

Egy terjedési folyamat elemzése

(A hólabda-modell)

Bevezetés

Az elmúlt években néhány alkalommal egy érdekes játék terjedt el hetek alatt szinte az egész országban. Egyre többen kaptak olyan levelezőlapot, amely a játékba való bekapcsolódásra szólított fel. Akar-e Ön 5 hét alatt 25 000 Ft-ot nyerni? Ha igen, küldjön 10 Ft-ot egy megadott címre és gondoskodjék arról, hogy további 5 személy bekapcsolódjék a játékba! (Meg kell jegyezni, hogy ez a „megadott cím” egy korábbi játékos címe volt.) Ha mindenki becsületesen játszik, 5 hét alatt 25 000 Ft üti a markát!

A szöveg természetesen nem pontosan így szólt, de a játék lényege ez volt. Egy vagy több személy elindította a „hólabdát”, s az „becsületesen” ment is egészen addig, amíg eljutott arra a pontra, ahol már szinte mindenki ismert, s ha pontosan nem is tudták, de érezték, hogy ezen már nem lehet nyerni. A játékot természetesen betiltották, hiszen könnyen belátható, hogy egyesek — a játék elindítói, illetve azok, akik az elején bekapcsolódtak — valóban nagy összeget nyernek, mások viszont, ha kicsiben is, de pontosan ezt az összeget vesztik el. A „hólabda” tehát megállt, de matematikai elemzése mégsem haszontalan, hiszen belátható, hogy sok társadalmi, gazdasági folyamat „hólabdaszerűen” zajlik le, s közös *fejlődési, növekedési törvényszerűségeik* éppen a „steril hólabda” modell segítségével írhatók le.

Az információ terjedése, a fogyasztói szokások (divat) elterjedése, a nagyértékű, tartós fogyasztási cikkek elterjedése éppen úgy összefüggésben van a hólabda folyamattal, mint például a fertőző betegségek elterjedése. Mindezen folyamatok közös jellemzője az, hogy adott egy véges számú elemet tartalmazó zárt közeg, melynek elemei közül az induló állapotban egy, vagy néhány bizonyos kitüntetett tulajdonsággal rendelkezik (az információ birtokában van, van színes televíziója, fertőző beteg stb.). Ez a kitüntetett tulajdonság terjed, mégpedig úgy, hogy időegység alatt minden elem átlagosan a további elemnek adja tovább egészen addig, ameddig az összes elem nem rendelkezik ezzel a tulajdonsággal. E modellel kapcsolatban több feladat is megfogalmazható. Keressük mindenképp előt a folyamatot leíró függvényt, vizsgáljuk ennek tulajdonságait, majd adott megfigyelések esetén becsüljük a folyamat paramétereit, előre jelezzük várható alakulását és a prognózisra megbízhatósági intervallumot készítünk.

A fenti leírás természetesen erősen leegyszerűsített, modellszerű. A későbbiek során — elsősorban a gazdasági alkalmazásokat tekintve — látni fogjuk, hogy a különböző folyamatok nem teljesen analógok a hólabda folyamattal, viszont az is nyilvánvaló, hogy *alapotívációjuk* megegyezik, így a hólabda folyamat elemzése jó kiindulópontot jelent ezeknek a folyamatoknak a vizsgálatához. Ezen belül is — most már a közgazdasági jellegű feladatnál maradva — első-

sorban a fogyasztás, a fogyasztói szokások, a divat és más hasonló folyamatok időbeli lefutásának elméleti és ökonometria vizsgálatára alkalmazható ez a modell.

A hasonló jelenségeket a matematikai-közgazdaságtan általában a *logisztikus-függvény* segítségével vizsgálja. Ennek a függvénynek a tulajdonságai általánosan ismertek [4], [7], paramétereinek becslésére különféle eljárásokat dolgoztak ki [1], [3], [6], és létjogosultságát nagyszámú, sikeres gyakorlati alkalmazás is bizonyítja. Valószínűnek látszik, hogy hasonló folyamatokkal más tudományok (pl. fizika, biológia stb.) is foglalkoznak, de ezeknek a kutatásoknak az eredményei a közgazdasági kutatók számára ismeretlenek, így természetesen eredményeiket sem használhatják. (Ennek kapcsán felmerül a különböző tudományok, tudományágak közös fejlődési törvényszerűségeit együttesen vizsgáló, interdiszciplináris, rendszerszemléletű megközelítés — bizonyos vonatkozásokban már régóta sürgetett — szükségessége.)

A cikk első része a hólabda modell elméleti vizsgálatával foglalkozik. Kimutatja, hogy a legegyszerűbb, elméleti esetben a folyamatra vonatkozó, egymástól csak csekély mértékben különböző feltételezések milyen modellre és milyen növekedési folyamatra vezetnek. Megvizsgálja tulajdonságait és a legegyszerűbb esetre megadja a növekedési függvény explicit megoldását is. A második fő rész a gyakorlati vonatkozásokat tárgyalja. Kiterjeszti az elméleti modellt úgy, hogy az már numerikus felhasználásra is alkalmas legyen, összehasonlítja a „hólabda” folyamatot a logisztikus folyamattal, majd konkrét példán is bemutatja a hólabda folyamat ökonometria alkalmazását. A cikket az eredményeknek és a megoldatlan problémáknak az összefoglalása és az irodalomjegyzék zárja.

I. Az elméleti modell

Vezessük be a következő jelöléseket:

N : az adott zárt közeg elemeinek száma

n : a vizsgált, kitüntetett tulajdonság terjedési sebessége; jelentése az, hogy Δt idő alatt egy kitüntetett elem n elemnek adja át ezt a tulajdonságot;

s_t : a kitüntetett tulajdonsággal rendelkezők száma a t -ik időpontban.

A fenti mennyiségekre az alábbi relációknak kell teljesülniük:

$0 < N < \infty$ (hiszen véges számú elemből áll az alaphalmaz);

$0 < n < \infty$ és integer (hiszen n darabszámot jelöl);

$0 < s_t < N$ (definíció szerint);

$0 < t < \infty$ és integer (hiszen a folyamat lefutását diszkrét időpontokon vizsgáljuk).

A további vizsgálatok során csak az $n = 1$ esettel foglalkozunk, ami az elméleti modell esetében nem jelenti az általánosság megsértését, hiszen Δt tetszés szerint választható meg, úgy, hogy az $n = 1$ éppen teljesüljön.

1. 1. A modellek megfogalmazása és a folyamat leírása

Az elméleti modellt — melynek alapkérdését már a bevezetőben megadtuk — két változatban fogalmazzuk meg. Mindkét változat közös jellemzője, hogy egy kitüntetett tulajdonsággal rendelkező elem ezt a tulajdonságát a folyamat során nem veszti el, sőt aktív marad, azaz a tulajdonság terjesztésében részt

vesz. Közös jellemzője a két változatnak az is, hogy a kitüntetett elemek a folyamat lezajlása során egyenlő valószínűséggel adják át tulajdonságukat az összes többi elemnek. Itt meg kell jegyezni, hogy a tulajdonság terjesztésének számszerűsítésekor valamely elem saját maga felé való terjesztését (azaz a tulajdonság terjesztésének elmaradását) is figyelembe vettük. Ez a megoldás vitatható, és elsősorban az így adódó egyszerűbb formulák indokolják, ugyanakkor megjegyzendő, hogy az „önfertőzés” kihagyása esetén a korrektebb megoldás elvi útja teljesen megegyezik az itt leírtakkal, gyakorlatilag pedig — főként elég nagy N esetén — a két megoldás egyenértékű.

