

## TERMÉSZETES KIVÁLASZTÁS ÉS RIEMANN-GEOMETRIA

FARKAS MIKLÓS

Felírjuk a szelekció Fisher-féle differenciálegyenletét, tárgyaljuk ennek legfontosabb tulajdonságait. Kimondjuk a Kimura-féle maximumelvet, és új metrika bevezetésével igazoljuk azt.

### 1. Bevezetés

Ha egyetlen fajt tekintünk, és változatlan körülmények között, a többi együttélő fajtól elkülönítve vizsgáljuk, a darwini természetes kiválasztás egy egyszerűsített változatát viszonylag könnyen modellezhetjük matematikailag. Feltételezzük, hogy genetikailag minden egyes egyed egy *diploid* sejtből fejlődik ki. A diploid *zigóta* két haploid ivarsejt, két *gaméta* összeolvadásából jön létre, amelyek az egyed két szülőjétől, az anyától és az apától érkeznek a reprodukció során és hordozzák a szülők genetikailag átörökíthető anyagát. A gamétákat *genome-típusokba* soroljuk; egy *i* típusú és egy *k* típusú gaméta egy *ik* típusú zigótát alkot. Az *ik* típust azonosnak tekintjük a *ki* típussal, vagyis nem különböztetjük meg a gamétákat aszerint, hogy az anyától vagy az apától érkeztek-e. A reprodukció során az *ik* típusú zigóta egy *i* és egy *k* típusú gamétát produkál, ezek azután más gamétapárt keresnek maguknak. Az egyszerűség kedvéért azonban úgy képzeljük, hogy az *ik* típusú zigóta egy ugyanolyan típusút reprodukál, és figyelmen kívül hagyjuk a mutációkat és a cross-vert (azt, hogy egyes gének véletlenszerűen kicserélődnek, illetve azt, hogy a zigótát alkotó gaméták kromoszómáin egyes szakaszok helyet cserélhetnek). A „létért való küzdelem” a zigóták között zajlik; az életképesebb típusok elszaporodnak, a gyengébbek kihalnak. Jelöljük  $b_{ik}$ -val, illetve  $d_{ik}$ -val az *ik* zigóta szaporodási, ill. halálozási rátáját; ekkor a

$$m_{ik} = b_{ik} - d_{ik}$$

különbséget az *ik* zigóta *fitnessének* nevezzük. Az előbbieket szerint  $m_{ik} = m_{ki}$ . Ez az időben állandónak tekintett érték határozza meg a zigóta sikerét a versenyben, vagyis azt, hogy melyik zigóta *genotípus* marad fenn és örökíti át, és melyik hal ki. Az itt szereplő fogalmakra és a következőkre nézve lásd [1] és az abban szereplő hivatkozásokat.

A következő pontban bevezetjük a gaméta genome típusok állapotterét és felírjuk a dinamikájukat leíró *Fisher-féle differenciálegyenletet*. A 3. pontban kimondjuk a *Kimura-féle maximumelvet*, bevezetjük a *Shahshahani-metrikát* és igazoljuk a maximumelv érvényességét.

## 2. Az állapotter és a Fisher-féle differenciálegyenlet

Legyen  $n$  a különböző, figyelembe vett gaméta genome típusok száma és  $x_i(t) \geq 0$  az  $i$ -edik genome típus mennyisége ( $i = 1, 2, \dots, n$ ) a  $t$  időpontban. Feltevézzük, hogy a populáció „jól kevert”, vagyis bármely gaméta egyenlő valószínűséggel találkozik bármelyik másik gametával, és a létszámok olyan nagyok, hogy a relatív gyakoriságokat egyenlőnek vehetjük a valószínűségekkel. Az  $ik$  típusú zigóták száma legyen  $x_{ik}$ , az összes zigóta száma  $\bar{x} = \sum_{i,k=1}^n x_{ik}$ , az összes gaméta száma  $2\bar{x} = \sum_{i=1}^n x_i$ . Az  $ii$  típusú *homozigóták* egy főre eső száma időegység alatt  $m_{ii}$ -vel, az ezeket alkotó  $i$  gaméták száma ennek kétszeresével változik. Az  $ik$  típusú *heterozigóták* egy főre eső száma időegység alatt  $m_{ik}$ -val, az ezekben lévő  $i$  gaméták száma ugyanennyivel változik. Az  $i$  gaméták számának dinamikáját a következő differenciálegyenlet-rendszer írja le:

$$\frac{dx_i}{dt} = 2m_{ii}x_{ii} + \sum_{k \neq i} m_{ik}x_{ik}, \quad i = 1, 2, \dots, n. \quad (1)$$

Minket a különböző genome típusok gyakorisága, a genome típusok eloszlása érdekel. Az  $i$  genome típus gyakorisága  $p_i = \frac{x_i}{2\bar{x}}$ , az  $ik$  és az  $ii$  zigóta genotípus gyakorisága

$$p_{ik} = \frac{x_{ik}}{\bar{x}} = 2p_i p_k, \quad i \neq k,$$

illetve

$$p_{ii} = \frac{x_{ii}}{\bar{x}} = p_i^2,$$

mivel ennyi annak valószínűsége, hogy egy  $i$  gamétát tartalmazó zigóta egy  $k$  (illetve egy ugyanolyan  $i$ ) gamétát tartalmazó zigótával találkozzék és ilyen utódokat hozzon létre (ez az ún. *Hardy-Weinberg-törvény*; a heterozigóták esetében ugyanaz a genotípus kétféleképpen jöhet létre: anyától-apatól, illetve apától-anyától). Az állapotter

$$R_+^n = \{(x_1, \dots, x_n) : x_i \geq 0, i = 1, \dots, n\},$$

a gyakoriságok dinamikája azonban e tér

$$S = \{(p_1, \dots, p_n) \in R_+^n : p_1 + \dots + p_n = 1\}$$

szimplexéhez van kötve. A következőkben  $(p_1, \dots, p_n)$  mindig e szimplex pontját jelöli. Nem túl nehéz számolás után (1)-ből levezethető a *természetes szelekció*

Fisher-féle differenciálegyenlete:

$$\frac{dp_i}{dt} = p_i \left( \sum_k m_{ik} p_k - \sum_{k,j} m_{kj} p_k p_j \right), \quad i = 1, \dots, n. \quad (2)$$

Az

$$m_i(p) = \sum_k m_{ik} p_k$$

kifejezés az  $i$  gamétát tartalmazó zigóták fitnessének súlyozott számtani közepe; ezt az  $i$  gaméta fitnessének nevezzük.

$$\bar{m}(p) = \sum_{k,j} m_{kj} p_k p_j = \sum_j m_j(p) p_j$$

a populáció átlagfitnessze. A (2) differenciálegyenlet ezekkel a mennyiségekkel még a következőképpen is felírható:

$$\dot{p}_i = p_i (m_i(p) - \bar{m}(p)), \quad i = 1, \dots, n. \quad (2')$$

A (2') egyenleteket összeadva látható, hogy ha a kezdeti érték  $p(0) \in S$ , akkor  $\sum_i \dot{p}_i(0) = 0$ , vagyis a megoldás trajektóriája rajta marad az  $S$  szimplexén. Továbbá, ha az  $i$  gaméta fitnessze nagyobb az átlagfitnessnél, akkor gyakorisága nő, ha pedig kisebb, akkor gyakorisága csökken. Egyensúlyi állapotba akkor kerül a rendszer, ha az összes (megmaradt) genome típus fitnessze egyenlő. Egyszerű számolás mutatja, hogy a tárgyalt esetben (szimmetrikus fitnessz mátrix) érvényes a következő:

2.1. TÉTEL. (A populációgenetika alaptétele). A (2) differenciálegyenlet megoldásai mentén a populáció átlagfitnessze növekedik.

