HABLICSEK LÁSZLÓ

Korspecifikus termékenységi arányszámok modellezése

Azokban a tudományágakban, ahol a vizsgálatok tárgyát képező jelenségek nem ismételhetők, a fizikai értelemben vett kísérletezés kizárt, állandó problémát jelent a jelenségek megvalósulásának modellezése, az empirikus adatok matematikai jellegű leírása.

Ez a demográfiában sincs másképp, annak ellenére, hogy általában bőséges népesedés-statisztikai adatok állnak rendelkezésre és a népességtudomány axiómarendszerét módszertanát és számítási eljárásait részletesen kidolgozták. Az alapprobléma az, hogy a társadalmi-gazdasági-demográfiai tényezők a jelenségek megvalósulását igen összetett, bonyolult módon határozzák meg, és a meghatározó formációk sokaságához képest a rendelkezésre álló konkrét realizációk száma kevés.

A tanulmányban demográfiai megoszlások modellezési lehetőségeit vizsgálom. A népmozgalom alapjelenségei, a *termékenység, a házassági és a vándormozgalom*, valamint a *halandóság*, mint az ember *életútjához kötött* jelenségek számos megoszlást produkálnak, például életkor szerint. Diszkrét megközelítésben tekinthetjük azt, mekkora valószínűséggel ad gyermeknek életet egy nő x és $(x + 1)$-éves kora között (f_x), milyen valószínűséggel köti meg első házasságát x évesen (n_x), mekkora „esélye" van arra, hogy x-dik és $(x + 1)$-dik születésnapja között meghaljon (q_x). A felsoroltak nem a valószínűségszámítás értelmében vett valószínűségeloszlások, hiszen összegük általában nem egy. Sorozatokról, illetve folytonos megközelítésben függvényekről van szó, melyek az adott jelenség életkor szerinti megvalósulását leírják.

A folytonos megközelítés az f, n, q függvényekkel operál, melyek az x változó (életkor) folytonos függvényei. Ebben az esetben az $(a, a + \Delta a)$ intervallumban az elhalálozás valószínűsége közelítőleg $q(a)\,\Delta a$-val egyenlő, vagyis valószínűségszámítási analógiával q „sűrűségfüggvény". Azt fogjuk mondani, hogy $q(a)$ az a életkorbeli elhalálozás valószínűsége.

Az empirikus adatok matematikai leírása, a valószínűségeket megadó függvénytípus keresése régóta tárgya demográfusok és matematikusok kutatásainak, jelentősége a népesedési folyamatok jobb megismerésében nyilvánvaló. Az adott jelenség megvalósulását leíró függvény

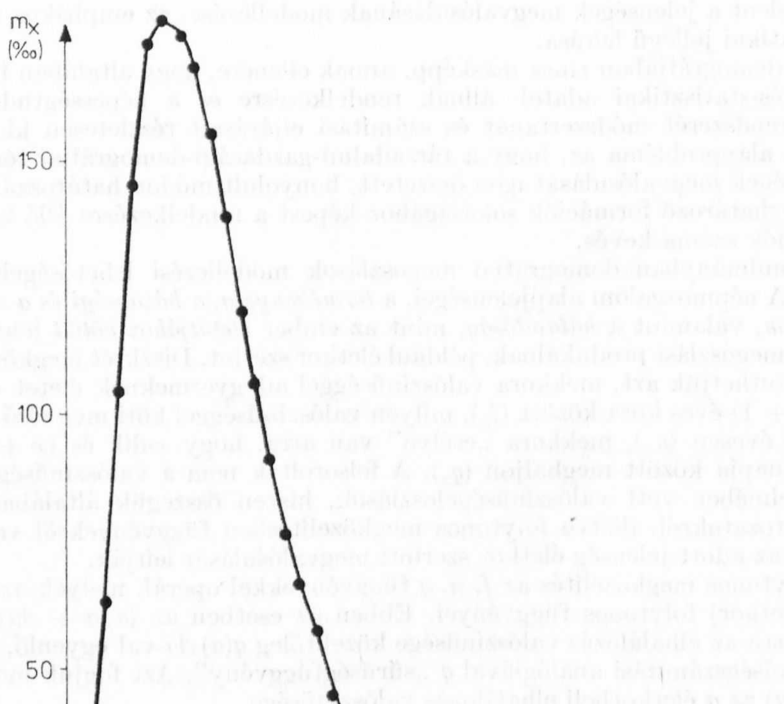
- megadja a jelenség lefolyását néhány paraméterrel;
- lehetővé teszi jelenségek kapcsolatának matematikai vizsgálatát;
- segítségével megbízhatóbb előrejelzés adható a jövőbeni alakulásra;
- vizsgálata közelebb visz a jelenség lényegének feltárásához.

A tanulmány annak a kutatásnak matematikai vonatkozású részeredményeit tartalmazza, melyet a KSH Népességtudományi Kutató Intézetben folytatunk népmozgalmi jelenségek modellezésével kapcsolatban.

A termékenységi függvény tulajdonságai

A következőkben a *korspecifikus termékenységi arányszámok* függvényszerű leírásával foglalkozunk, ami egy x éves nőre – adott naptári évben, vagy másik szemléletmód szerint x éves korában – jutó élveszületések száma (m_x). Élveszületési valószínűségek helyett azért célszerű arányszámokat tekinteni, mert a népmozgalmi statisztikában közvetlenül hozzáférhetőek és előreszámítási célokra is alkalmasabbak. A továbbiakban az m_x, $x = 15, 16, \ldots, 49$ adatsort termékenységi arányszámoknak, a megfelelő folytonos függvényt termékenységi függvénynek nevezzük.