Az első (ún. alapmodell) változat fontos feltételezése az, hogy a tulajdonság terjedése meghatározott idő-intervallumokban történik, egy időintervallumon belül a fertőzés időbeli sorrendiségétől eltekint, hiszen ezen az intervallumon belül az újonnan megfertőzött elemek még nem aktívak, a kitüntetett tulajdonság csak bizonyos késleltetéssel, lappangási idővel terjed. Ez az eset fordul elő a hólabda folyamatnál, ahol — legalábbis elvben — a kitüntetett tulajdonság egy hetes kvantumokban terjed, de közel áll ehhez a modellhez a lappangási idővel járó fertőző betegségek terjedési folyamata is.

A második (az ún. továbbfejlesztett) modell ezzel szemben a megfertőzött elemek közt időbeli sorrendiséget tételez fel, azaz úgy számol, hogy egy elem megfertőződésekor azonnal aktív lesz, így a következő fertőzés kialakításakor már szerepet játszhat. Ez az előző — szakaszos — terjedéssel szemben valamilyen folyamos terjedést ír le. Úgy tűnik, hogy ez a második eset realisztikusabb, az alapmodell kezelése, megoldása viszont egyszerűbb. Mint később kimutatjuk, a két változat egymáshoz közel álló eredményeket szolgáltat. Végül utalni kell arra, hogy a modell az itt specifikált két változaton kívül nyilvánvalóan több változatban is felírható. Egy ilyen változat lehet például az, amikor egy elem egy bizonyos mennyiségű információt ad át a másiknak, amely azonban csak akkor aktivizálódik, ha a kapott információ volumene meghalad egy meghatározott határt. Valószínűnek látszik, hogy ezek a feladatok is hasonló folyamatokhoz vezetnek, bár itt csak az említett két változatot vizsgáltuk meg részletesebben.

1.1.1. Az alapmodell

Keressük tehát a kiemelt tulajdonsággal rendelkező elemek számának időbeli alakulását, feltéve, hogy a tulajdonság a Δt időintervallumon belül időbeli sorrendiség feltételezése nélkül terjed. Ekkor a kitüntetett elemek száma $t = 1$ időpontban

$$s_1 = s_0 + s_0 - g_0 = 2s_0 - g_0$$

lesz, azaz a kiinduló ($t = 0$) időpontban a kitüntetett elemek száma (s_0), azok száma akiknek ezek továbbadták s_0 , (hiszen $n = 1$ feltétel volt) és levonandó azon elemek száma, amelyek a $t = 0$ időpontban már rendelkeztek a tulajdonsággal és másodszer is megkapták. Látható, hogy g_0 valószínűségi változó, mely a $(0; s_0)$ intervallumban vehet fel egészszámú értéket. Mivel a további elemzések során g_0 (illetve g_t) fontos szerepet játszik, vizsgáljuk meg eloszlását!

Azonnal látható, hogy g_0 eloszlásának meghatározásakor egy egyszerű mintavételi feladattal állunk szemben. Adott az összes elem száma N , mely két csoportba sorolható

- a kiemelt tulajdonsággal rendelkezik s_0 db;
- nem rendelkezik ezzel a tulajdonsággal $N - s_0$ db.

Vegyünk egy s_0 elemű mintát, és vizsgáljuk meg mi a valószínűsége annak, hogy a mintában éppen k db a kiemelt tulajdonsággal rendelkező elem lesz!

$$\begin{aligned}
 P(g_0 = 0) &= \binom{s_0}{0} \left(\frac{s_0}{N}\right)^0 \left(\frac{N-s_0}{N}\right)^{s_0} \\
 P(g_0 = 1) &= \binom{s_0}{1} \left(\frac{s_0}{N}\right)^1 \left(\frac{N-s_0}{N}\right)^{s_0-1} \\
 &\vdots \\
 &\vdots \\
 P(g_0 = k) &= \binom{s_0}{k} \left(\frac{s_0}{N}\right)^k \left(\frac{N-s_0}{N}\right)^{s_0-k} \quad (k \leq s_0) \\
 &\vdots \\
 &\vdots \\
 P(g_0 = s_0) &= \binom{s_0}{s_0} \left(\frac{s_0}{N}\right)^{s_0} \left(\frac{N-s_0}{N}\right)^0.
 \end{aligned}$$

A fenti eloszlás ismeretében a folyamat már számítógépen szimulálható, de tulajdonságainak jobb megismerése érdekében határozzuk meg a várható értéket is. Ismeretes, hogy a fenti binomiális eloszlás várható értéke

$$E(g_0) = s_0 \frac{s_0}{N} = \frac{s_0^2}{N},$$

s ezt felhasználva s_1 várható értéke

$$E(s_1 | s_0) = s_0 + s_0 - E(g_0) = 2s_0 - \frac{s_0^2}{N} \quad \text{lesz.}$$

Mivel a folyamat a többi időpontra is hasonlóan írható fel:

$$s_2 = 2s_1 - g_1$$

\vdots
 \vdots

$$s_{t+1} = 2s_t - g_t,$$

a várható értékekre is igaz, hogy

$$E(s_2 | s_1) = 2s_1 - \frac{s_1^2}{N}$$

\vdots
 \vdots

$$E(s_{t+1} | s_t) = 2s_t - \frac{s_t^2}{N}.$$

A folyamat időbeli lefutásának *várható értékét* tehát egy elsőrendű, nemlineáris rekurzív formula határozza meg, ami természetesen felírható differencia-egyenlet alakjában is az alábbi módon:

$$\Delta s_t = s_{t+1} - s_t = s_t - \frac{s_t^2}{N} = s_t \left(1 - \frac{s_t}{N} \right).$$

Ez a forma egyértelműen meghatározza a folyamat várható értékét minden t időpontra, de mielőtt elemeznénk, illetve explicite időfüggvényre megoldanánk, felírjuk a továbbfejlesztett feladatot.

1.1.2. *A továbbfejlesztett modell*

Ez a modell — mint arra már utaltunk — csupán abban különbözik az előzőtől, hogy itt egy elem a megfelelő tulajdonságot nem csak egy másiktól kaphatja meg. A modellt — az előző jelölések megtartásával — a következőképpen írhatjuk fel:

$$\begin{aligned} s_1 &= s_0 + s_0 - \tilde{g}_0 = s_0 + \tilde{h}_0 \\ s_2 &= s_1 + s_1 - \tilde{g}_1 = s_1 + \tilde{h}_1 \\ &\vdots \\ s_{t+1} &= s_t + s_t - \tilde{g}_t = s_t + \tilde{h}_t, \end{aligned}$$

ahol \tilde{h}_t azon elemek számát jelöli, amelyek a $(t; t + 1)$ időintervallumban a kiemelt tulajdonságot megkapták. Keresendő \tilde{h}_t eloszlása és várható értéke, melyek segítségével a folyamat szimulálható, illetve várható értéke felírható. (Az itt következő levezetés során az egyszerűség kedvéért a t indexet elhagyjuk.)

Definiáljunk egy s elemű $z = (\xi_1 \xi_2 \dots \xi_s)$ vektort a következőképpen

- $\xi_i = 0$, ha az i -edik elem olyan elemnek adja át a tulajdonságot, amely már rendelkezett vele;
- $\xi_i = 1$, ha a tulajdonságot olyan elem kapja, amely még nem rendelkezett vele.

