FOGALMAK ÉS MÓDSZEREK

Meszéna György

Valószínűségeloszlások és idősorok felbontása

Az adatokon nyugvó elemző tevékenység, a közgazdász munkájának mindenkor visszatérő, központi jelentőségű részét alkotja. A gazdasági valóság belső összefüggéseinek feltárása gyakran igen sok tényező együttes hatásának vizsgálatát, az összetevők számának, súlyának megállapítását, kívánja meg. Az adatgyűjtés közismert technikai és gazdasági nehézségei, a sokféle véletlen hatás megismerése mintavételi eljárások alkalmazására, reprezentatív megfigyelésekre vezet. Ha az éppen vizsgált alapsokaság valószínűségeloszlásában különböző tényezők hatása keveredik, a mintából szerzett információk ezt tükrözni fogják. A mélyreható elemző munkához ilyenkor szolgálnak jó segítségül a valószínűségeloszlások felbontására kidolgozott, hatékony módszerek. Lehetőség nyílik az egyes összetevők jellegének, paramétereinek meghatározására, ezáltal terveinkbe külön-külön is beépíthetők s figyelembevehetők lesznek.

A legkülönbözőbb tartalmú és szintű ipari és mezőgazdasági termelési, beruházási, bel- és külkereskedelmi adatok, indexek illetve mutatók, stb. időbeni változásait a matematika speciális sztohasztikus folyamatok, idősorok alakjában tárgyalja. Az egyes társadalmi, gazdasági, sőt természeti törvényszerűségek hatása, a véletlen ingadozásokkal együtt, szuperponálódva jelenhet meg megfigyelt alapadatainkban. A jelenségek alapos megismerése, reális közgazdasági tartalommal rendelkező magyarázata, nem képzelhető el, az együttesen ható összetevő tényezők szétválasztása nélkül.

Egy eredő hatás felbontása, — bármely területen is jelentkezik, — általában nem könnyű, vagy legalábbis, minden további nélkül, nem egyértelmű feladat. Ezek a nehézségek indokolják egyrészt a terület távolról sem lezárt voltát, (napjainkban is széleskörű kutatás tárgyát képezi), másrészt azt a tényt is, hogy a meglevő eredményeknek is csak kisebb részét tekinthetjük közismertnek. Ugyanakkor társadalmi-gazdasági törvényszerűségeink jelentős része is sztochasztikus hatásokat tartalmazó, tendenciaszerűen ható, összetett jelenség, s így alaposan, csak az összetevők birtokában ismerhetjük meg. Ez a magyarázata a felbontás-problémakör egyre mélyülő tanulmányozásának.

Az alábbiakban nem törekszünk teljességre. Célunk az, hogy a problémát előtérbe állítsuk, felhívjuk a figyelmet egyes módszerekre és közgazdasági alkalmazási lehetőségeikre. A matematikai apparátus részleteit illetően, esetenkint a vonatkozó szakirodalomra utalunk.

1. Összetett valószínűségeloszlások vizsgálata

Nagy vonalakban áttekintve a valószínűségeloszlások kialakulásának történetét, abban *három szakaszt* különböztethetünk meg. Kezdetben a „harang alakú” Gauss-görbével mint sűrűségfüggvénnyel rendelkező normális eloszlás, szinte egyeduralkodónak volt tekinthető. Ha egy tapasztalati adathalmazban a szimmetricitás követelménye csorbát szenvedett, az különböző mértékszámokkal számszerűsíthető volt, s kifejezte a ,,normális eloszlástól való eltérés'' mértékét. Később sikerült a legkülönbözőbb igényeket kielégítő asszimmetrikus eloszlások konstrukciója, így nagymértékben bővült az egyes esetekben alkalmazható hipotézisek választéka is. Napjainkban viszont már az összetett eloszlások előállítása, felbontása az aktuális problémakör.

Empirikus eloszlásfüggvény Bruns-sorba fejtése

Első lépésként egy sajátos, régebbi eljárás rövid ismertetésével szeretnénk foglalkozni. Gyakorlati kivitele formálisan igen hasonlít más, később kialakult, s elvileg eltérő módszerekhez.

Ha egy tapasztalati eloszlás nem, vagy csak közelítőleg nevezhető normálisnak, a ,,kumulált'' relatív gyakoriságok lépcsős függvényét (empirikus eloszlásfüggvény), a normális eloszlás eloszlásfüggvénye nem fogja jól közelíteni. A lépcsős függvényt $V(x)$-el jelölve, az valamely $N(x)$ alapfüggvény szerint a következőképpen fejthető sorba:

$$V(x) - N(x) = c_0 N'(x) + c_1 N''(x) + \dots$$

A c_k együtthatókra megfelelő képletek állnak rendelkezésre, alapfüggvényül viszont a standard normális eloszlás eloszlásfüggvényét fogjuk választani. A deriválás során szorzó tényezőként a Hermite-polinomoktól csak kevéssé eltérő polinom kifejezéseket kapunk, ezek az egész eljárással együtt, praktikusan, táblázatosan elrendezhetők, számolhatók.

A sorbafejtést tetszőleges lépésig folytathatjuk, s az így adódó azonos típusú ,,komponensek'' segítségével az igényeknek megfelelő legjobb közelítést előállíthatjuk.

1. Példa:

Az alábbi példa uránlelőhelyek keresésének egy gyakorlati eljárása keretében készült, természetes vizekben található uránnyomok statisztikai értékelése során [11], [12]. A közgazdasági vonatkozások nyilván nem igényelnek bővebb magyarázatot. (A szivárgó víz a kőzetek tulajdonságaitól — pl. szemcsenagyság— függően, oldja a talajban levő uránt, s így a geológiai viszonyok ismeretében a vízminták speciális, a további következtetésekhez jól felhasználható térképek készítésére alkalmasak.)