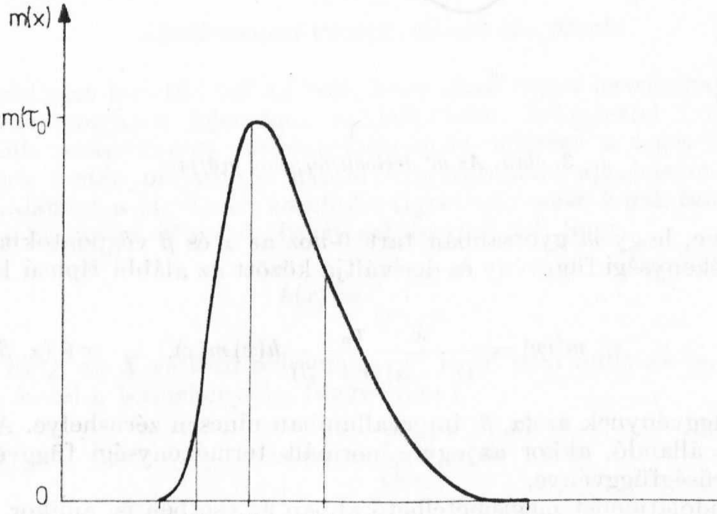
A termékenységi függvény lényeges külső jegyeinek leírására függvényelemzési módszer alkalmas. Bár itt csak a termékenységgel foglalkozunk, az alkalmazott gondolatmenet más jelenségekre is analóg módon alkalmazható.



1. ábra. Élveszületési arányszámok az anya kora szerint, 1978.
Forrás: Demográfiai Évkönyv, 1978

Az újabb magyar termékenységi arányszámok jellegzetes egymóduszú, aszimmetrikus görbe diszkrét pontjai. A gyermekszülés 13–14 éves korban kezdődik, a függvény igen meredeken emelkedik körülbelül 21–22 éves korig, ahol maximumát eléri. Ezen *legvalószínűbb szülési kor* után szintén meredeken, a felfutó ághoz képest azonban kevésbé gyorsan csökken és 50 év környékén gyakorlatilag 0-vá válik (1. ábra).

Ennek alapján az m termékenységi függvény egy (α, β) intervallumban, a nők úgynevezett propagatív periódusán értelmezett pozitív függvény. Az értelmezési tartomány határpontjaiban egyoldali limesszel rendelkezik: $m(\alpha + 0) = m(\beta - 0) = 0$. A függvények egyetlen szélsőértékhelye a τ_0 pont, ami maximumhely és $\tau_0 < \mu_1$, az átlagos szülési kor. Létezik továbbá két olyan pont, τ_1 és τ_2 $(\tau_1 < \tau_0 < \tau_2)$, melyekre:
— m konvex (α, τ_1)-ben
— konkáv a (τ_1, τ_2) szakaszon
— konvex τ_2 és β között (2. ábra).



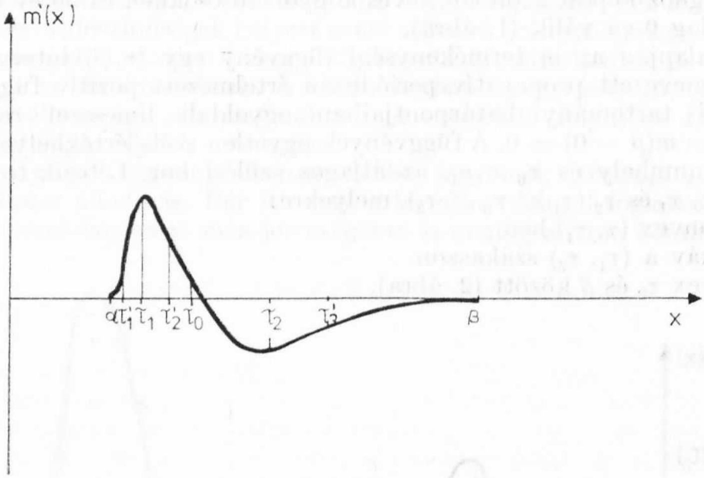
2. ábra. A termékenységi függvény gráfja

Figyelembe véve, hogy m helyettesítési értéke τ_0-ban $m(\tau_0)$ feltehetően nem függvénye a szóban forgó paramétereknek, valamint, ha ezek a paraméterek egymástól is függetlenek, a termékenységi függvény legegyszerűbb modelljének egy hatparaméteres függvény adódik:

$$m(x) = m(x, \alpha, \beta, \tau_0, \tau_1, \tau_2, m(\tau_0)). \tag{1}$$

Vizsgáljuk meg az m' deriváltfüggvény lefutását. Az m felsorolt tulajdonságaiból következik, hogy az (α, β) intervallumban m'-nek egyetlen zérushelye a τ_0 pont, τ_1-ben és τ_2-ben vannak szélsőértékhelyei (τ_1-ben maximuma, τ_2-ben minimuma). Továbbá m' értelmezési tartománya szintén (α, β) és a tapasztalat szerint $m'(\alpha + 0) = m'(\beta - 0) = 0$. A deriváltfüggvény az (α, τ_0)

intervallumban csak pozitív, a (τ_0, β) szakaszon csak negatív értékeket vesz fel. Léteznek továbbá a $\tau_1', \tau_2', \tau_3'$ pontok, melyekre m' konvex az (α, τ_1') és (τ_2', τ_3') szakaszokon és konkáv a (τ_1', τ_2'), (τ_3', β) intervallumokban (3. ábra).



3. ábra. Az m' deriváltfüggvény gráfja

Hozzávéve, hogy m gyorsabban tart 0-hoz az α és β végpontokban, mint m', a termékenységi függvény és deriváltja között az alábbi típusú kapcsolat áll fenn:[1]

$$m'(x) = \frac{x - \tau_0}{(x-\alpha)(x-\beta)} h(x) m(x), \qquad x \in (\alpha, \beta) \tag{2}$$

ahol a h függvénynek az (α, β) intervallumban nincsen zérushelye. Amennyiben $h(x) =$ állandó, akkor az egyre normált termékenységi függvény a β-eloszlás sűrűségfüggvénye.

Ez a gondolatmenet megismételhető abban az esetben is, amikor a termékenységi periódus nem véges ($\beta = +\infty$). Ekkor

$$m'(x) = \frac{x - \tau_0}{x - \alpha} h(x) m(x),\text{[2]} \tag{3}$$

amiből $h(x) =$ állandó esetén az egyre normált termékenységi függvényre a Γ-eloszlás sűrűségfüggvénye adódik. Tekintsünk még egy speciális esetet. Legyen

$$h(x) = -\frac{3}{2} \frac{x + b}{x - \alpha}, \tag{4}$$

[1] A (2) típusú egyenletnél — figyelembe véve a termékenységi függvény tulajdonságait a h függvény pozitív. Ugyanis m nem-negatív, $(x-\alpha)(x-\beta)$ negatív, $x - \tau_0$ negatív, ha $x < \tau_0$ és pozitív, ha $x > \tau_0$ és m'-nek pozitívnak kell lenni (α, τ_0)-ban, negatívnak (τ_0, β)-ban.