Nyilvánvaló, hogy \tilde{h} a 0 értéket csak úgy veheti fel, ha minden elem csak már „fertőzött” elemnek adja át a tulajdonságot, azaz valamennyi ξ_i 0 értéket vesz fel, tehát

$$P(h = 0) = P(\xi_1 = 0; \dots \xi_s = 0).$$

\tilde{h} az 1 értéket többféleképpen veheti fel. Vagy az első elem talál „új, még nem fertőzött” elemet és a többi nem, vagy a második stb . . . Fontos megjegyezni, hogy az egyes elemek aktivizálódása közt *időbeli sorrendiséget* tételeztünk fel. Ez jogos lehet abban az esetben, ha egy homogén sokaságot vizsgálunk, hiszen két „fertőzés” tökéletes időbeli egybeesése gyakorlatilag kizárt. Ezért pl. a második elem „fertőzési esélyei” függnak attól, hogy az első elem talált-e új, még nem fertőzött elemet a mintában, azaz ξ_1 és ξ_2 valószínűségi változók egymástól *nem függetlenek*. A megfelelő valószínűségek az elmondottak értelmében a következőképpen írhatók fel:

$$\begin{aligned}
 P(\tilde{h} = 1) &= P(\xi_1 = 1) P(\xi_2 = 0 \mid \xi_1 = 1) P(\xi_3 = 0 \mid \xi_1 + \xi_2 = 1) + \dots \\
 &+ P(\xi_1 = 0) P(\xi_2 = 1 \mid \xi_1 = 0) P(\xi_3 = 0 \mid \xi_1 + \xi_2 = 1) + \dots \\
 &\vdots \\
 &+ P\left(\xi_s = 1 \mid \sum_{i=1}^{s-1} \xi_i = 0\right) \\
 &\vdots \\
 P(\tilde{h} = k) &= P(\xi_1 = 1) P(\xi_2 = 1/\xi_1 = 1) \dots P\left(\xi_{k+1} = 0 \mid \sum_{i=1}^k \xi_i = k\right) + \dots \\
 &\vdots \\
 P(\tilde{h} = s) &= P(\xi_1 = 1) P(\xi_2 = 1 \mid \xi_1 = 1) \dots P\left(\xi_s = 1 \mid \sum_{i=1}^{s-1} \xi_i = s-1\right).
 \end{aligned}$$

Látható, hogy \tilde{h} eloszlása, bár meglehetősen bonyolult, a folyamat szimulációjára alkalmas¹, de zárt formában aligha oldható meg. Meghatározható viszont \tilde{h} várható értéke az alábbi módon:

$$\begin{aligned}
 E(\tilde{h}) &= E(\xi_1 + \xi_2 + \dots + \xi_s) = E(\xi_1) + E(\xi_2) + \dots + E(\xi_s) \\
 E(\xi_1) &= \frac{N-s}{N} \\
 E(\xi_2) &= E(\xi_2 \mid \xi_1 = 0) P(\xi_1 = 0) + E(\xi_2 \mid \xi_1 = 1) P(\xi_1 = 1) = \\
 &= \frac{N-s}{N} \cdot \frac{s}{N} + \frac{N-s-1}{N} \cdot \frac{N-s}{N} = \frac{N-s}{N} \left(\frac{N-1}{N}\right) = \\
 &= \frac{N-s}{N} - \frac{1}{N} \frac{N-s}{N} = E(\xi_1) - \frac{1}{N} E(\xi_1).
 \end{aligned}$$

Általánosságban is felírható, hogy

$$\begin{aligned}
 E(\xi_k) &= E\left(\xi_k \mid \sum_{i=1}^{k-1} \xi_i = 0\right) P\left(\sum_{i=1}^{k-1} \xi_i = 0\right) + \\
 &+ E\left(\xi_k \mid \sum_{i=1}^{k-1} \xi_i = 1\right) P\left(\sum_{i=1}^{k-1} \xi_i = 1\right) + \\
 &\vdots \\
 &+ E\left(\xi_k \mid \sum_{i=1}^{k-1} \xi_i = k-1\right) P\left(\sum_{i=1}^{k-1} \xi_i = k-1\right) = \\
 &= \sum_{j=0}^{k-1} E\left(\xi_k \mid \sum_{i=1}^{k-1} \xi_i = j\right) P\left(\sum_{i=1}^{k-1} \xi_i = j\right). \quad (1)
 \end{aligned}$$

¹ Erre F. Liska Tibor végzett kísérleti számításokat.

Könnyen belátható, hogy

$$E\left(\xi_k \mid \sum_{i=1}^{k-1} \xi_i = j\right) = \frac{N - s - j}{N}$$

(ez ξ_i definíciójából következik), s így az (1) formula az alábbi alakot ölti:

$$\begin{aligned} E(\xi_k) &= \sum_{j=0}^{k-1} \frac{N - s - j}{N} P\left(\sum_{i=1}^{k-1} \xi_i = j\right) = \\ &= \sum_{j=0}^{k-1} \frac{N - s}{N} P\left(\sum_{i=1}^{k-1} \xi_i = j\right) - \sum_{j=0}^{k-1} \frac{j}{N} P\left(\sum_{i=1}^{k-1} \xi_i = j\right). \end{aligned}$$

Mivel

$$\sum_{j=0}^{k-1} P\left(\sum_{i=1}^{k-1} \xi_i = j\right) = 1$$

és

$$\sum_{j=0}^{k-1} j P\left(\sum_{i=1}^{k-1} \xi_i = j\right) = E\sum_{i=1}^{k-1} \xi_i = \sum_{i=1}^{k-1} E(\xi_i),$$

ξ_k várható értékre az alábbi összefüggés adódik

$$E(\xi_k) = \frac{N - s}{N} - \frac{1}{N} \sum_{i=1}^{k-1} E(\xi_i). \quad (2)$$

A (2) képletből teljes indukcióval belátható, hogy az alábbi rekurzió érvényes

$$E(\xi_k) = \left(1 - \frac{1}{N}\right) E(\xi_{k-1}), \quad (3)$$

hiszen

$$E(\xi_{k-1}) = \frac{N - s}{N} - \frac{1}{N} \sum_{i=1}^{k-2} E(\xi_i)$$

és

$$\begin{aligned} E(\xi_k) &= \frac{N - s}{N} - \frac{1}{N} \sum_{i=1}^{k-1} E(\xi_i) = \frac{N - s}{N} - \frac{1}{N} \sum_{i=1}^{k-2} E(\xi_i) - \frac{1}{N} E(\xi_{k-1}) = \\ &= E(\xi_{k-1}) - \frac{1}{N} E(\xi_{k-1}). \end{aligned}$$

A (3) képlet felhasználásával

$$E(\xi_1) = \frac{N - s}{N}$$

$$E(\xi_2) = 1 - \frac{1}{N} E(\xi_1)$$

⋮

$$E(\xi_k) = \left(1 - \frac{1}{N}\right) E(\xi_{k-1}) = \left(1 - \frac{1}{N}\right)^{k-1} E(\xi_1) = \left(1 - \frac{1}{N}\right)^{k-1} \left(\frac{N - s}{N}\right)$$

adódik. Ekkor \bar{h} keresett várható értéke már megkapható az alábbi módon²:

$$\begin{aligned} E(\bar{h}) &= \sum_{i=1}^s E(\xi_i) = \frac{N-s}{N} \sum_{i=1}^s \left(1 - \frac{1}{N}\right)^{i-1} = \frac{N-s}{N} \frac{1 - \left(1 - \frac{1}{N}\right)^s}{1 - \left(1 - \frac{1}{N}\right)} = \\ &= (N-s) \left[1 - \left(1 - \frac{1}{N}\right)^s\right]. \end{aligned}$$