A feldolgozásban a Hegyalja eruptív kőzeteiből fakadó forrás- és kútvizei képezik az alapsokaságot, ezekből történt a mintavétel, s az urán-koncentráció alkalmas egységekben való meghatározása, az empirikus eloszlás előállításával együtt.

A hét tagig felírt Bruns-sor a következő volt:

$$B(x) = \Phi(x) + \frac{1}{\sqrt{\pi}} e^{-x^2} [-0{,}01787\, p_2(x) - 0{,}002479\, p_3(x) +$$

$$+ 0{,}000789\, p_4(x) - 0{,}0000367\, p_5(x) + 0{,}00005008\, p_6(x) + 0{,}0000\,138\, p_7(x)].$$

A szereplő „Hermite" polinomok:

$$p_2(x) = 4\,x^2 - 2$$

$$p_3(x) = 8\,x^3 + 12\,x$$

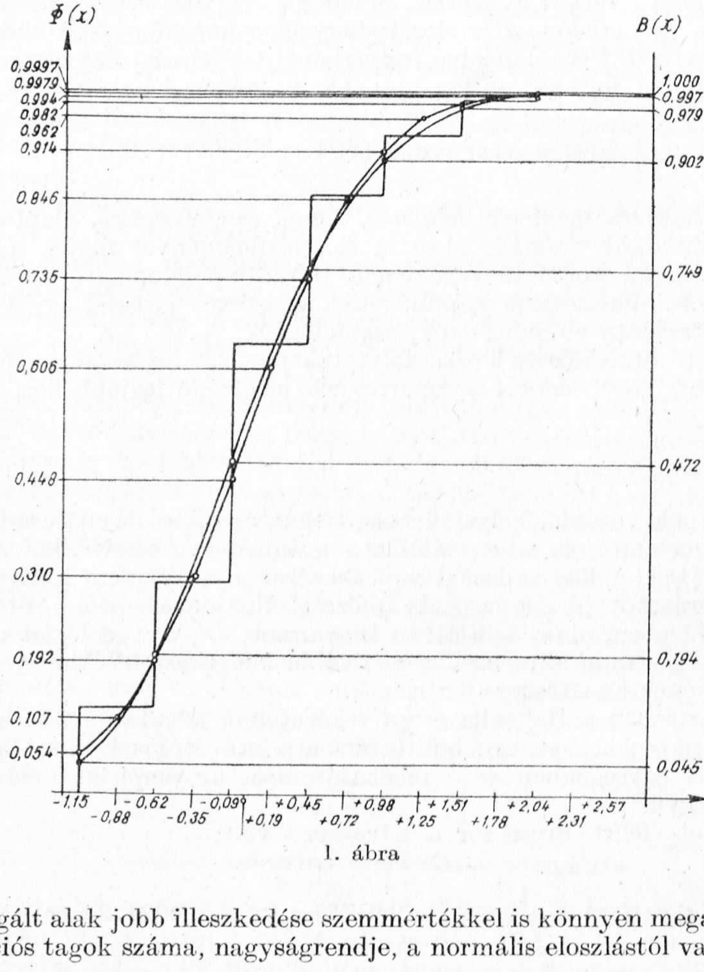
$$p_4(x) = 16\,x^4 - 48\,x^2 + 12$$

$$p_5(x) = -\,32\,x^5 + 160\,x^3 - 120\,x$$

$$p_6(x) = 64\,x^6 - 480\,x^4 + 720\,x^2 - 120$$

$$p_7(x) = -\,128\,x^7 + 1344\,x^5 + 1680\,x$$

Az empirikus eloszlás, a Gauss-görbe és a Bruns-sorral korrigált alak ábráit egy rendszerben felrajzolva (a szükséges transzformációk elvégzésével) a következő képet kapjuk:



1. ábra

A korrigált alak jobb illeszkedése szemmértékkel is könnyen megállapítható. A korrekciós tagok száma, nagyságrendje, a normális eloszlástól való eltérésre jellemző.

„Gauss analízis”

Ha $f(x, a, b)$ egy kétparaméteres, egycsúcsú sűrűségfüggvény, akkor a

$$k(x) = \sum_{k=1}^{N} p_k \cdot f(x; a_k, b_k) \qquad (p_k > 0)$$

összefüggés által meghatározott $k(x)$ függvényt, az $f(x;\ a_k,\ b_k)$ komponensek p_k súlyokkal képezett szuperpozíciójának nevezzük [19].

Több helyi maximummal rendelkező „hullámos” empirikus sűrűségfüggvények elemzésénél szoktunk találkozni ezzel a problémával. Ha az alapsokaság ismeretében feltételezhető, hogy az eredő több normális eloszlás keverékeként jött létre, érdekes és fontos feladatként vetődik fel a komponens-eloszlások, s azok paramétereinek meghatározása.

Ekkor tehát a keverék alakja:

$$f(x) = \sum_{k=1}^{N} A_k \cdot \frac{\exp\left(-\dfrac{(x-m_k)^2}{2\,\sigma_k^2}\right)}{\sqrt{2\pi}\,\sigma_k}$$

s feladatunk $f(x)$ ismeretében az $A_k;\ m_k;\ \sigma_k$ paraméterek értékének megállapítása lesz (7).

E területen kiemelkedő, úttörő munkát, MEDGYESSY PÁL végzett [1—9] [18]. Az igen sokrétűen felvetődő, és nagy apparátussal tárgyalható anyag részleteit illetően, ismételten a megadott szakirodalomra utalunk, az alábbiakban néhány példát szeretnénk előnyben részesíteni.