[2] Végtelen propagatív periódus esetén a h függvény negatív.

ekkor az ún. Hadwiger-féle eloszláshoz jutunk, melynek általános alakja — b megfelelő megválasztásával —

$$(5)\qquad m(x) = \frac{A}{(x-\alpha)^{3/2}}\exp\left[\frac{B}{x-\alpha}+\frac{x-\alpha}{C}\right]$$

Mindhárom szóban forgó, függvénytípussal jó eredményeket érték el különböző termékenységi függvények, modellezésénél ([1], [2], [3]).

A függvényvizsgálati módszerrel kapott (2) illetve (3) differenciálegyenlet az α, β, τ_0 paraméterekkel és egy h függvénnyel jellemzi a termékenységi függvényeket. Jelen pillanatban még kevés az ismeretünk ahhoz, hogy h-t meghatározhassuk. Nyitott kérdés, hogy h *levezethető-e* általános demográfiai összefüggésekből vagy pedig az egyes konkrét termékenységi függvényekhez kell becsülnünk azt.

Először h becslésével foglalkozunk.

Általánosított Pearson-típusú eloszlások

Az elsődleges kutatási cél az volt, hogy általánosan használható módszert adjunk népmozgalmi jelenségek modellezésére. Tekintettel arra, hogy az előző szakaszban kapott differenciálegyenlet-módszer a legegyszerűbb feltételezések esetén megadta a három leggyakrabban alkalmazott függvénytípust, valamint a (4) típusú közelítést figyelembe véve h-nak racionális törtfüggvény-közelítése célszerű. Tehát azt tesszük fel, hogy

$$(6)\qquad h(x) = \frac{P^h(x)}{Q^h(x)}$$

ahol P^h és Q^h az X változó polinomjai, P^h és Q^h nem nulla az (α, β) intervallumban. Ezzel a termékenységi függvényre a

$$(7)\qquad m'(x) = \frac{P(x)}{Q(x)}m(x)$$

differenciálegyenletet kapjuk, ahol P és Q polinomok, P-nek egyszeres és egyetlen gyöke (α, β)-ban τ_0, $Q(x) \neq 0$, ha $x \in (\alpha, \beta)$, továbbá véges propagatív periódus esetén $Q(\alpha) = Q(\beta) = 0$, nem végesnél $Q(\alpha) = 0$.

Amennyiben P elsőfokú, Q pedig legfeljebb másodfokú polinom, akkor a (7) differenciálegyenlet a Pearson-típusú eloszlásokat adja:

$$(8)\qquad m'(x) = \frac{x-\tau_0}{ax^2+bx+c}m(x).$$

Ismeretes, hogy a legfontosabb folytonos eloszlások, így a normális, az exponenciális, a Γ-, a β-, a Student-, a χ^2 és a Cauchy-eloszlás Pearson-típusúak (m a sűrűségfüggvény).

A (7) differenciálegyenletnek eleget tévő, valamely (α, β) értelmezési tartományú sűrűségfüggvényekkel megadott eloszlásokat *általánosított Pearson-típusú eloszlásoknak*, vagy röviden és jellegükre utalva *P/Q-eloszlásoknak* fogjuk nevezni. A (7) differenciálegyenlet egy tetszőleges megoldására a *P/Q-függvény* elnevezést használjuk ([4]).

Eddigi meggondolásaink alapján tehát célszerűnek és hatékonynak látszik népmozgalmi megoszlások közelítése P/Q-függvényekkel. Két kérdés merül fel: 1. Hogyan valósítható meg egy P/Q-függvény illesztése a gyakorlatban? 2. Megfelelnek-e a P/Q-függvények családjának tulajdonságai a népmozgalmi megoszlások közti összefüggéseknek?

Az 1. kérdéssel kapcsolatban eljárást adunk a P/Q-függvények paramétereinek becslésére, aminek következtében adott p és q fokszámok esetében kiválasztható a legjobban illeszkedő P/Q-függvény. A 2. kérdésnél azt fogjuk bizonyítani, hogy a népmozgalmi megoszlások bizonyos kapcsolatba hozhatók ilyen típusú függvényekkel.

P/Q-függvények illesztése

Adott p és q fokszámok esetében vizsgáljuk egy P/Q-függvény paramétereinek becslését. Legyen $\tau \in (\alpha, \beta)$ és adjuk meg az $m(\tau)$ értéket. Ekkor a (7) differenciálegyenlet egyértelműen megoldható és

$$m(x) = m(\tau) \exp \int_{\tau}^{x} \frac{P(y)}{Q(y)} dy. \tag{9}$$

Ha adottak az m_x, $x = 15, 16, \ldots, 49$ arányszámok, akkor az elméleti m függvényt közelítő P/Q-függvény becsléséhez a P és Q polinomok együtthatóit és egy $\hat{m}(\tau)$ értéket kell „legjobban" megválasztani. Legyen

$$P(x) = \sum_{i=0}^{p} a_i x^i \quad \text{és} \quad Q(x) = \sum_{i=0}^{q} b_i x^i, \tag{10}$$

ahol feltételezhetjük, hogy $a_p = 1$. Mivel a Q polinom gyökei az α és a β értékek és ebben az esetben a (7) differenciálegyenletnek nincs értelme, a becsléseket valamely (α_1, β_1) intervallumban végezzük el, ahol $\alpha < \alpha_1 < \beta_1 < \beta$. A gyakorlatban $\alpha_1 = 15$ és $\beta_1 = 50$ megfelelő választásnak mutatkozik. Mivel $Q(x) \neq 0$, ha $x \in (\alpha_1, \beta_1)$, ezért a (7) differenciálegyenletből a k nemnegatív egész értékekre a

$$x^k Q(x) m'(x) = x^k P(x) m(x) \tag{11}$$

összefüggések kaphatók. Vezessük be az m függvénynek az intervallumra vonatkozó „momentumait":

$$M_i(\alpha_1, \beta_1) = \int_{\alpha_1}^{\beta_1} x^i m(x) dx. \tag{12}$$

Integráljuk a (11) egyenletek mindkét oldalát az (α_1, β_1) intervallumban, a bal oldalakat parciálisan. Ekkor k különböző értékeire fennállnak a

$$m(\beta_1)\beta_1^k Q(\beta_1) - m(\alpha_1)\alpha_1^k Q(\alpha_1) - \sum_{i=0}^{q} b_i(i+k) M_{i+k-1}(\alpha_1, \beta_1) = \sum_{i=0}^{p} a_i M_{i+k}(\alpha_1, \beta_1) \tag{13}$$

egyenlőségek, ahol M_{-1} (α_1, β_1)-et 0-nak értelmezzük ([5]).