A kapott formulát érdemes tovább vizsgálni a szögletes zárójelben levő kifejezés binomiális kifejtésével:

$$\begin{aligned} 1 - \left(1 - \frac{1}{N}\right)^s &= 1 - \left[\binom{s}{0} - \binom{s}{1} \frac{1}{N} + \binom{s}{2} \frac{1}{N^2} - \dots \right] = \\ &= 1 - \left(1 - \frac{s}{N} + \binom{s}{2} \frac{1}{N} \dots\right) = \frac{s}{N} + R_1, \end{aligned}$$

\bar{h} várható értéke ekkor az

$$E(\bar{h}) = (N-s) \frac{s}{N} + R_2 \quad (4)$$

alakot ölti, ahol R_2 a maradéktag és nagyságára *becslés konstruálható*.

A továbbfejlesztett modell bemutatása után megmutatjuk, hogy az alapmodell ennek speciális esete. Az itteni felírásnak megfelelően az alapmodell esetében $h = s - g$ és az itt definiált ξ_i karakterisztikus valószínűségi változók az alapmodell esetében egymástól függetlenek. A függetlenség abból következik, hogy a Δt időintervallumon belül nem tételeztük fel az újonnan fertőzött elemek aktivizálódását.

Ekkor, mivel

$$P(\xi_i = 1) = \frac{N-s}{N}$$

és

$$P(\xi_i = 0) = \frac{s}{N},$$

s így

$$E(\xi_i) = \frac{N-s}{N}$$

a $h = \sum_{i=1}^s \xi_i$ változó binomiális eloszlást követ $(N-s)/N$, illetve s/N paraméterekkel, és várható értéke $E(h) = s(N-s)/N$ lesz.

Ekkor az is látható, hogy a két feladatban központi fontosságú $E(h)$, illetve $E(\bar{h})$ érték csak egy maradéktagban tér el egymástól, amiből arra lehet következtetni, hogy egyik a másikkal jól közelíthető.

Visszatérve a továbbfejlesztett modellre, annak általános tagja az alábbi rekurzióval fejezhető ki

$$E(s_{t+1} | s_t) = s_t + (N - s_t) \left[1 - \left(1 - \frac{1}{N}\right)^{s_t}\right].$$

²Hasonló, de nem teljesen azonos feladat megoldását idézi más levezetéssel Prékopa. A [5].

Ez a formula nem túlságosan áttekinthető, éppen ezért a továbbiakban az elemzést a lényegesen egyszerűbb alapmodellen végezzük el és numerikus számításokra is azt használjuk. Ahhoz azonban, hogy az alapmodell tulajdonságait a továbbfejlesztett modellre is kiterjesszük, becsülni kellene az R_2 maradéktagot és elemezni viselkedését. Ez a vizsgálat eddig nem készült el.

1. 2. Az alapmodell tulajdonságai és explicit megoldása

Az $s_{t+1} = 2s_t - \frac{s_t^2}{N}$ alakú alapmodell olyan folyamatot ír le, melynek legfontosabb tulajdonságai az alábbiak:

- s_t monoton növekvő;
- s_t felső korlátja N , és végül
- s_t határértéke $t \rightarrow \infty$ esetén N .

A fenti tulajdonságok közül az első kettő bizonyítható a rekurzív formula alapján. Az első tulajdonság azonnal belátható, ha az egyenletet differenciaegyenletként írjuk fel, hiszen a $\Delta s_t = s_t(1 - s_t/N)$ formában $0 \leq s_t < N$ (ami kiinduló feltétel volt) biztosítja Δs_t pozitivitását, ami s monoton növekedését jelenti.

A második tulajdonság bizonyítására írjuk fel az $s_t < N$ kiinduló feltételt az alábbi módon:

$$s_t + \varepsilon = N, \quad \varepsilon > 0$$

Azt kell belátni, hogy $s_{t+1} < N$. A bizonyítás igen egyszerű, hiszen

$$\begin{aligned} s_{t+1} &= 2s_t - \frac{s_t^2}{N} = 2(N - \varepsilon) - \frac{(N - \varepsilon)^2}{N} = \\ &= 2N - 2\varepsilon - \frac{N^2 - 2\varepsilon N + \varepsilon^2}{N} = 2N - 2\varepsilon - N + 2\varepsilon - \frac{\varepsilon^2}{N} = N - \frac{\varepsilon^2}{N} < N, \end{aligned}$$

azaz, $s_{t+1} < N$ adódik, ami az N felső korlát voltát bizonyítja.

A harmadik tulajdonság bizonyítását az explicit megoldott időfüggvény alapján lényegesen egyszerűbb elvégezni. Ezért, most előbb megadjuk az alapmodell megoldását, azaz a folyamat várható értékének explicit időfüggvényét³.

Kiindulva az $s_{t+1} = 2s_t - s_t^2/N$ alakból, szorozzuk meg mindkét oldalt $-N$ -nel és adjunk mindkét oldalhoz N^2 -et, így a jobb oldal teljes négyzetté válik

$$N^2 - Ns_{t+1} = s_t^2 - 2s_t N + N^2,$$

azaz

$$N(N - s_{t+1}) = (N - s_t)^2.$$

Bevezetve a $Z_t = N - s_t$ helyettesítést, azt kapjuk, hogy

$$Z_{t+1} = \frac{Z_t^2}{N}.$$

³ Ez a megoldási mód Tihanyi Ambrustól származik.

Kiindulva Z_0 értékéből sorozatosan képezhetjük a Z_i értékeket

$$Z_1 = \frac{Z_0^2}{N}; \quad Z_2 = \frac{Z_1^2}{N} = \frac{\left(\frac{Z_0^2}{N}\right)^2}{N} = \frac{Z_0^4}{N^3}; \quad \dots \quad Z_t = \frac{Z_0^{2^t}}{N^{2^t-1}}.$$

Visszatérve az eredeti változóra

$$N - s_t = \frac{(N - s_0)^{2^t}}{N^{2^t-1}},$$

és átlakítások után

$$s_t = N \left[1 - \left(1 - \frac{s_0}{N} \right)^{2^t} \right]$$

adódik.

Ez a növekedési folyamatok leírására a szakirodalomból már ismert *Gompertz függvény* [7] egy *speciális változata*. Speciális két okból:

- egyrészt azért, mert az ökonometria ezeket a függvényeket folytonos függvényként értelmezi, míg itt *csak diszkrét t értékek mellett vesz fel értékeket*;
- másrészt azért, mert az *időváltozó a 2 alapú hatvány kitevőjeként szerepel*, míg az általános esetben az időváltozó hatványalapja tetszőleges pozitív értéket vehet fel.