Valóban kiadódik, hogy

$$\bar{m}'(p(t)) = 2D^2 \geq 0, \quad \text{ahol} \quad D^2 = \sum_i p_i \left( m_i - \sum_k p_k m_k \right)^2$$

a genome fitnesszek eloszlásának szórásnégyzete. Egyenlőség csak akkor áll fenn, ha a szórás zérus. Az előző egyenlőtlenségből az is látható, hogy az átlagfitnessz annál gyorsabban növekszik, minél nagyobb a szórás. Ekkor gyorsan csökken az átlag alatti gaméták gyakorisága és növekszik az átlag fölöttieké.

### 3. A Kimura-féle maximumelv, a Shahshahani-metrika

A Kimura-féle maximumelv, amelynek érvényességét intuitive elvárjuk, azt mondja ki, hogy az evolúció során nemcsak az igaz, hogy a populáció átlagfitnessze állandóan növekedik, hanem az is, hogy az eloszlás a leggyorsabb növekedés irányában változik. Matematikailag ezt a következőképpen fogalmazhatjuk meg.

3.1. TÉTEL. A Fisher-féle differenciálegyenlet megoldásai trajektóriáinak érintővektora, vagyis a (2) differenciálegyenlet jobboldalán álló vektor párhuzamos az átlagfitnessz gradiens vektorának az  $S$  szimplexre vetett vetületével.

Ha elvégezzük a számításokat, kiderül, hogy a Tétel állítása nem teljesül. Látszólag vagy a Kimura-féle maximumelvet, vagy a Fisher-féle differenciálegyenletet el kellene vetni. Az ellentmondást Shahshahani oldotta fel [2] dolgozatában azzal, hogy a gaméta genome típusok mennyiségeinek fázisterében más metrikát kell bevezetni. Az  $R_+^n$  fázistérben két „szomszédos” pont,  $(x_1, \dots, x_n)$  és  $(x_1 + dx_1, \dots, x_n + dx_n)$  távolságának négyzete, az „ívelemnégyzet” (azért, hogy összhangban maradjunk a korábbiakkal, a koordináták indexeit továbbra is alulra írjuk, és nem fogjuk az Einstein-konvenciót használni):

$$ds^2 = \frac{1}{x_1} dx_1^2 + \dots + \frac{1}{x_n} dx_n^2, \quad x_i > 0, \quad i = 1, \dots, n. \quad (3)$$

Ez a *Shahshahani-metrika*, amit fázisterünkben használni kell. Intuitív jelentése az, hogy a  $dx_i$  megváltozás annál jelentősebb, minél kisebb  $x_i$ , és nagy  $x_i$  esetén elhanyagolható. A (3) metrika bevezetésével a fázistér látszólag Riemann-térre alakul. Azonban az

$$x_i = \frac{\tilde{x}_i^2}{4}, \quad dx_i = \frac{\partial x_i}{\partial \tilde{x}_i} d\tilde{x}_i = \frac{1}{2} \tilde{x}_i d\tilde{x}_i, \quad i = 1, \dots, n \quad (4)$$

koordinátatranszformáció a (3) ívelemnégyzetet a

$$ds^2 = d\tilde{x}_1^2 + \dots + d\tilde{x}_n^2$$

alakra transzformálja, ami azt jelenti, hogy a tér euklideszi maradt, bevezethető Descartes-féle koordinátarendszer, csupán az eredeti (mármint az  $x_i$  rendszer) nem az.