A (13) egyenletrendszer a P és Q polinomok együtthatóiban lineáris. Amennyiben ismertek az $M_i(\alpha_1, \beta_1)$, $i = 0, 1, \ldots, p + q + 1 + \max(p, q)$ momentumok, valamint az $m(\alpha_1)$, $m(\beta_1)$ értékek, akkor a P és Q polinomok együtthatói meghatározhatók. A gyakorlatban az $M_i(\alpha_1, \beta_1)$ elméleti momentumokat ($\alpha_1 = 15$, $\beta_1 = 50$ esetén) a

$$(14) \qquad \widehat{M}_i(15, 49) = \sum_{x=15}^{49} (x + 0{,}5)^i m_x$$

tapasztalati értékekkel helyettesítjük, az $m(15)$ és $m(50)$ értékeket pedig valamilyen meggondolás, interpoláció stb. útján vehetjük fel. A termékenység esetében például $m(15) = m_{15}/2$ és $m(50) = 0{,}0$ jó becslésnek látszik. Felmerül a kérdés, $m(\alpha_1)$-nek és $m(\beta_1)$-nek a gyakorlatban komoly problémát okozó becslését, hogyan lehetne kiküszöbölni.

A (13) egyenletből szemeljük ki a k-dik, $(k + 1)$-dik és $(k + 2)$-dik egyenletet (ami egyben jelentse a k, $k + 1$ és $k + 2$ értékekhez tartozó egyenleteket). A tömörebb írásmód kedvéért vezessük be az

$$A_Q^k = \sum_{i=0}^{q} b_i(i + k) M_{i+k-1}(\alpha_1, \beta_1) \quad \text{és a} \quad B_P^k = \sum_{i=0}^{p} a_i M_{i+k}(\alpha_1, \beta_1)$$

jelöléseket. Ekkor a (13) egyenletrendszer így módosul:

$$(16) \qquad m(\beta_1)\beta_1^k Q(\beta_1) - m(\alpha_1)\alpha_1^k Q(\alpha_1) = A_Q^k + B_P^k.$$

ahol $k = 0, 1, \ldots$

A $(k + 1)$-dik egyenletet osszuk el β_1-el és vonjuk ki a k értékhez tartozóból. Ekkor az

$$-\alpha_1^k \left(1 - \frac{\alpha_1}{\beta_1}\right) m(\alpha_1)Q(\alpha_1) = A_Q^k + B_P^k - \frac{1}{\beta_1}(A_Q^{k+1} + B_P^{k+1})$$

összefüggést kapjuk. Ugyanez k helyett $(k + 1)$-el:

$$-\alpha_1^{k+1} \left(1 - \frac{\alpha_1}{\beta_1}\right) m(\alpha_1)Q(\alpha_1) = A_Q^{k+1} + B_P^{k+1} - \frac{1}{\beta_1}(A_Q^{k+2} + B_P^{k+2})$$

A második egyenletet osszuk el α_1-el, vonjuk ki az elsőből, ekkor a két egyenletből megkapjuk a

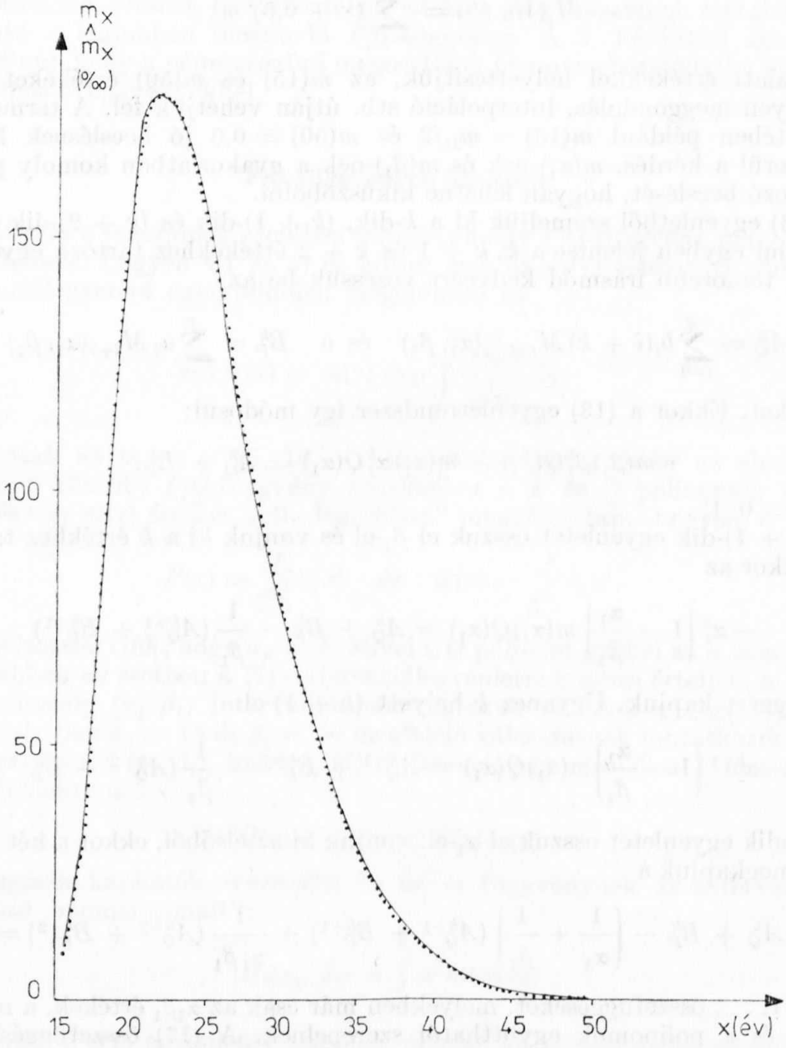
$$(17) \qquad A_Q^k + B_P^k - \left(\frac{1}{\alpha_1} + \frac{1}{\beta_1}\right)(A_Q^{k+1} + B_P^{k+1}) + \frac{1}{\alpha_1\beta_1}(A_Q^{k+2} + B_P^{k+2}) = 0,$$