2. Példa:

A (7) közleményben található az alábbi érdekes alkalmazás. Az előzőekben leírt, sok véletlen hatással is zavart, összetett spektrumot mértek egy berendezéssel. Az eszköz pontossága sem volt a legmegfelelőbb. Ilyen esetekben új problémaként jelentkezik az egyes komponensek egzisztenciája, következésképpen az összetevők száma is. A vizsgálatot Gauss-analízissel végezték, így válaszolni tudtak a vitatott kérdésekre. Később lehetőség nyílt nagyobb pontosságú, tökéletesebb mérőberendezés alkalmazására. s ez igen szépen igazolta a számítási eredményeket. „A matematikai eljárás mintegy pótolta a nagyobb felbontóképességű készüléket.”

3. példa:[1]

Egy gyárban 903 munkás dolgozik az alábbi besorolásban:
takarító személyzet
segédmunkások
betanított munkások
szakmunkások

Az egyik bérfizetés alkalmával készült az alábbi összeállítás.

[1] BÉKÉSI GÁBOR tud. diákköri dolgozatából

Havi kereset Ft	Munkások száma	Havi kereset Ft	Munkások sz.
576— 625	2	1376—1425	106
626— 675	3	1426—1475	109
676— 725	4	1476—1525	102
726— 775	8	1526—1575	92
776— 825	16	1576—1625	91
826— 875	34	1626—1675	91
876— 925	48	1676—1725	83
926— 975	61	1726—1775	68
976—1025	88	1776—1825	54
1026—1075	105	1826—1875	42
1076—1125	99	1876—1925	32
1126—1175	85	1926—1975	23
1176—1225	74	1976—2025	18
1226—1275	71	2026—2075	14
1276—1325	75	2076—2125	10
1326—1375	92	2126—2175	7
		2176—2225	5

Feladatunk ennek a táblázatnak az ismeretében megállapítani az egyes besorolási kategóriák átlag fizetéseit.

Felhasználjuk még a következőket:

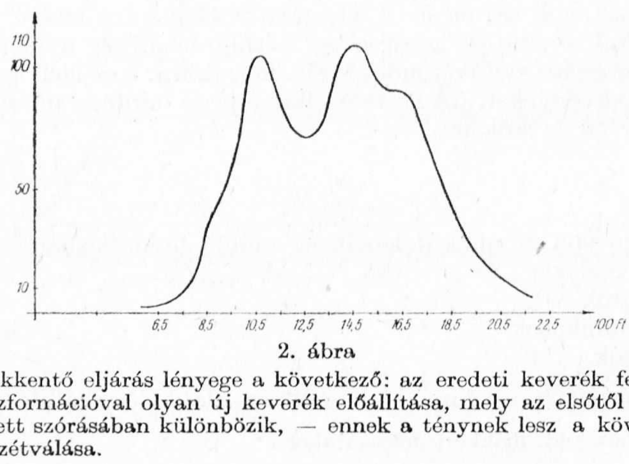
a) A négy besorolás fizetési kategorizálást is jelent. (Tehát a kategóriák átlagfizetései nem egyezhetnek meg.)

b) Az egyes kategóriákon belül a munkások fizetése normális eloszlást követ.

Amint az a mellékelt ábrákból is megállapítható, az eljárás szóráscsökkentéssel[2] dolgozik, az alapgörbéből kiindulva, az egyes lépések után előálló, egyre „hegyesebb tűk" jelzik a kopmonensek elhelyezkedését. A példában két lépés végigszámolása történt meg, (munkaigénye logarléc alkalmazásával kb. 2^{h}). A folytonos görbék a szemléletesség céljából szerepelnek.

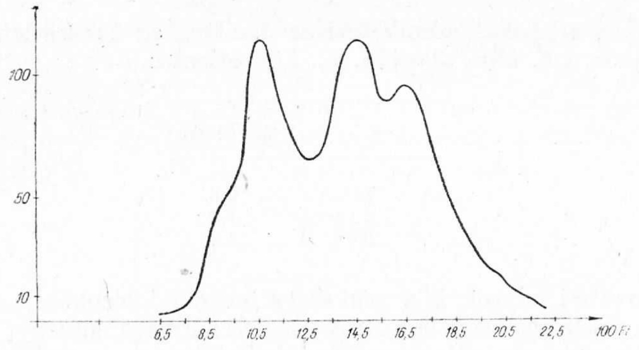
Az alkalmazott közelítés mellett az eredményül kapott átlagok:

takarító személyzet:	900 Ft	betanított munk.:	1400 Ft
segédmunkások:	1050 Ft	szakmunkások:	1650 Ft

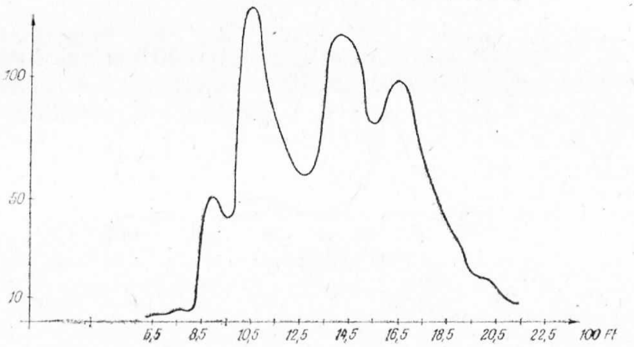


2. ábra

[2] A szóráscsökkentő eljárás lényege a következő: az eredeti keverék felhasználásával, megfelelő transzformációval olyan új keverék előállítása, mely az elsőtől csak az összetevők csökkentett szórásában különbözik, — ennek a ténynek lesz a következménye, a komponensek szétválása.