Az explicit megoldásfüggvény ismeretében most már a harmadik tulajdonság is könnyen belátható, hiszen

$$\lim_{t \rightarrow \infty} N \left[1 - \left(1 - \frac{s_0}{N} \right)^{2^t} \right] = N - N \lim_{t \rightarrow \infty} \left(1 - \frac{s_0}{N} \right)^{2^t} = N.$$

Az eddigiek alapján a hólabda folyamat elméleti vizsgálata az alábbi *összefoglaló következtetések*hez vezetett:

- az *alapmodell Gompertz típusú növekedést mutat*, a kiemelt tulajdonsággal rendelkezők száma monoton növekszik, eleinte lassú, majd gyorsuló, végül ismét lassuló ütemben, és elég sok lépés után közelíti meg a felső korlátot, mely az összes elem számával egyenlő.
- a továbbfejlesztett modell hasonló tulajdonságot mutat az alapmodellhez, a kettő minden lépésben csak egy maradéktagban tér el egymástól, de erre a maradéktagra eddig jó becslést nem tudunk adni;
- a fenti modell alapján ki lehet mutatni a hólabda játék „inkorrekt” voltát, pontosabban azt, hogy kinek milyen esélye van a nyeresre és a vesztesre, de ez természetesen nem célja e cikknek. Sokkal inkább célunk az, hogy az eddigi eredményeket kiterjesszük egy olyan formára, amely már konkrét *ökonometriai számítások* céljára is alkalmas. Ezzel foglalkozik a következő érsz.

2. Gyakorlati alkalmazások

A gyakorlati alkalmazáshoz először a modellnek azt a módosítását készítjük el, amelynek alapján numerikus számítások végezhetőek, majd rámutatunk az itteni eredmények és az ökonometriában hagyományosan alkalmazott eljárás közti különbségre, végül pedig számszerű példát mutatunk be az elmondottak illusztrálására.

2.3. A statisztikai modell

Az alapmodell (és a továbbfejlesztett modell) azért nem alkalmas közvetlenül numerikus számításokra, mivel feltételeztük, hogy időegység alatt az átadások száma (a terjedési sebesség) egységnyi. Ezt egy elméleti modellben fel lehetett tételezni, hiszen az időegységet tetszés szerint választottuk meg, de *gyakorlati számítások esetében az időegységet a statisztikai adatbázis határozza meg* és semmi sem garantálja, hogy a terjedési sebesség egységnyi legyen.

A probléma megoldása érdekében definiáljuk n -t a korábbiaktól eltérően olyan 0 és 1 közé eső számként, amely azt mondja meg, hogy egy elem átlagosan milyen valószínűséggel adja tovább a kiemelt tulajdonságot. Az elméleti modellek esetében ez a szám 1 volt, azaz a továbbadást biztos eseménynek feltételeztük. Ez a némiképp zavaró jelölés azért indokolt, mivel n ez esetben is a terjedési sebesség mutatója.

Definiáljuk az új $\bar{\xi}_i$ ($i = 1, 2, \dots, s$) változókat (a t index átmeneti elhagyásával) a korábbi ξ_i változókhoz hasonló módon, de már az új, fenti feltételek mellett.

$\bar{\xi}_i = 0$ ha az i -edik elem olyan elemnek adja át a tulajdonságot, amely már rendelkezett vele;

$\bar{\xi}_i = 1$ ha a tulajdonságot olyan elem kapja, amely még nem rendelkezett vele

Mivel $\bar{\xi}_i = n \xi_i$ és $E(\bar{\xi}_i) = n E(\xi_i)$ az *alapmodell* esetében $E(\bar{h}) = \sum_{i=1}^s E(\bar{\xi}_i) = n \sum_{i=1}^s E(\xi_i) = nE(h) = ns(N - s)/N$, avagy a t indexek bevezetésével:

$$E(\bar{h}_t) = ns_t \frac{N - s_t}{N}. \tag{5}$$

Az (5)-ből pedig felírható az alapmodell statisztikai változatának differenciaegyenlete az alábbi módon:

$$\Delta s_t = s_{t+1} - s_t = ns_t \left(1 - \frac{s_t}{N} \right),$$

avagy

$$s_{t+1} = (n + 1) s_t - \frac{ns_t^2}{N}. \tag{6}$$

Erre a folyamatra a korábbiakhoz hasonló módon belátható az s_t értékek monotonitása. A forma explicit megoldása ezideig nem sikerült és semmi sem bizonyítja, hogy ez a megoldás (t , illetve s_0 függvényében) zárt alakban előállítható. Ezért a második és harmadik tulajdonságot, nevezetesen azt, hogy N a sorozat felső korlátja és egyben határértéke $t \rightarrow \infty$ esetén, a (6) formula alapján bizonyítjuk. Ehhez ismét az

$$s_t = N - \varepsilon, \quad 0 < \varepsilon < N$$

felírást használjuk. Ekkor

$$s_{t+1} = (n + 1) (N - \varepsilon) - \frac{n(N - \varepsilon)^2}{N} = N - \varepsilon \left(1 - n + \frac{n\varepsilon}{N} \right).$$

A zárójelben levő kifejezésre az alábbi egyenlőtlenség érvényes:

$$0 < 1 - n + \frac{n\varepsilon}{N} < 1,$$

hiszen $n-1 < n\varepsilon/N$, ami a baloldali egyenlőtlenség, valamint $-n + n\varepsilon/N < 0$, ami pedig a jobboldali egyenlőtlenség fennállását igazolja. A baloldali egyenlőtlenségből az következik, hogy $s_{t+1} < N$, azaz N a sorozat felső korlátja. A jobboldali egyenlőtlenség pedig azt jelenti, hogy N -hez tetszés szerint közel eső s_t esetére is igaz az, hogy s_{t+1} még közelebb esik N -hez, azaz N a sorozat supremuma. Ez egyben azt is jelenti, hogy N a sorozat, illetve a folyamat határértéke.

A továbbfejlesztett modell statisztikai változata némiképp bonyolultabb, de levezetése teljesen analóg az elméleti változatával.

$$\text{Ismeretes, hogy } P(\bar{\xi}_1 = 1) = nP(\xi_1 = 1) = n - \frac{N-s}{N}$$

$$P(\bar{\xi}_1 = 0) = 1 - P(\bar{\xi}_1 = 1) = 1 - n \frac{N-s}{N},$$

s ezekből

$$E(\bar{\xi}_1) = n \frac{N-s}{N}.$$

A $\bar{\xi}$ változóknak a korábban leírt összefüggése alapján

$$\begin{aligned} E(\bar{\xi}_2) &= E(\bar{\xi}_2 | \bar{\xi}_1 = 0) P(\bar{\xi}_1 = 0) + E(\bar{\xi}_2 | \bar{\xi}_1 = 1) P(\bar{\xi}_1 = 1) = \\ &= n \frac{N-s}{N} \left(1 - n \frac{N-s}{N} \right) + n \frac{N-s-1}{N} \cdot n \frac{N-s}{N} = \\ &= n \frac{N-s}{N} \left(1 - \frac{n}{N} \right) = E(\bar{\xi}_1) - \frac{n}{N} E(\bar{\xi}_1). \end{aligned}$$

Ekkor (2)-nek megfelelően:

$$E(\bar{\xi}_k) = n \frac{N-s}{N} - \frac{n}{N} \sum_{i=1}^{k-1} E(\bar{\xi}_i)$$

$$E(\bar{\xi}_k) = n \left(1 - \frac{n}{N} \right)^{k-1} \left(\frac{N-s}{N} \right)$$

$$E(\bar{h}) = (N-s) \left[1 - \left(1 - \frac{n}{N} \right)^s \right].$$

Ez sorbafejtés után $E(\bar{h}) = ns \frac{N-s}{N} + R_3$ alakra írható.