$k = 0, 1, \ldots$ összefüggéseket, melyekben már csak az $\alpha_1 \beta_1$ értékek, a momentumok és a polinomok együtthatói szerepelnek. A (17) összefüggésben az $M_i(\alpha_1, \beta_1) \approx \widehat{M}_i(\alpha_1, \beta_1)$ helyettesítéseket végrehajtva végül a paramétereket a

$$\sum_{i=0}^{q} b_i \left((i + k)\widehat{M}_{i+k-1} - (i + k + 1)\left(\frac{1}{\alpha_1} + \frac{1}{\beta_1}\right)\widehat{M}_{i+k} + \frac{1}{\alpha_1\beta_1}(i + k + 2)\widehat{M}_{i+k+1}\right) +$$

$$(18) \qquad + \sum_{i=0}^{p} a_i \left(\widehat{M}_{i+k} - \left(\frac{1}{\alpha_1} + \frac{1}{\beta_1}\right)\widehat{M}_{i+k+1} + \frac{1}{\alpha_1\beta_1}\widehat{M}_{i+k+2}\right) = 0$$

egyenletekből számíthatjuk, ahol $k = 0, 1, \ldots, (p+q)$. A kapott $p+q+1$ egyenlet $p+q+1$ ismeretlent tartalmaz ($a_p = 1$), és a gyakorlatban általában egyértelműen megoldható. A P és Q polinomok együtthatóinak ismeretében numerikus módszerrel meghatározhatók a polinomok zérushelyei, valamint a közelítő értékek.



4. ábra. P/Q-függvény illesztése az 1978. évi korspecifikus termékenységi arányszámokhoz ($p = 1, q = 4$)

P/Q-függvények illesztésére a (18) egyenletrendszert használva számítógépes programot készítettünk FORTRAN nyelven, amely a KSH Számítóközpontjában futott.

A program alapján először a momentumok becslésére, majd a (18) egyenletrendszer megoldására kerül sor. A kapott együtthatók alapján numerikus eljárás becsli a P és Q polinomok gyökeit, valamint az

$$\int_{\alpha_1}^{x} \frac{P(y)}{Q(y)} dy$$

(x egész) integrálokat. Mivel a termékenységi arányszámok *összege*, az ún. *teljes termékenység* fontos mutató a demográfiában, az m_x arányszámokhoz úgy illesztettük az

$$\hat{m}_x^0 = \exp\left(\int_{\alpha_1}^{x} \frac{P(y)}{Q(y)} dy\right)$$

becsléseket, hogy a szorzótényező egyúttal a becsült *teljes termékenység* legyen. Ehhez egy σ szorzót határoztunk meg a

$$(19) \qquad \sum_{x=15}^{49} \left(m_x - \sigma \frac{\hat{m}_x^0}{\sum\limits_{y=15}^{49} \hat{m}_y^0} \right)^2 = \min.$$

feltétellel. Az illeszkedés jóságát az eltérések négyzetösszegével és a két adatsor közti korrelációval mértük.

Az 1. táblázat és a 4. ábra az 1978. évi korspecifikus termékenységi arányszámokra a $p = 1$ és $q = 4$ fokszámú P/Q-függvény illesztésének eredményeit tartalmazza.

Az eredmények azt mutatják, hogy a $p = 1$, $q = 4$ fokszámú P/Q-függvények családja minimális hibával írja le az 1978. évi termékenységi arányszámokat. Lényegében az illeszkedés olyan jó, hogy az eltérések okának az empirikus adatok véletlen ingadozását is tekinthetjük.

Összehasonlítás végett megjegyezzük, hogy az 1961. évi korspecifikus termékenységi arányszámok Γ-eloszláson alapuló modellje esetében a négyzetes eltérésre 1491,6 adódott ([1]), míg a $p = 1$, $q = 4$ fokszámú P/Q-függvény esetében 139,98, azaz utóbbi több mint 10-szer jobb illeszkedést produkál.

P/Q-függvények tulajdonságai

Jelöljük $\mathfrak{M}$-el azon P/Q-függvények családját, melyeknek értelmezési tartománya $(\alpha, \beta) \subset (0, +\infty)$, és értelmezési tartományuk minden pontjában pozitív értéket vesznek fel. Az $\mathfrak{M}$ függvényosztály a függvényszorzásra nézve csoportot alkot.

Valóban ha f és $g \in \mathfrak{M}$, akkor $f \cdot g$ értelmezési tartománya is részhalmaza $(0, +\infty)$-nek és $f \cdot g$ pozitív. Ha f-et a P^f és Q^f, g-t a P^g és Q^g polinomok határozzák meg, akkor

$$(fg)' = f'g + fg' = \left(\frac{P^f}{Q^f} + \frac{P^g}{Q^g}\right)(fg)$$

1. táblázat

P/Q függvény illesztése az 1978. évi korspecifikus termékenységi arányszámokhoz
$(p = 1, q = 4)$

A $P(x)$ polinom együtthatói: $a_0 = -21{,}84$ $a_1 = 1{,}0$
A $Q(x)$ polinom együtthatói: $b_0 = -68{,}67$ $b_1 = 10{,}2$ $b_2 = -0{,}435$
$b_3 = 0{,}00176$ $b_4 = 0{,}0000723$
A P polinom gyöke (módusz): 21,84
A Q polinom gyökei: 13,91*, 48,14**