3. ábra



4. ábra

4. példa:

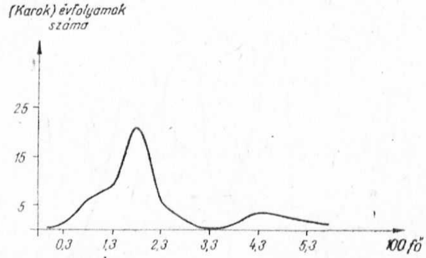
Az 1966/67-es tanévben Magyarországon a tudományegyetemek nappali tagozatán 12 kar, és az esetek többségében karonként 5 évfolyam volt található. Ha tekintjük az egyes karok létszámadatait, évfolyamok szerinti bontásban, 55 adatot kapunk:

Nappali hallgatók száma	karok számának gyakorisága
6— 55	1
56—105	6
106—155	9
156—205	21
206—255	6
256—305	2
306—355	0
356—405	1
406—455	4
456—505	3
506—555	2
	55

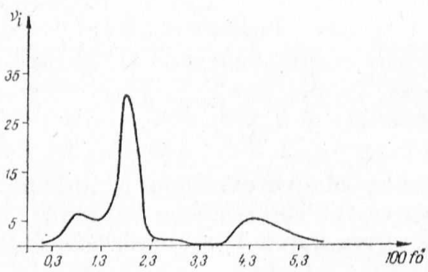
Az empirikus adatokat szemlélteti az 5. ábra, az ismeretlen paraméterek megállapíthatók a 6. ábra alapján, pl. közvetlenül:

k	m_k átlagok
1	0,8
2	1,8
3	4,3

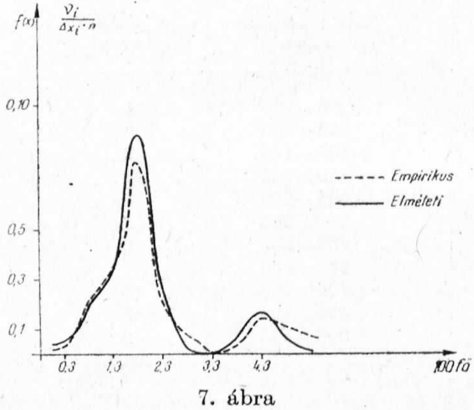
Valamint további adatok is a szükséges meggondolásokkal). Az empirikus, és az eljárás révén illesztett összetett elméleti sűrűségfüggvényt a 7. ábra tartalmazza. (Az egyes esetekben nem indokolt folytonos vonalak, csak a szemléltetést szolgálják.)



5. ábra



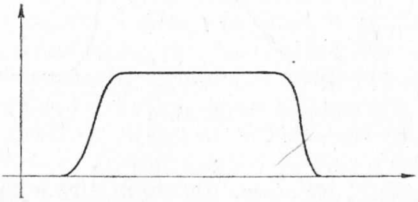
6. ábra



7. ábra

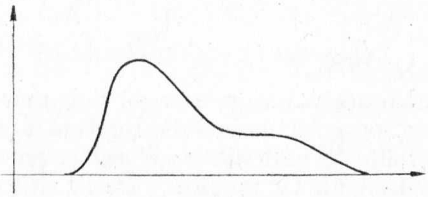
Keverékeloszlás típusok

Elméletileg az egyszerű eloszlások száma is végtelen. Gyakorlatilag azonban 10—20 eloszlástípus a problémák döntő többségében hozzásegít a megoldáshoz. Ugyanilyen okokból célszerű számbavenni a fontosabb keverékeloszlások alakjait is.



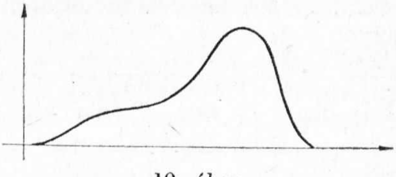
8. ábra

Ilyen típusú keverékeloszlás akkor keletkezik, ha a szóbajövő időtartam alatt az eredeti eloszlás középértékét meghatározó véletlen viszonyokra stabil, egyenletes változás szuperponálódik.



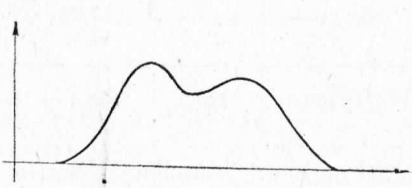
9. ábra

Ez a típus akkor keletkezik, ha a tekintett időtartam alatt az *a)* pont alatti egyenletes változás helyett kezdetben gyorsabb, majd egyre lassúbb változás lép fel.



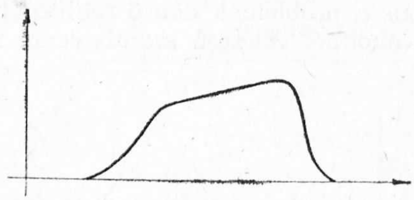
10. ábra

Ilyen típusú keveréket kapunk akkor, ha a változás a *b)* pont alatt leírttal ellentétes irányú, kezdetben lassú, majd egyre gyorsabb lesz.



11. ábra

Ezt a keverékeloszlás típust kapjuk, ha az előzőekben leírt változás üteme kezdetben gyorsuló, majd később lassulóvá válik.



12. ábra

Ezt az eredményt kapjuk, ha az *a)* pontban leírt körülmények mellett, valamilyen okból a szórás egyenletesen növekszik.

Ha a felbontást el akarjuk végezni, a komponensekre hipotéziseket kell felvennünk, de ebben az esetben is komoly számítástechnikai nehézségekre lehet számítani. A következő pontban egy ilyen természetű példával foglalkozunk. További elméleti megfontolásokat mindezzel kapcsolatban pl. a [14], [15] könyvekben találhatunk.

Összetett jövedelem-eloszlások

A társadalom egyes osztályai vagy rétegei más-más megoszlást mutatnak készpénzbevételeiknek kategóriánkénti összetétele szempontjából. Sőt, egy választott csoporton belül is változik az eloszlás, paraméterei, sőt típusa is módosulhat az idő múlásával. Ugyanakkor ezek a változások a vásárlóképesség alakulásával, az életszínvonal emelkedésével, stb. kapcsolatos törvényszerűségek lényeges részét képezik, megismerésük tehát igen fontosnak tekinthető.