Ez a forma analóg az alapmodell megfelelő formájával, de részletes elemzésére csak az R_3 maradéktag ismeretében lehet vállalkozni.

2.4. A hólabda folyamat és a logisztikus növekedés

A telítettségi szinttel rendelkező gazdasági folyamatokat (például a tartós fogyasztási cikkek állományának időbeli alakulását) az ökonometriában a leggyakrabban a *logisztikus függvény*nel írják le. A logisztikus függvényt a *Robertson differenciálegyenlet*ből [4] szokták levezetni, amelynek alakja:

$$\frac{dy}{dt} = \frac{b}{k} y(k - y).$$

Ennek megoldása:

$$y = \frac{k}{1 + ae^{-bt}}$$

az említett logisztikus függvény. Érdemes rámutatni a Robertson egyenlet és a (6) képlet közti hasonlóságra, amely az alábbi alakba is felírható:

$$\Delta s_t = \frac{n}{N} s_t(N - s_t).$$

Még szemléletesebb az analógia, ha a logisztikus függvény diszkrét pontokra felírt formáját tekintjük. (Ezt a formát a logisztikus függvény paramétereinek becslésekor szokták használni [6].)

$$y_{t+1} = \left(\frac{1 - e^b}{e^b} + 1 \right) y_t - \frac{1}{k} \cdot \frac{1 - e^b}{e^b} y_t y_{t+1}.$$

Ha elvégezzük az $\frac{1 - e^b}{e^b} = \tilde{n}$ és a $k = \tilde{N}$ helyettesítést, akkor az

$$y_{t+1} = (\tilde{n} + 1) y_t - \frac{\tilde{n}}{\tilde{N}} y_t y_{t+1} \tag{7}$$

alakhoz jutunk, melynek „párja” a hólabda esetében az

$$s_{t+1} = (n + 1) s_t - \frac{n}{N} s_t \cdot s_t \tag{8}$$

alak. A két folyamat összevetéséből az alábbi következtetések vonhatók le:

- a két folyamat egymáshoz való viszonya a *folytonos és a diszkrét* modell, illetve megoldás közötti különbségre vezethető vissza;
- a „hólabda”-modell, mint diszkrét folyamat jobban reprezentálja a gazdasági valóságot, elméletileg megalapozottabb és ennek folytán alkalmas arra, hogy a terjedési folyamatokhoz minden egyes időpontban valószínűség-eloszlást és ennek segítségével konfidencia intervallumot határozzunk meg;
- a logisztikus függvény folytonosságánál fogva bizonyos esetekben kényelmesebben kezelhető, explicit időfüggvénye könnyebben megoldható, a folyamat tulajdonságai jobban elemezhetőek;
- a (7) formula alapján látható, hogy amennyiben a megfigyelési időpontokat „sűrítjük”, a két függvény közti különbséget okozó tagok egymáshoz közelítenek, hiszen

$$\lim_{\Delta t \rightarrow 0} y_t y_{t+\Delta t} = y_t^2;$$

- bár a paraméterbecslésről — amely a numerikus alkalmazhatóság egyik fontos kérdése — csak a későbbiekben lesz szó, azt már az eddigiek alapján is megállapíthatjuk, hogy mindkét folyamat alkalmazásának vannak előnyös és hátrányos oldalai. Ezért nem is vetjük fel azt a kérdést, hogy melyik jobb, hanem arra hívjuk fel a figyelmet, hogy *a két elemzés egymást kiegészíti*; ismereteinket teljesebbé, a gyakorlati alkalmazást hatékonyabbá teszi az együttes vizsgálat;
- visszatérve az alapmodell megoldására a Gompertz és a logisztikus függvény bizonyos kapcsolatára is fény derül. Megjegyezzük, hogy korábban [7] a két függvény között — más megközelítés alapján — rokonságot lehetett kimutatni.

A továbbiakban a paraméterbecsléssel kapcsolatos problémákat tekintjük át, s az elmondottakat számszerű példával is illusztráljuk.

2.5. Paraméterbecslési eljárások és numerikus eredmények

Az ökonometria gyakorlata a logisztikus függvény paramétereit két fő módon *becsüli*: az explicit időfüggvényből nemlineáris normálegyenletrendszer felhasználásával (közönséges legkisebb négyzetek módszerével), vagy a (7) rekurzív összefüggésrendszer felhasználásával két paraméterre linearizált formából, a harmadikra pedig iterációval. Ezek az eljárások meglehetősen bonyolultak, bár a számítástechnika fejlődésével egyre hatékonyabbá válnak. A konvergenciájuk sincs megnyugtató módon tisztázva, ezért alkalmazásuk nem minden probléma nélküli. A paraméterbecslési eljárások összefoglaló ismertetése és értékelése részletesen megtalálható a szakirodalomban [3], [7].

A *hólabda folyamatra javasolt paraméterbecslési eljárás* előnye egyszerűségében van. Kiindulva a statisztikai modell (6) összefüggéséből az alábbi lineáris regresszió írható fel:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + u$$

ahol $y = s_t$, $x_1 = s_{t-1}$, $x_2 = (s_{t-1})^2$, u pedig a maradékváltozót jelenti.

Az ismeretlen α_1 és α_2 paraméterek a *közönséges legkisebb négyzetek* módszerével becsülhetők, majd a becsült paraméterekből meghatározhatók az eredeti folyamat paraméterei is az alábbi módon:

$$\hat{n} = \hat{\alpha}_1 - 1$$

$$N = \frac{\hat{\alpha}_1 - 1}{\hat{\alpha}_2}.$$

A folyamat harmadik paramétere (az induló állapotot kifejező paraméter) közvetlenül adott, mint a megfigyelési értékek idősorának első tagja. Megjegyzendő, hogy az induló állapot ilyen meghatározását a folyamat tisztán rekurzív jellege teszi indokolttá.

Az elmondottakat a rádióelőfizetők számának alakulását leíró függvény példáján illusztráljuk⁴. A rádióelőfizetők számának alakulása tipikus „hólabda” folyamat, hiszen jól definiálható felső korláttal rendelkezik (az ország családjainak száma), és a tulajdonság (a rádiózás, mint szokás) elterjedése

⁴A példát egy korábbi, részben e témával foglalkozó tanulmányból [2] vettem át.