Kor	Empirikus	Becsült
	arányszámok (‰)	
15	10,6	8,5
16	29,8	29,9
17	62,9	67,0
18	104,7	109,3
19	145,0	145,0
20	173,3	168,0
21	177,7	177,4
22	175,0	176,1
23	168,7	167,2
24	154,9	153,8
25	138,3	138,2
26	119.7	121,9
27	105,9	106,1
28	90,4	91,1
29	75,3	77,5
30	66,0	65,3
31	56,6	54,5
32	43,7	45,1
33	37,2	37,0
34	31,1	30,0
35	23,3	24,1
36	19,7	19,1
37	14,2	14,9
38	11,5	11,5
39	8,9	8,6
40	5,9	6,3
41	4,7	4,5
42	3,2	3,0
43	2,0	1,9
44	0,9	1,1
45	0,4	0,5
46	0,2	0,3
47	0,1	0,0
48	0,0	0,0
49	0,0	0,0

Négyzetes eltérés: 95,7‰²
Korreláció: 0,9996

* 13,91 valójában Q szélsőértékhelye, de $Q(13{,}91)$ közel 0.
** A $\beta = 48{,}14$ érték mutatja, hogy a $\beta_1 = 50$ érték magas.

miatt az $f \cdot g$ függvényre (7) típusú differenciálegyenlet áll fenn, tehát $fg \in \mathfrak{M}$. Hasonlóan igazolható, hogy ha $f, g \in \mathfrak{M}$, akkor $f/g \in \mathfrak{M}$. Az azonosan 1 függvény nyilván $\mathfrak{M}$ eleme, és egyben egységelem. Az f függvény inverze az $\mathfrak{M}$ csoportban $1/f$.

A P/Q-függvények csoporttulajdonságának alapvető demográfiai párhuzama van. A demográfia egyik alapfelvetése szerint a népmozgalmi jelenségek *függetlenek*. Ez az ún. *függetlenségi hipotézis* azt jelenti, hogy a jelenségek hatása izolálható. Ha A és B két jelenség, akkor létezik A-nak és B-nek $q_A(x)$ és $q_B(x)$ x-beni *megvalósulási* valószínűsége, azzal a feltételezéssel, hogy a másik jelenséget teljesen kiküszöböljük. Ekkor a *független*, $p_A(x) = 1 - q_A(x)$, $p_B(x) = 1 - q_B(x)$ *megmaradási* valószínűségek[3] szorzata a valóságban tapasztalt megmaradási valószínűséggel egyenlő.

$$p_A(x) \cdot p_B(x) = p_{AB}(x), \tag{20}$$

vagyis a megmaradási valószínűségek halmaza zárt a szorzásra nézve.

A P/Q-függvények csoporttulajdonsága és a népmozgalmi jelenségek megmaradási valószínűségeinek szorzattulajdonsága arra utal, hogy P/Q-függvények a népmozgalmi megoszlások egészével kapcsolatba hozhatók.

Összefoglalva az eddigieket, a modellezésnél a függvényvizsgálati módszerrel indulva egy differenciálegyenlethez jutottunk. A (2) illetve (3) differenciálegyenletekben az ismeretlen h függvény legegyszerűbb közelítései megadták a szakirodalom által javasolt három legfontosabb modellt. A h függvény közelítése racionális törtfüggvénnyel célszerű, ennek alapján a termékenységi arányszámokat P/Q-függvénnyel becsültük. A P/Q-függvények illesztése számítógép igénybevételével lényegében problémamentesen megoldható. Az eredmények azt mutatták, hogy plusz három paraméter bevezetése 10-szeres illeszkedésjavulást eredményez, és lényegében az alkalmazott függvény elméleti modellnek tekinthető. Ezenkívül a P/Q-függvények struktúrája, tulajdonságai analógiát mutatnak a népmozgalmi megoszlások tulajdonságaival.

A megoszlás külső jegyeinek elemzése mellett a jelenség természetére vonatkozó feltételezések alapján is megkísérelhetjük a modellalkotást. A következőkben ezzel foglalkozunk.

A termékenység autoregresszív modellje

A termékenységnek, mint demográfiai jelenségnek megvalósulását általánosságban három faktor befolyásolja.

1. természetes termékenység
2. demográfiai tényezők
3. társadalmi-gazdasági hatások.

A három befolyásoló tényező közül az első kettő *funkcionális* jellegű, azaz meghatározott nagyságukhoz a termékenységi arányszámok meghatározott nagysága tartozik. A társadalmi-gazdasági körülmények *adaptív* jellegűek, tehát „sztochasztikusan" befolyásolják a termékenységet. Ennek alapján a következő modellt vehetjük fel:

Legyen (Ω, A, P) valószínűségi mező, ahol Ω a propagatív korú nők halmaza. Jelölje adott $\omega \in \Omega$ esetén $\xi_t(\omega)$, $\eta_t(\omega)$, $\zeta_t(\omega)$, $\alpha_t(\omega)$ rendre a t éves korban az ω nő által szült gyermekek számát, a természetes termékenység hatá-

[3] Tehát annak valószínűsége, hogy az A illetve a B jelenség nem valósul meg.

sát, a demográfiai tényezőket és a társadalmi-gazdasági faktort. Ekkor

(21) $$\xi_t = \int_0^t [a(\tau)\eta_{t-\tau} + b(\tau)\zeta_{t-\tau}]d\tau + \int_0^t c_\tau d\alpha_\tau$$

írható homogén megközelítésben, ahol a és b determinisztikus függvények, c_τ pedig valószínűségi változó. Kézenfekvő felvenni, hogy a természetes termékenység és a demográfiai tényezők a megelőző $\xi_{t-\tau}$ értékben összpontosulnak és hogy a nők egészére nézve a „maradék" társadalmi-gazdasági hatás várható értéke zérus körül van.[4] Emiatt a

(22) $$\xi_t = \int_\alpha^t A(t-\tau)\xi_\tau d\tau + \int_0^t B_\tau d\alpha_\tau$$

előállítás lesz érvényes, ahol $M(\int_0^t B_\tau d\alpha_\tau) = 0$. A (22) összefüggésben a determinisztikus részre azt lehet mondani, hogy $A(\tau)$ már igen kicsi, ha $(t-\tau)$ nagy, vagyis, hogy ξ_t kielégítően megadható néhány megelőző érték lineáris kombinációjával. Emiatt — feltéve, hogy egy konstans p idő elteltével már a megelőző ξ_t értékek hatása elenyésző vagyis $A(t-\tau) \approx 0$, ha $t-\tau \geqq p$,

(23) $$\xi_t = \int_{t-p}^t A(t-\tau)\xi_t d\tau + \int_0^t B_\tau d\alpha_\tau$$

adódik. Azaz ξ_t-t *autoregresszív módon* közelítjük. Diszkrét megközelítésben tehát ξ_x-re azt vesszük fel, hogy

(24) $$\xi_x = \sum_{i=1}^p a_i \xi_{x-i} + \varepsilon_x$$

ahol $a_1, a_2, \ldots, a_p$ valós számok és (ε_x) 0-várható értékű sztochasztikus sorozat. A (24) egyenletből várható értékre áttérve a

(25) $$m_x = \sum_{i=1}^p a_i m_{x-i}$$

összefüggés adódik, vagyis modellünk a következő: a termékenységi arányszám a megelőző p számú termékenységi arányszám lineáris kombinációja és az együtthatók nem függnek az x értéktől.