5. Példa:

Anélkül, hogy a számszerű adatok becslési módszereit részleteznénk, tekintsük az alábbi táblázatot:

Pénzbevételi kategóriák a parasztság esetében:	1961	1962	1963	1964	1965
—3600	30,1	25,2	25,5	31,4	36,7
3601—4000	15,2	14,2	11,3	9,3	7,4
4001—4400	13,2	12,7	10,4	9,1	7,0
4401—4800	11,5	11,3	10,1	8,9	7,1
4801—5200	11,5	12,3	11,8	10,5	8,9
5201—5600	10,0	11,8	13,3	12,3	12,4
5601—	8,5	12,5	17,6	18,5	20,5
	100,0	100,0	100,0	100,0	100,0

A táblázat későbbi éveinél az empirikus eloszlásban *második csúcs* alakul ki. (Az osztályközök határai nem szerencsések, de az eredeti adatok hiányában már nem voltak megváltoztathatók). Jövedelemeloszlási problémák esetében elsősorban a logaritmikus normális eloszlást szoktuk alkalmazni, s valószínűségszámítási megfontolásokból kiindulva, a második csúcsot egy χ_n eloszlás alkalmas súlyozással történő keverésével vehetjük tekintetbe. A felbontás megvalósítása érdekében egyes konstansokat bizonyos feltételek alapján közvetlenül, másokat alkalmas egyenletrendszer felírásával, s annak közelítő módszerekkel való megoldásával határozhatunk meg. Ha az eljárást minden évben elvégezzük, s a paraméterek változását függvényalakban feltüntetjük, extrapolálással külön-külön megbecsülhetjük ezek későbbi értékeit is. Ezen adatok birtokában felírva a keverékeloszlás kifejezését, „megjósolhatjuk" a pénzbevételek kategóriánkénti $\%$-os alakulását az elkövetkező években. — Az eljárás nem mentes számítástechnikai nehézségektől, nem is kívánjuk részletezni, csak szerettünk volna bemutatni egy, az összetevőkre bontással kapcsolatos, jellegzetes, gyakorlati gondolatmenetet.

2. Idősorok vizsgálata

Statisztikai adatok egy sorozatát *idősornak* nevezzük, ha ezek az adatok valamely mutató különböző időpontokban mért értékei, és az adatokat a növekvő idő szerint rendeztük [10].

A matematikai statisztika egyik leginkább fejlődésben lévő és eredményeit is a legutóbbi évtizedekben elért területe az *idősorok analízise.* Nehezíti ugyanis az idősorok problémáinak a megoldását az a tény, hogy az esetenkint rendelkezésünkre álló egyetlen realizáció kizárólagos birtokában kell a kérdésekre válaszolnunk. Az időtengely alkalmas pontjaihoz tartozó idősor-adatok ugyanis valószínűségi változók, s éppen megfigyelt értékeik a lehetséges értékek összességéből egy mintaelemet képviselnek. Ráadásul a dolog természetéből még az is következik, hogy az egyes időpontokban nincs lehetőség újabb információk szerzésére, a megfigyelés ismételt elvégzésére. Erre vezethetők vissza az idősorok statisztikai analízisében megnyilvánuló nehézségek, s a fejlődés is ezért indult csak később meg. Felettébb megtévesztő lehet a mindennapi gyakorlatban elterjedt néhány egyszerű módszer, s ezek mechanikus használata (pl. lineáris trendvonal közvetlen alkalmazása, stb.). Gyakran találkozhatunk az idősorok vizsgálatát erre a szintre egyszerűsítő szemlélettel.

Az idősor-analízis legfontosabb problémái a paraméterek meghatározására, az összetevők szétválasztására (szűrésre), extrapolációra, és interpolációra terjednek ki. Hazai szakirodalmunkból is hiányzik még az idősorok elméleti és gyakorlati kérdéseivel behatóan foglalkozó könyv, ami az egyre szaporodó alkalmazások hatékonyságát és megbízhatóságát igen megnövelné.

Felbontás egyszerűbb esetekben

Több komponensből álló, összetett idősor felbontása esetén is a konkrét terület ismerete alapján, feltételekkel kell élnünk az egyes összetevők természetét illetően. Az egyik legegyszerűbb esetben például feltételezhetjük egy *lineáris alaptendencia* létezését (egyenes trendvonal), egy ismert periodicitással fellépő *periódikus komponens* szereplését, valamint az előzőek összegére természetszerűleg még szuperponálódó *véletlen hatásokat.* (Ezen utóbbiról szoktuk a stacionárius jelleget feltenni.)

Ebben az esetben a trend meghatározható pl. egyszerűen mozgó átlagolással, ezután kivonással leválasztható az eredeti idősorról, s már csak a másik két komponens összege áll előttünk. A periódikus összetevőt az azonos „fázishoz" tartozó adatok átlagolásával kapjuk meg, s ismételt kivonás után tisztán megmarad a véletlen komponens.

Meg kell jegyezni, hogy a bevezetőben mondott nehézségek, valamint a mozgó átlagolásból származó adatveszteségek miatt minden statisztikai feldolgozás esetén minél hosszabb idősor biztosítása kívánatos.

Kicsit is igényesebb munkák esetén lehetőség van pl. a trendvonal *megbízhatóságának* meghatározására, úgynevezett *konfidenciasáv* alakjában, a felbontás igazában csak ennek birtokában értékelhető. Fokozott mértékben merül fel ez az igény, ha extrapolálni akarjuk eredményeinket. Rövid idősor esetén ugyanis a megbízhatóság rohamosan romlik a szélső pontokban, nem is beszélve az extrapolált szakaszokról [10], [16].

Nem-lineáris alaptendenciák

Egyes területeken elméleti eredmények indokolják meghatározott görbetípusok szerepeltetését, más esetekben a tapasztalati adatok ábrázolása, vagy a mozgó átlagokkal való kiegyenlítés után dönthetünk alkalmasnak látszó függvények választásáról. Bár a két eset elméletileg lényegesen eltér, számítástechnikai szempontból hasonlóan kezelhető.