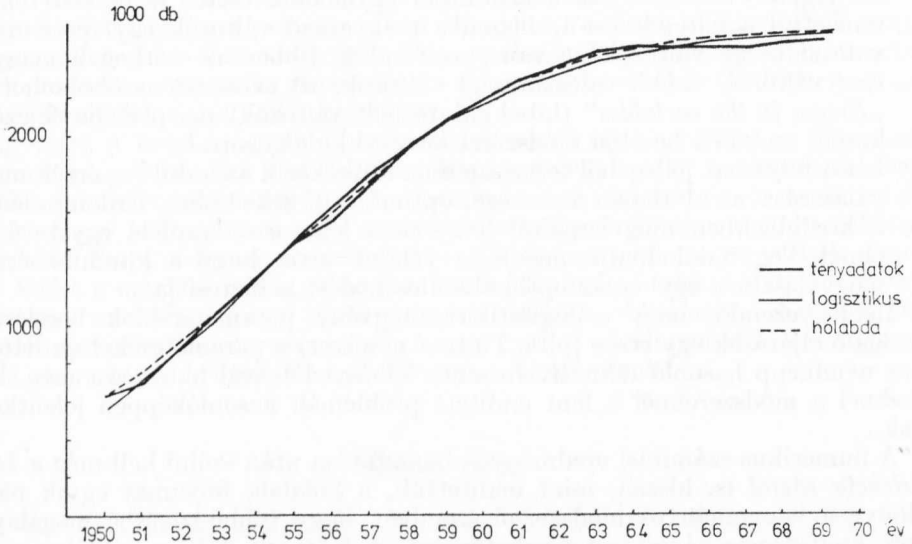
valóban a feltételezett egyszerű törvényszerűséget követi. A megfigyelési adatokat a KSH megfelelő évkönyvéből [8] vettük. A megfigyelési időszak 1950–1969-ig terjedt. A kapott regressziós függvény

$$\hat{y} = 1.2903 x_1 + 0.0001126 x_2$$

volt, ahonnan $n = 0.2903$, $N = 2578$ adódott. Tehát a folyamat

$$\hat{s}_{t+1} = 1.2903 s_t - \frac{0.2903 s_t^2}{2578}$$

alakot ölti. (A rádióelőfizetők számát 1000 db-ban fejeztük ki.) A függvény (l. 1. ábra) igen kedvező statisztikai mutatókkal rendelkezik. Az R^2 értéke 0.99 felett van, a relatív hiba pedig 1,5%, ami végső soron azt jelenti, hogy *a vizsgált jelenség lényeges jellemzői a modell feltételezéseivel nagymértékben konformak.*



1. ábra

Az összehasonlítás kedvéért bemutatjuk az ábrán az ugyanebből az idősből becsült logisztikus függvény értékeit is. Ez a függvény ugyancsak jól illeszkedik a tényadatokhoz, relatív hibája 1,49%, tehát valamivel kisebb, mint az előbbi folyamatnál. Ez feltehetően abból adódik, hogy a logisztikus függvény paramétereit meglehetősen bonyolult eljárással, minden paraméterre optimálisan becsültük [1], míg a „hólabda” s_0 paraméterét nem becsültük, hanem közvetlenül az idősből vettük át. A két becslés közötti igen csekély eltérés arra is utal, hogy az itt javasolt becslési eljárás (legalábbis ebben az esetben) lényegileg egyenértékű a lényegesen bonyolultabb és áttekinthetlenebb eljárással.

A bemutatott példa kapcsán néhány megjegyzést kell tenni. Egyrészt arra hívjuk fel a figyelmet, hogy a numerikus alkalmazáshoz alaposan meg kell fontolni, hogy *a vizsgált idősor valóban az elméleti modellben megfogalmazott viselkedést tükrözi-e.* Ha például ez esetben nem a rádióelőfizetők számát,

hanem a rádiókészülékek állományát vizsgáltuk volna, nem kaptunk volna ilyen jó eredményt, hiszen akkor a második, harmadik stb. . . készülék belépése megzavarta volna a folyamatot. Másodszer utalni kell arra, hogy még a leg gondosabb előkészítés mellett is számolni kell azzal, hogy a gazdaságban — és általában a társadalomtudományokban — *a törvényszerűségek tendenciákban érvényesülnek*, ami az eredmények más szemléletű értékelését követeli meg, mint a természettudományokban. Végül megemlítjük, hogy a becslés során nem vettünk figyelembe egy sor olyan, ökonometria jellegű problémát, mely a becslési eredményeket ronthatja, így

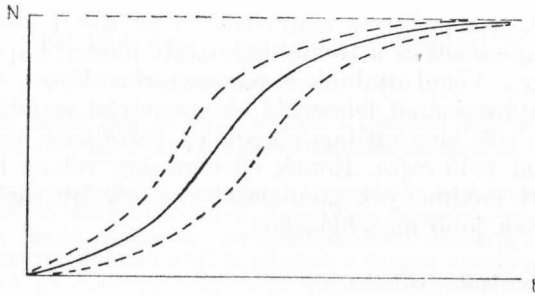
- a magyarázó változók s_i és s_i^2 közt káros multikollinearitás léphet fel és ez jelentősen ronthatja a becslések értékét;
- a hagyományos ökonometria általában feltételezi, hogy a magyarázó változók nem valószínűségi változók, és ilyen feltételek mellett ad idősoros modellek esetén a közönséges legkisebb négyzetek módszere konzisztens, sőt legjobb becslést. Már a szimultán egyenletek esetén is el kell tekinteni ettől a feltételezéstől, hiszen a magyarázó változók egy része ott is valószínűségi változóként van specifikálva. Ebben az esetben a magyarázó változók valódi valószínűségi változók, ezt az esetet az ökonometria „*Errors in the variables*” (hibákkal terhelt változók) c. speciális fejezete kezeli, melynek becslési módszerei kevésbé kidolgozottak;
- bár a folyamat jellegéből természetesen következik az induló s_0 érték megválasztása, az általában nem vezet optimális illeszkedéshez. Érdeemes lenne a későbbiekben megvizsgálni, hogy nem lehet-e valamiféle egyszerű és elméletileg indokolható megoldást találni arra, hogy a kiinduló érték megválasztása egyben az optimális illeszkedést is biztosítja.

Megjegyezendő, hogy a logisztikus függvény paramétereinek becslésére szolgáló eljárások egy része (pl. a Tintner-módszer) a paramétereket az ittenihez némiképp hasonló rekurzív formula felhasználásával határozza meg. Így ezeknél a módszereknél a fent említett problémák hasonlóképpen jelentkeznek.

A numerikus számítási eredmények bemutatása után szólni kell még a *konfidencia sávról* is, hiszen, mint említettük, a hólabda folyamat egyik nagy előnye a hagyományos módszerrel szemben, hogy többé-kevésbé megalapozott konfidencia sávot tud meghatározni. A fenti numerikus példához nem tudunk konfidencia sávot mellékelni, mivel a statikus modell esetére ($n \neq 1$), nincs elég jól működő gépi szimulációs program. Ehelyett egyszerű esetben, fiktív példán szimuláltuk az alapmodell konfidencia sávjának alakulását.⁵

Helyhiány miatt nem mutatjuk itt be ezeket a számításokat, csupán egy fontos tapasztalatra utalunk. Nevezetesen arra, hogy a konfidencia sáv a folyamat középső értékeinél a legnagyobb, az alsó és felső szakasznál egyre szűkebb, azaz a folyamat „szabadsága” a középső szakaszon a legnagyobb. Ez érdekes és fontos eredménynek tűnik, hiszen azt jelenti, hogy a sáv nem monoton tágul, hanem egy idő után összeszűkül, azaz az *előrebecslés* a folyamat felső szakaszában lényegesen megbízhatóbb lesz, mint a középső szakaszon, sőt *megbízhatósága az inflexió ponton túl egyre növekszik*. A konfidencia-sáv alakját a 2. ábra jellemzi.

⁵ A gépi számításokat F. Liska Tibor végezte, akinek ezért, valamint a megoldás egyéb részeihez kapcsolódó ötleteiért köszönettel tartozom.



2. ábra

A konfidenciasáv említett tulajdonsága egyébként logikailag könnyen belátható. A konfidenciasáv az előrebecsült értékeken is megkonstruálható. Így az előrejelzés jobban értelmezhető és megbízhatóbb lesz, mint a hagyományos előrejelzések, amelyek „csak” az előrejelzés várható hibáját adják meg, vagy esetleges konfidenciasávjuk erős, gyakorlatilag nehezen értelmezhető feltételezéseken nyugszik. A becslés és előrebecslés kérdését röviden még az alábbi két gondolattal kell kiegészíteni.