Bár az alkalmazott feltételezések, melyek a (25) egyenletekhez vezettek, eléggé „ad hoc" jellegűek, mégis a választott autoregresszív megoldás kielégítően jellemzi a termékenységi arányszámokat.

Foglalkozzunk először az együtthatók becslésével. A legkisebb négyzetek módszere alapján az $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p$ becsléseket a

(26) $$\sum_{x=p+1+\alpha}^{\beta} \left(m_x - \sum_{i=1}^p \hat{a}_i m_{x-i}\right)^2 = \min$$

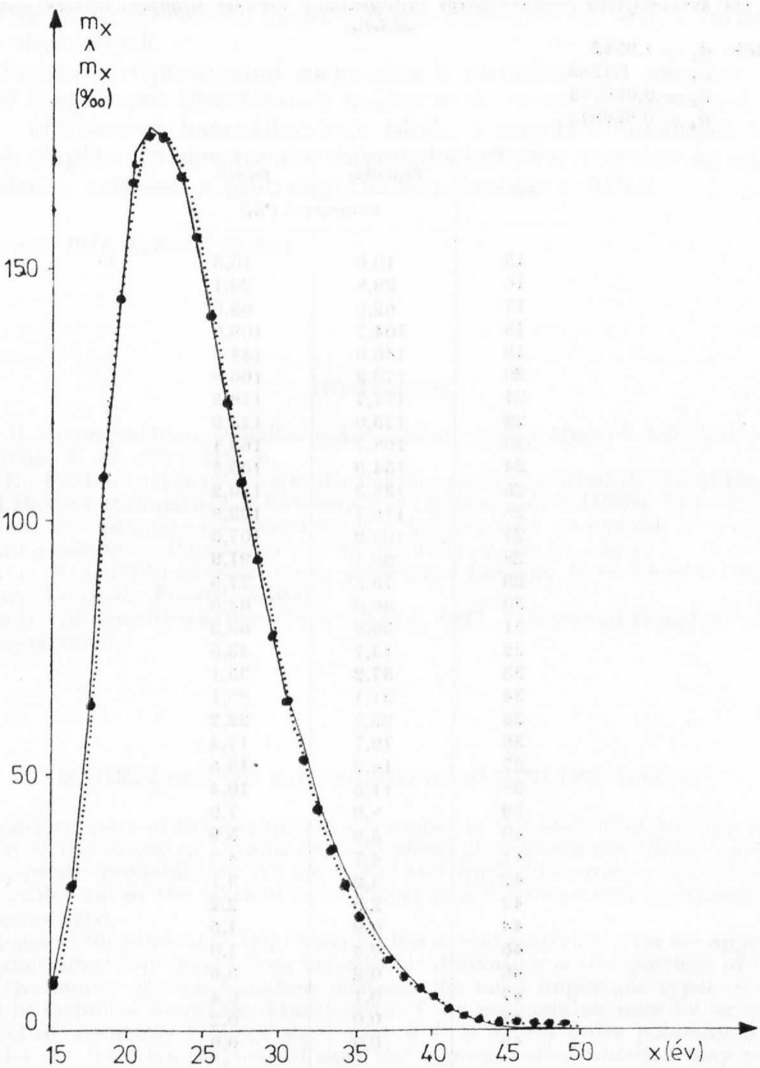
[4] „Maradék" hatáson azt értjük, amit a megelőző $t-\tau$ *életkorokhoz* tartozó $\xi_{t-\tau}$ értékek nem tartalmaznak. Ilyen értelemben mondjuk, hogy — a pillanatnyi — „maradék" hatás elenyésző.

feltétellel kaphatjuk meg, amit $\hat{a}_k$ szerint differenciálva az együtthatókra a

$$(27) \qquad \sum_{x=p+1+\alpha}^{\beta} m_x m_{x-k} - \sum_{i=1}^{p} \hat{a}_i \sum_{x=p+1+\alpha}^{\beta} m_{x-i} m_{x-k} = 0,$$

$k = 1, 2, \ldots, p$ egyenletrendszer adódik. Ennek megoldásai az $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p$ együttható-becslések.

Ezek után az $m_\alpha, m_{\alpha+1}, \ldots, m_{\alpha+p-1}$ értékekből és a kapott együtthatókból a (25) összefüggés alapján valamennyi termékenységi arányszám becsülhető.



5. ábra. Az 1978. évi korspecifikus termékenységi arányszámok becslése négyparaméteres autoregresszív modellel

Pontosabb becsléseket kaphatunk, ha nemcsak az $\hat{a}_1, \ldots, \hat{a}_p$ hanem az $\hat{m}_\alpha, \ldots, \hat{m}_{\alpha+p-1}$ kezdőértékeket is megbecsüljük a (26) feltétel alapján. Ennek az eljárásnak előnye, hogy valóban autoregresszív sort kapunk, hátránya viszont, hogy a becslési eljárás jóval bonyolultabb.