Röviden összefoglaljuk a leggyakrabban feltételezett nemlineáris alaptendencia-típusokat: (a többváltozós esetekre nem térünk ki):

$y = a \cdot t^b$ hatványfüggvény;

$y = a \cdot b^t$ exponenciális függvény;

$y = \dfrac{1}{a + b \cdot t}$ reciprok vagy hiperbolikus függvény;

$y = a_0 + a_1 \cdot t + a_2 \cdot t^2 + \ldots + a_n \cdot t^n$ n-ed foku parabola

$y = \dfrac{k}{1 + b \cdot e^{-a \cdot t}}$ logisztikus trend;

Természetesen mindig felléphetnek periódikus komponensek is, és ugyancsak szerepel a véletlen hatásokat tartalmazó összetevő is. Ennek megfelelően egy nem lineáris, összetett esetre vonatkozó feltételezés, pl. a következőképpen alakulhat:

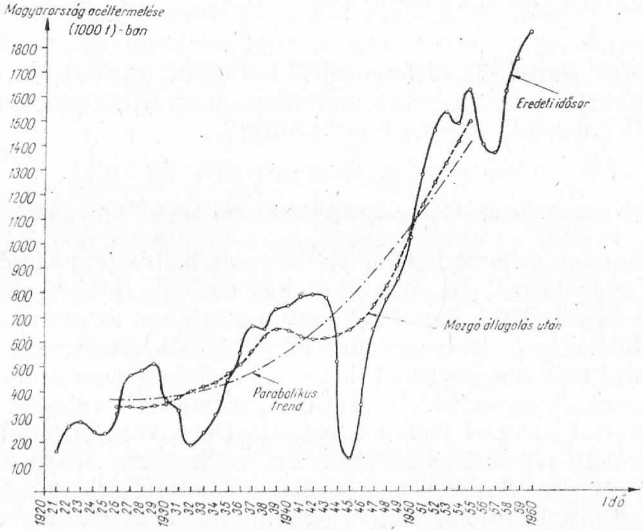
$$X_t = a \cdot t^b + d_t + Y_t$$

itt:

X_t = az idősor „t" időponthoz tartozó értéke;
$a \cdot t^b$ = az alaptendencia „t" időponthoz tartozó értéke;
d_t = a periódikus komponens időponthoz tartozó értéke;
Y_t = a véletlen hatásokat magába foglaló összetevő.

6. példa:

Tekintsük pl. az 1921—1960 között Magyarország évi acéltermelését megadó idősort. Az adatok és mellékszámítások nagy mennyisége, s így mellőzése miatt a példát egy ábrában foglaljuk össze. Bár ebben az esetben a periodicitás konjunkturális jellegű, s így hossza ingadozó, első lépésben mozgó átlagolást alkalmaztunk, s az ebből adódó görbéhez illesztettünk analitikusan, parabolikus trendet. A periódikus komponenseket a leírt módon egyszerűen meg tudtuk határozni. (Az idősor és a felbontás további diszkussziót igényelne, erre azonban a példa szemléltető jellege miatt nem térünk ki.)



13. ábra

Illeszkedő polinom fokszámának megállapítása

A kérdést gyakran közvetlen szemlélettel döntik el. Megnyugtató elintézése pl. a szukcesszív differenciák módszerével történhet [10]. Egy k-ad fokú polinom k-adik differenciálhányadosaként nyerhető konstans analógiájára, egy egész t értékeken értelmezett k-adfokú $q(t)$ polinom, k-adik differenciája is konstans lesz. Ha tehát rendre elkészítjük a vizsgált idősor növekvő rendű differenciáit, az így leszármaztatott idősorok tagjai a keresett fokszámnak megfelelő számú lépés után, véletlen ingadozásoktól, eltekintve, stabilizálódnak. A szórás minden lépés utáni kiszámításával a kívánt helyzetet könnyen felismerhetjük. [$q(t)$ első differenciája: $\Delta q(t) = q(t+1) - q(t)$, második differenciája: $\Delta^2 q(t) = \Delta q(t+1) - \Delta q(t) = q(t+2) - q(t+1) - [q(t+1) - - q(t)] = q(t+2) - 2\,q(t+1) + q(t)$ stb. A módszer igazolásához szükséges: a differenciák általános képzése (mivel ezek az eredeti idősor elemeihez hasonlóan valószínűségi változók), meg kell még keresni várható értéküket és szórásnégyzetüket.]

7. Példa:

Az alábbi példa[3] az USA-ban 1919 és 41 között évente fogyasztott húsmennyiség idősora alapján készült:

[3] (Átvéve a [10]-ből)

	Szórásnégyzet:
Az eredeti idősor adatai esetén:	62,25174
Az első differenciákból készült idősor esetén:	23,58636
A második differenciákból készült idősor esetén:	17,04112
A harmadik differenciákból készült idősor esetén:	15,73921
A negyedik differenciákból készült idősor esetén:	15,28964
Az ötödik differenciákból készült idősor esetén:	15,30118

A stabilizálódás a harmadik differenciától kezdődik, a trend-vonal egyenletéül tehát harmadfokú polinomot célszerű alkalmaznunk. A felbontás további lépései azután már a leírtak alapján folytatódnak.

