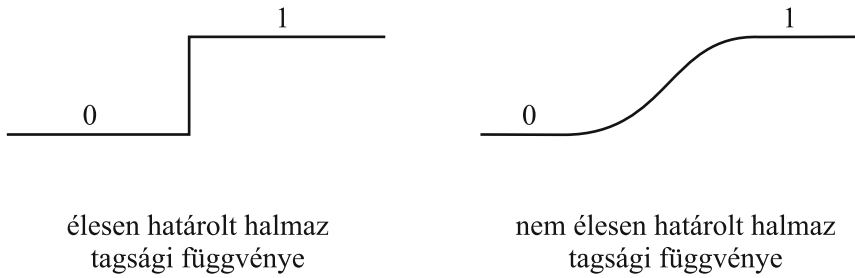
$$A = \{(d; h) | d \in D, h \in H\},$$

ahol A a $\{(d; h_A(d))\}$ értékpárok halmaza, $h_A(d)$ a tagsági függvény, amelynek értéke 0 és 1 között változik.

A tagsági függvényeknek különböző fajtái léteznek. Néhány tagsági függvényt Bartelme (2005) alapján a 10.21. ábrán mutatunk be. Az első bemutatott függvény az élesen határolt (crisp) halmazokat jellemző tagsági függvény.

A térinformatikában használt különböző jellegű tagsági függvényekről és felhasználási területükről hasznos összeállítást közöl Raines et al. (2011).

A fuzzy-halmaz és tagsági függvénye között egy-egy értelmű megfeleltetés létezik. Ebből következik, hogy két fuzzy-halmaz akkor azonos, ha tagsági függvényük azonos. A fuzzy-halmazokkal ugyanolyan műveletek végezhetők, mint az élesen határolt halmazokkal. Azonban e műveletek többféleképpen definiálhatók. A leggyakoribb műveletek egy lehetséges definíciója Borgulya (1999) és Kóczy, Tikk (2000) alapján a következő. Legyen A, B két fuzzy-halmaz az X alaphalmaz felett $d_A(x)$ és $d_B(x)$ tagsági függvényekkel.



10.21. ábra. Élesen határolt és fuzzy halmazok tagsági függvényei

- Két halmaz metszete (minimum-operátor)

$$d_{A \cap B}(x) = \min[d_A(x), d_B(x)],$$

- Két halmaz egyesítése (maximum-operátor)

$$d_{A \cup B}(x) = \max[d_A(x), d_B(x)],$$

A két művelet további definíciói (például T -norma, S -norma), valamint egyéb műveletek definíciói megtalálhatók az előbb idézett művekben. A most leírt műveleteket Raines et al. (2010) Fuzzy AND és Fuzzy OR elnevezéssel látja el.

Irodalomjegyzék

- Akadémiai kislexikon L-Z* (1990), Akadémiai Kiadó, Budapest, pp. 1-978.
- Bartelme, N. (2005): *Geoinformatik, Modelle – Strukturen – Funktionen*, Springer-Verlag, Berlin, Heidelberg, pp. 1-454.
- Bill, R (1999): *Grundlagen der Geo-Information-Systeme*, Band 2. Wichmann Verlag, Heidelberg, pp. 1-475.
- Bill, R., Fritsch, D. (1991): *Grundlage der Geo-Information-Systeme*, Wichmann Verlag, Heidelberg.
- Borgulya, I. (1999): *Neurális hálók és fuzzy-rendszerek*, Dialóg Campus Kiadó, Budapest-Pécs, pp. 1-226.
- Britannica Hungarica* 7. kötet (1997), Magyar Világ Kiadó, Budapest. pp. 1-782.
- Chang, K.(2004): *Introduction to Geographic Information Systems*, McGraw Hill, Boston, etc. pp. 1-400.
- Detrekői Á. (1991): *Kiegyenlítő számítások*, Tankönyvkiadó, Budapest, pp. 1-685.

- Detrekői Á., Szabó Gy. (2002): *Térinformatika*, Nemzeti Tankönyvkiadó Rt., Budapest pp. 1-380.
- Dijkstra, E. W. (1959): A Note on Two Problems in Connexion with Graphs, *Numerische Mathematik 1*. 269-271.
- Farkas M. (1972): *Matematikai kislexikon*, Műszaki Könyvkiadó, Budapest, pp. 1-444.
- Horvai, Gy, (szerk.) (2001): *Sokváltozós adatelemzés (kemometria)*, Nemzeti Tankönyvkiadó, Budapest, pp. 1-356.
- Katona Gy. Y., Recski A, Szabó Cs. (2002): *A számítástudomány alapjai*, Typotex, Budapest, pp. 1-190.
- Korn, G. A., Korn, T. M. (1975): *Matematikai kézikönyv műszakiaknak*, Műszaki Könyvkiadó, Budapest, pp. 1-995.
- Longley, P. A. et al. (2011): *Geographic Information System&Science*, John Wiley&Sons. pp. 1-539.
- Raines, G.L. et al. (2011): New fuzzy logic tools in ArcGIS 10, *ArcUser, Spring 2010*, pp. 8-12.
- Reiman I. (1986): *A geometria és határterületei*, Gondolat, Budapest. pp. 1-419.
- Rónyai L. et al. (1999): *Algoritmusok*, Typotex, Budapest, pp. 1-349.
- Schenk, T. (1999): *Digital Photogrammetry*, Vol. I., TerraScience, Laurelville, OH. pp. 1-428.
- Schmied, W. (2001): *Berechnung kürzeste Wege in Strassennetzen mit Weg-everboten*, DGK, Reihe C Nr. 538, Verlag der Bayerischen Akademie der Wissenschaften München.
- Természettudományi lexikon*, 4. kötet, Akadémiai Kiadó. Budapest, pp. 1-736.
- Van Dijk, M.J., et al. (1999): Geostatistics. In: Bähr, H.-P., Vögtle, Th. (eds): *GIS for Enviromental Monitoring*, E. Schwizerbarsche Verlagsbuchhandlung, Stuttgart pp. 174-182.