A 2. táblázat és az 5. ábra egy olyan becslési eljárás eredményeit tartalmazza amelyik az említett két lehetőség között van. Ennek során először a (27) egyenletekből megbecsültük az $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p$ együtthatókat, majd ezeket

2. *táblázat*

Az 1978. évi korspecifikus termékenységi arányszámok becslése négyparaméteres autoregresszív modellel

Együtthatók: $\hat{a}_1 = 1{,}9542$
$\hat{a}_2 = -1{,}1283$
$\hat{a}_3 = 0{,}090783$
$\hat{a}_4 = 0{,}056915$

Kor	Empirikus	Becsült
	arányszámok (‰)	
15	10,6	10,5
16	29,8	29,1
17	62,9	63,6
18	104,7	109,0
19	145,0	144,5
20	173,3	166,8
21	177,7	176,5
22	175,0	175,9
23	168,7	168,1
24	154,9	155,5
25	138,3	140,2
26	119,7	123,9
27	105,9	107,5
28	90,4	91,9
29	75,3	77,5
30	66,0	64,6
31	56,6	53,3
32	43,7	43,5
33	37,2	35,1
34	31,1	28,1
35	23,3	22,2
36	19,7	17,4
37	14,2	13,5
38	11,5	10,4
39	8,9	7,9
40	5,9	5,9
41	4,7	4,3
42	3,2	3,1
43	2,0	2,2
44	0,9	1,5
45	0,4	1,0
46	0,2	0,6
47	0,1	0,4
48	0,0	0,2
49	0,0	0,0

Korreláció: 0,9990
Négyzetes eltérés: 132,6‰²

elméleti értéknek fogadva el, a (26) feltételt átalakítva becsültük az $\hat{m}_\alpha$, $\hat{m}_{\alpha+1}$. . ., $\hat{m}_{\alpha+p-1}$ kezdő értékeket.

A 2. táblázatot az 1. táblázattal összehasonlítva látható, hogy a P/Q-függvénnyel kapott becslés csak valamivel jobb az autoregresszív modell eredményeinél és — tekintettel az autoregresszív becslési eljárásnál tett megjegyzésre — ez is abból adódhat, hogy nem tudtuk megbecsülni a legjobban illeszkedő autoregresszív sorozatot. Ezek alapján — legalábbis az 1978. évi élveszületési arányszámokra — az autoregresszív modell elméleti modellnek tekinthető.

A modell összhangban van a termékenységi folyamat azon nyilvánvalónak tűnő ismérvével, hogy a nő vagy házaspár termékenységi illetve családtörténetében a gyermek születésének esélyét a megelőző néhány év (esetünkben 4 év) termékenységtörténete meghatározza, a korábbi hatások a várható érték szintjén elmosódnak.

Összefoglalva: népmozgalmi megoszlások modellezésére mindkét módszer — a P/Q-függvények illesztésének módszere és az autoregresszivitás feltételezése — általánosan használhatónak tűnik, a kapott eredmények igen jók. Mindezek alapján érdemes tovább folytatni a kutatást, egyrészt az alkalmazások területén, másrészt a jobb elméleti megalapozás céljából.

(Beérkezett: 1981. december 20-án)

IRODALOM

1. TEKSE K.: Korspecifikus születési arányszámok demográfiai modelljeiről — *Demográfia,* 1965. 2. sz. 201—219. p.
2. GILJE, E.: Fitting curves to age-specific fertility rates *Statistical Review* of the National Central Bureau of Statistics of Sweden, Third Series, Vol. 7. (1969). 118—134. p.
3. DUCHENE, J.—GILLET—DE STEFANO, S.: Adjustement analytique des courbes de fécondité générale — *Population et Famille,* 1974. 2. sz. 53—93 p.
4. KENDALL, M. G.: *The advaced theory of statistics* London, 1948. Charles Griffin Company Limited. Fourth Edition.
5. JORDAN K.: *Matematikai statisztika.* Budapest, 1927. Athenaeum Irodalmi és Nyomdai Részvénytársulat.

MODELLING OF AGE-SPECIFIC FERTILITY RATES

An important part of demographical researches is the modelling and the functional description of the course of various demographical phenomena according to cohorts and of the age-specific probabilities. We developed and applied two methods, to describe age-specific fertility rates: the method of so called P/Q functions and an approach of the autoregressive type.

The essence of the method of P/Q functions lies in that empirical data are approximated by such continuous functions whose logarithmic derivative is the quotient of two polynomials. The family of such functions contains the most important types of functions suggested in technical literature. Coefficients of the polynomials may be estimated by the method of moments. Fittings show that if P is a first order polynomial and Q a fourth order the iterative polynomial then the approximating function may practically be regarded as a theoretical model. Besides, properties of the family of P/Q functions show some analogy to an important characteristic of demographical distributions.

The other procedure, the approach of autoregressive type starts from the assumption that at a given age the fertility of women is determined by that of a few preceding years (age). Specifically a linear relationship with constant coefficients is supposed. In the course of practical application coefficients were estimated by least squares method, then the same procedure was used for estimating initial values. In case of four years the model produces as good a fit as the method of P/Q functions.

МОДЕЛИРОВАНИЕ ВОЗРАСТНО-СПЕЦИФИЧЕСКИХ КОЭФФИЦИЭНТОВ ПЛОДОВИТОСТИ

Одной из важных глав фундаментальных исследований в области демографии является вероятностное моделирование, определяемое изложонно в виде функции протекания различных явлений по народонаселению с учетом возраста формирования возрастно-специфических возможностей реализации. В аспекте описания возроста-специфических коэффициентов плодовитости мы разработали и используем два метода, т. е. т. н. метод функций P/Q и авторегрессивный по своему типу подход.

Суть метода функций P/Q кроется в том, что к эмпирическим данным мы описываем посредством такой непрерывной функции, логарифмической производной которой является частное двух полиномов. Группа функций, обладающих такими свойствами включает в себя наиболее важные типы функций, предлагаемых в специальной литературе. Коэффициенты полиномов могут оцениваться методом моментов. Результаты показывают, что в случае полинома первой степени P и четвертой степени Q приближенная функция может рассматриваться, по существу, в качестве теоретической модели. Наряду с этим свойства семейства функций P/Q показывают аналогию с одной существенной характерной чертой дифференциации движения народоноселения.

Другой используемый метод, т. е. авторегрессивный по своему типу подход исходит из того, что в отношении какого-то определенного возраста фертильность женщин определяется фертильностью нескольких предшествующих лет (возрастов). В отношении порядка определения предлагается наличие постоянной коэффициентной линейной связи. В ходе практического использования коэффициенты оценивались методом наименьших квадратов и в последующем эта методика используется и при оценке начальных значений. В разрезе четырех лет данная модель дает такие же хорошие результаты, что и метод функций P/Q.