Nagyobb számú periódikus komponens előfordulásának vizsgálata

Amilyen könnyen felismerhető a tiszta periodicitás, éppen olyan nehezen kezelhető ez a probléma, ha csak akár két periódikus komponens egyidejű hatásáról van is szó. Több összetevő esetén általában az eredőről semmilyen periodicitást közvetlenül leolvasni nem lehet. Ebből következik, hogy hosszú, a trend hatástól már megtisztított idősorok esetében igen érdekes vizsgálati lehetőséget jelent az úgynevezett ,,rejtett periódusok" keresése és meghatározása. Tekintettel az ezzel járó sok számítási munkára, kiindulásul célszerű statisztikai próbát végezni, ellenőrizve azt a hipotézist, hogy idősorunkban nincsenek rejtett periódusok. Ha a próba a hipotézisnek megfelelő szinten ellene mond, akkor elkészítjük a ,,periodogramot" melyről leolvasható a keresett periódikus komponensek száma és periódus hossza [10]. Ezután nyílik lehetőség a kapott eredmények például közgazdasági értékelésére. A módszerben rejlő sajátságok, lehetőségek, matematikai igazolása igen komoly apparátust igényel, egy adott esetben történő végigszámoláshoz viszont a gimnáziumi matematikaanyag is elegendő. A számítások mennyisége általában indokolja számítógép használatát.

8. példa:

A közelmúltban nagyobb meteorológiai adatgyűjtést végeztem a következő meggondolásokkal: az ország két megyényi — több szempontból meteorológiai egységet képező — területén a mérőállomások tavasztól őszig terjedő csapadékadatait 5 napos egységekre számítva rendeztem, 1900-ig visszamenően. (Az adatok végül 36 db 17×64-es típusú mátrixban helyezkedtek el.) Ezen az úton lehetőség nyílik az egyes évek ugyanazon 5 napos egységeiből valószínűségeloszlások konstruálására, s az egymásután következő 5 napos egységekhez tartozó eloszlások közötti kapcsolat, törvényszerűség keresésére. Másrészt felvetődik az a kérdés, találhatók-e rejtett periódusok egy meghatározott 5 napos egység hosszú távon való követésében, ill. van-e kapcsolat az egyes egységekhez tartozó rejtett periódusok között. Az eredményeket igen jól fel lehetne használni sztohasztikus, mezőgazdasági, távlati tervezési modellek készítésében. Érdekes módon kapcsolódik az előbbi adatgyűjtéshez egy párhuzamos vízügyi felmérés is, jelenleg az adatok elektronikus gépi feldolgozása folyik, az eredmények közlésére visszatérünk.

9. példa:

Szeretnék utalni a vonatkozó szakirodalom egy klasszikus példájára is, melyet W. BEVERIDGE közölt [10]. Ezek az adatok az angliai walesi búza ár-indexek trendhatástól már megszűrt értékeit tartalmazzák 1500-tól 1869-ig. A hosszúsága miatt feldolgozásra különösen alkalmas idősor periodogramja alapján 20 db eredetileg rejtett periódikus komponens (s ezek különböző periódus hossza) volt felismerhető.

Megjegyezzük, hogy a probléma sajátos vonását az idősor elemeinek valószínűségi változó jellege adja. Ha ugyanis az összetett görbe determinisztikus természetű, a „Fourier-analízis" eljárása a komponensek megkeresésének minden igényt kielégítő, általános módszere.

Trendszámítás ortogonális polinomokkal

Vizsgálni kívánt idősorunk ismert realizációja tartalmazzon n elemet. Kimutatható, hogy minden n-hez meghatározható a

$$\Phi_0(t) = 1; \; \Phi_1(t); \; \Phi_2(t), \ldots$$

polinomoknak egy olyan rendszere, melyekre:

$$\sum \Phi_i(t) \cdot \Phi_k(t) = 0 \qquad (i \neq k)$$

ezeket a polinomokat ortogonális polinomoknak nevezzük.[4]

Említettük már az előzőekben az idősorok analízisének azt az esetét, amikor a vizsgált terület ismerete alapján semmilyen görbetípus eleve nem rögzíthető, a tapasztalati adatokból kell tehát kiindulni. Ugyanakkor lehetséges, hogy ezek egy bonyolult görbét határoznak meg, s egy magasabb fokú polinom illesztése igen fáradságos lehet. Ilyenkor igen jó hasznát vehetjük az ortogonális polinomoknak [10], [13]. Technikailag igen jól kezelhetők, használatukhoz kényelmes képletek állnak rendelkezésre, helyettesítési értékeik széles határok között táblázva vannak [17], s belőlük a bonyolult görbék is jól előállíthatók. Sok szempontból nem jó, de egy lényeges vonásra mégis rámutató hasonlattal élve, idézhetjük a modern elektrotechnika „építőkocka elv" szerint felépített konstrukcióit. Előre alkalmasan elkészített, egymáshoz módszeresen illeszthető elemekből igen összetett rendszereket tudnak előállítani. Bizonyos határokon belül lehetőség nyílik arra is, hogy az egyes „kockák" tervezéséhez, előállításához szükséges ismeretek nélkül — jellemzőik ismeretében —, konstruktív munkánkban, felhasználhassuk őket. Bizonyos gyakorlattal és óvatossággal, a táblázatok és képletek használatával hasonló helyzetet teremthetünk az ortogonális polinomok alkalmazásánál is. Az előállított közelítő függvény ábrázolása az alapadatokat tartalmazó rendszerben, várhatóan úgyis felhívja a figyelmet az elkövetett elvi vagy számolási hibákra.

10. Példa:

Vizsgálva az Egyesült Királyság hajógyártásának adataiból készült idősort, bruttó regiszter tonnában:

[4] Ha $n = 2m$ páros szám, akkor: $t = -(2m-1), \; -(2m-3), \ldots, -1, \; +1, \ldots, 2m-3, \; 2m-1$; Ha $n = 2m+1$, páratlan szám, akkor: $t = -m, \; -m+1 \ldots, m$

a) 1861—1901 közötti 50 évben,
b) 1861—1960 közötti 100 évben,

(adatokat lásd a „Világgazdasági idősorok" c. könyvben), megállapítható, hogy a hosszú időszak ellenére a trendvonal egyenlete a legegyszerűbb összetevőkből felépíthető.

a) használva az ortogonális polinomokra bevezetett: $\Phi_k(t)$ jelölést, a közelítést

$$X_t = a_0 + a_1 \cdot \Phi_1(t) + a_2 \cdot \Phi_2(t)$$

alakban keressük, ahol:

$$\hat{a}_i = \frac{\sum\limits_t X_t \cdot \Phi_i(t)}{\sum\limits_t \Phi_i^2(t)}$$

($\hat{a}_i$ a_i becslését jelöli).

$$\Phi_1(t) = \lambda_1 \cdot t$$

$$\Phi_2(t) = \lambda_2 \cdot \left(t^2 - \frac{1}{12}(n^2 - 1)\right)$$

λ_1 és λ_2 tetszőleges, de praktikusan megválasztható, s a táblázatokban szintén közölt konstansok.