Egyrészt a konfidenciasáv megkonstruálásánál eltekintettünk a folyamat paramétereinek becslési hibáitól. Elméleti konfidenciasávot határoztunk meg azzal a feltételezéssel, hogy a paraméterek értékei rögzítettek. A másik megjegyzés azzal kapcsolatos, hogy a paraméterbecslések hibája nagy mértékben függ attól, hogy a folyamat mekkora szakaszát ismerjük megfigyelések alapján. Ezért a megbízhatóság vizsgálatakor a jelenleginél sokkal mélyebb elemzésre van szükség, s úgy tűnik, hogy a hólabda folyamattal kapcsolatos további kutatásokkal éppen ezeket a — gyakorlati alkalmazhatóság szempontjából alapvető — kérdéseket kell elsősorban tisztáznunk.

Összefoglalás

A cikk témája egy eléggé általános terjedési folyamat vizsgálata és elemzése a matematikai modell kidolgozásától egészen a numerikus számítás problémájáig. A folyamat ilyen vizsgálata némiképp eltér a hagyományos ökonometriai kezeléstől, s ezáltal egy sor érdekes eredménnyel járul hozzá a folyamat megismeréséhez és alkalmazásához.

A folyamatot alapvetően nem időfüggvényként, hanem *rekurzív sémaként fogtuk fel*. Ennek nagy előnye, hogy megmagyarázza a folyamat alakulását akkor is, ha a konkrét időbeli lefutásban valami külső ok miatt zavaró hatás lépne be. A folyamat valószínűségszámítású megalapozása lehetőséget adott *konfidenciasáv* meghatározására, amely elsősorban az előrejelzések készítésénél játszik fontos szerepet. A rekurzív felírás további előnye az is, hogy *a paraméterek becslésére igen egyszerű eljárás volt konstruálható*. Végül eredménynek tekinthető az is, más hasonló folyamatok hasonló szellemű megoldásához is irányt mutathat ez az elemzés.

A feladat megoldása során természetesen *megoldatlan problémák* is maradtak. Nem foglalkoztunk a folyamattal $n > 1$ esetén (ennek értelmezése sincs teljesen tisztázva). Nem sikerült explicit megoldást találni sem a továbbfej-

lesztett feladatra, sem pedig az alapfeladatra az $n \neq 1$ esetben⁶. Nincs még becslésünk az alapmodell és a továbbfejlesztett modell kapcsolatát meghatározó maradéktagra. Végül utalunk a paraméterbecslésnél, valamint a konfidenciasáv meghatározásánál felmerülő ökonometriai problémákra. A feladat így meglehetősen sok heurisztikus megoldást tartalmaz, matematikailag sok még a kiegészíteni való rajta. Ennek ellenére úgy vélem, hogy a kérdés felvetése és az elért eredmények gondolatébresztőek lehetnek az ökonometria egyes problémáinak jobb megoldásához.

(*Beérkezett: 1977. október 13.-án.*)

IRODALOMJEGYZÉK

1. ÉLTETŐ Ö.—HUNYADI L.: On the Estimation of the Parameters of the Logistic Function. Paper presented to the European Meeting of the Econometric Society, Budapest, 5—8 September 1972. 12 p.
2. HUNYADI L.: Folytonos és diszkrét gazdasági modellek összehasonlító elemzése (Doktori értekezés). Budapest, 1974. 90 o.
3. KREKÓ B.—THEISZ É.—PÁRNICZKI L.—PINTÉR S.: Korreláció és trendszámítás. Közgazdasági és Jogi Könyvkiadó, Budapest, 1958. 318 o.
4. LANGE, O.: Введение в эконометрию. Прогресс. Москва, 1964. 294 o.
5. PRÉKOPA A.: Valószínűségelmélet. Műszaki Könyvkiadó, Budapest, 1974. 440 o.
6. PRÉKOPA A.—ÉLTETŐ Ö.: Matematikai jegyzetek IV. Matematikai Statisztika (kézirat). Statisztikai Kiadó, Budapest, 1961. 127 o.
7. Az ártervezés ökonometriai modelljének munkanyagai VII.: Növekedési függvények elemzése. Országos Anyag- és Árhivatal és INFELŐR Rendszerttechnikai V., Budapest, 1971. 51 o.
8. Statisztikai Évkönyv 1970. Központi Statisztikai Hivatal, Budapest, 1971. 563 o.

ANALYSIS OF A SPREADING PROCESS

(The snowball model)

Starting from the basic idea of „snowball” game the article examines a spreading process suitable for the analysis of various economic and biological problems.

In Part One the theoretical model is formulated in two variants. The theoretical model relies on the assumption that there is a given closed medium containing a finite number of elements where a marked property is spreading among the elements subject to simple laws. The basic model (first variant) assumes that this property is spreading discontinuously with a lag, while the improved model (second variant) reckons with continuous spreading. In Part One of the article beside reviewing the models also their difference equation is presented and the characteristics of the time-path of the process are analyzed.

Part Two deals with the practical application of the models. Firstly theoretical models are transformed in such a way that they become suitable for numerical computation. The close connection between the „snowball” process and the logistic function is demonstrated. As a consequence it is pointed out that the spreading process examined can be used for the analysis of the time-paths of contagious diseases, fashion or consumption patterns as well. Finally, a simple procedure is presented by which parameters of the process can be estimated in possession of empirical data, furthermore, important statements are made concerning the shape of the confidence interval too. The article presents an illustrative numerical problem describing the development of the number of radio subscribers.

⁶ Egy speciális esetre Danes István készített szellemes megoldást, melyet azonban itt nem tudunk közölni.

АНАЛИЗ ПРОЦЕССА РАСПРОСТРАНЕНИЯ

(Модель снежного кома)

В рассматриваемой статье, исходя из идеи игры в снежный ком рассматривается процесс распространения, пригодный для анализа различных экономических и биологических проблем.

В первой части в двух вариантах дается формулировка теоретической модели. Теоретическая модель исходит из той предпосылки, что налицо замкнутая система, включающая в себя определенное число элементов, распространяется какое-то особое свойство по простым закономерностям. Базовая модель (первый вариант) предполагает, что это свойство распространяется по-этапно, с некоторым замедлением, в то время как усовершенствованная модель (второй вариант) считается с непрерывным процессом распространения. В первой части статьи наряду с изложением самых моделей приводятся также и их дифференциальные уравнения и с их помощью анализируются характерные черты протекания этого процесса во времени.

Во второй части рассматриваются вопросы конкретного использования моделей. В-первых, теоретические модели перестраиваются так, чтобы они были пригодны для выполнения численных расчетов. Указывается на тесную связь между процессом «снежного кома» и логической функции. В силу этого указывается также и то, что изучаемый процесс распространения может быть использован применительно, например, для изучения распространения развития во времени инфекционных заболеваний, мода или потребительских привычек. В заключении приводится простой метод, с помощью которого на основании эмпирических данных станет возможным оценивать параметры процесса а также несколько важных замечаний относительно формы доверительного интервала. В конкретном разделе статьи, рассматривающей практическое применение, все изложенное проиллюстрировано на численном примере формирования численности радиослушателей (подписчиков).