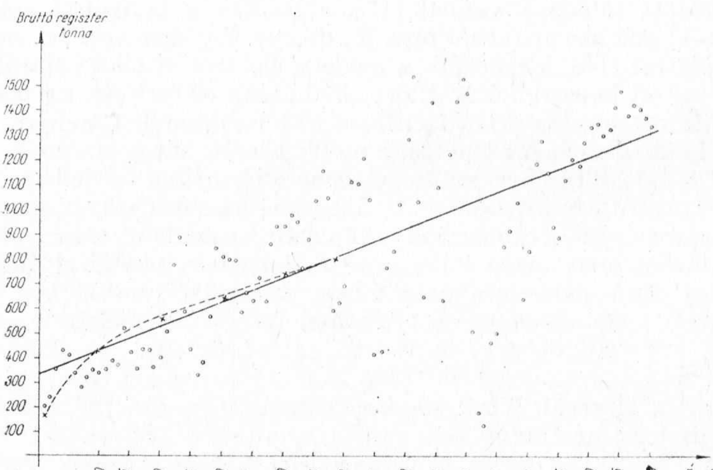
A számítások elvégzése után:

$$X_t = 594{,}66 + 4{,}770\,\Phi_1(t) + 0{,}103\,\Phi_2(t)$$

alakban adódott a trend.

b) Hosszabb távon a közelítés egész egyszerűvé vált, és

$$X_t = a_0 + a_1 \cdot \Phi_1(t)$$



14. ábra

alakban volt kereshető. Az együtthatók kiszámítása után:

$$X_t = 822{,}3 + 4{,}8494\,\Phi_1(t)$$

összefüggést kapjuk.

(t értéke ebben az esetben: $-99, -97, -95, \ldots -1, +1, \ldots, +97, +99$; értékeken van értelmezve.)

Végül szeretnék köszönetet mondani Chikán Attilának és Sólyom Csabának, akik a példák összeállításában voltak segítségemre.

(Beérkezett: 1968. VIII. 12.)

IRODALOM

[1] Doetsch, G.: Zerlegung einer Funktion in Gauss'sche Fehlerkurven und zeitliche Zurückverfolgung eines Temperaturzustandes, Math. Zeitschrift 41. 1936. p. 283–318.
[2] Gnyegyenko–Kolmogorov: Független valószínűségi változók összegeinek határeloszlásai Budapest, 1951.
[3] John, F.: Numerical solution of the equation of heat cond. Annali di Matematica Pura ed Applicata Serie IV. XL. 1955. p 129–142.
[4] Medgyessy, P.: Egy konvolúciós típusú integrálegyenlet numerikus megoldása MTA III. Oszt. Közleményei 16. 1966. 47–64. p.
[5] Medgyessy P.: Valószínűség-eloszlásfüggvények keverékének felbontása összetevőire. MTA Alk. Mat. Intézet Közl. II. 1952. 165–177. p.
[6] Medgyessy, P.: Anwendungsmöglichkeiten der Analyse der Wahrscheinlichkeitsdichtefunktionen Zeitschr. für angew. Math. und Mech. Band 37. 1957. Nr. 3/4. März–Apr.
[7] Medgyessy P.: Újabb eredmények val.-eloszlásfüggvények keverékének összetev. bontásával kapcs. MTA Alk. Mat. Int. Közl. III. 1954. 155–170. p.
[8] Medgyessy, P.: Decomposition of superposition of distr. functions Akadémiai Kiadó 1961.
[9] Medgyessy, P.: Stabil val. függvényekre fennálló parc. differenciálegyenletek és alkalmazásaik. MTA Alk. Mat. Kutató Közleményei I. 1956. 489–518. p.
[10] Prékopa A.–Éltető Ö.: Matematikai Jegyzetek IV. (Matematikai Statisztika) Statisztikai Kiadó Vállalat
[11] Meszéna Gy.: Számítások Bruns-sorokkal a természetes vizekben található uránnyom statisztikai értékeléséhez. Atomki Közlemények II. kötet 1960. 2. sz. MTA Atommagkutató Intézet Debrecen 99–107. p.
[12] Scherf E.–Meszéna Gy.: Matematikai statisztikai vizsgálatok a természetes vizek uránban való feldúsulásának fizikai feltételeiről. Atomki Közlemények II. kötet 1960. 2. sz. 109–143. p.
[13] Natanszon, I. P.: Konstruktív függvénytan, Akadémiai Kiadó, Budapest, 1952.
[14] Vincze I.: Statisztikai minőség ellenőrzés Közgazdasági és Jogi Könyvkiadó, Budapest, 1958.
[15] Smirnow, N. W.–Dumin, I. W.–Barkowski: Mathematische Statistik in der Technik. WEB. Deutscher Verlag der Wissenschaften. Berlin, 1963.
[16] Köves–Párniczki: Általános Statisztika. Tankönyvkiadó, Budapest, 1960.
[17] Pearson, E. S.–Hartley, H. O.: Biometrika tables for statisticians Cambridge-University Press, 1956.
[18] Medgyessy P.: Sűrűségfüggvény szuperpozíciók felbontásának egy lényegileg új módszeréről. MTA III. Osztály közleményei 17. 1967.