Ezek után kiszámítjuk az átlagfitnessz gradiensét és annak vetületét az  $S$  szimplex érintősíkjára. A hullámos, Descartes-féle koordinátarendszerben  $S$  egyenlete

$$\tilde{x}_1^2 + \dots + \tilde{x}_n^2 = 4, \quad \tilde{x}_i \geq 0, \quad i = 1, \dots, n,$$

vagyis a Shahshahani-metrikában  $S$  az  $(n-1)$  dimenziós, 2 sugarú, origó középpontú gömbfelületnek a pozitív ortánsba eső része. Az átlagfitnesszt áttranszformáljuk az új koordinátarendszerbe:

$$\bar{m}(\tilde{x}) = \sum_{i,k} m_{ik} x_i(\tilde{x}) x_k(\tilde{x}) = \frac{1}{16} \sum_{i,k} m_{ik} \tilde{x}_i^2 \tilde{x}_k^2.$$

Ha egy Riemann-térben egy skalármező gradiensét kiszámítjuk, akkor elsődlegesen a vektor kovariáns koordinátáit kapjuk meg. Euklideszi térben, Descartes-féle

koordinátarendszerben azonban ezek megegyeznek a kontravariáns koordinátákkal. Tehát az átlagfitnessz gradienseinek kontravariáns koordinátái

$$\tilde{\nabla}\bar{m}(\tilde{x}) = \frac{1}{4} \left[ \sum_k m_{1k} \tilde{x}_1(\tilde{x}_k)^2, \sum_k m_{2k} \tilde{x}_2(\tilde{x}_k)^2, \dots, \sum_k m_{nk} \tilde{x}_n(\tilde{x}_k)^2 \right].$$

E vektor  $S$  érintősíkjára vett vetületének kiszámításához látnunk kell, hogy az  $S$  gömbfelület  $\tilde{x}$  pontjában a normálvektora  $\tilde{x}$ , ennek abszolút értéke 2, vagyis a normál egységvektor  $\frac{\tilde{x}}{2}$ . A gradiens vetülete  $S$  érintősíkjára tehát

$$\begin{aligned} \tilde{v}(\tilde{x}) &= \tilde{\nabla}\bar{m}(\tilde{x}) - \left( \tilde{\nabla}\bar{m}(\tilde{x}) \cdot \frac{\tilde{x}}{2} \right) \frac{\tilde{x}}{2} = \tilde{\nabla}\bar{m}(\tilde{x}) - \bar{m}(\tilde{x})\tilde{x} \\ &= \left[ \tilde{x}_1 \left( \sum_k m_{1k} \frac{(\tilde{x}_k)^2}{4} - \bar{m}(\tilde{x}) \right), \dots, \tilde{x}_n \left( \sum_k m_{nk} \frac{(\tilde{x}_k)^2}{4} - \bar{m}(\tilde{x}) \right) \right]. \end{aligned}$$

A (4) transzformációs formulát alkalmazva a kontravariáns vektorkoordinátákra

$$\begin{aligned} v^i(x) &= \frac{\tilde{x}_i}{2} \tilde{x}_i \left( \sum_k m_{ik} \frac{(\tilde{x}_k)^2}{4} - \bar{m}(\tilde{x}) \right) \\ &= 2x_i \left( \sum_k m_{ik} x_k - \bar{m}(x) \right) \\ &= 2x_i(m_i(x) - \bar{m}(x)), \end{aligned}$$

ahol  $\sum_i x_i = 1$ . Ez (2') jobb oldalának kétszerese. Ezzel a 3.1. tételt bebizonyítottuk.

Vannak, akik a természetes kiválasztáson alapuló evolúcióból azt a következtetést vonják le, hogy az egyes fajok változatlan körülmények között a lehető legéletpesebb színvonalra fejlődnek, illetve fejlődtek, hogy a ma élő orosz lán, szúnyog, tölgyfa, vagy ember maga a tökéletesség. Ez hamis következtetés. Amikor hosszú időre állandó körülmények jönnek létre (időjárás, együttélő fajok, stb), az adott faj az ún. *fitness táj* (fitness landscape) egy meghatározott pontján található. A fitness táj az átlagfitnessz grafikonja a gamétaeloszlások szimplexe fölött. Úgy képzeljük ezt el, mint egy holdbéli tájat, vagy egy vulkánikus csúcsokkal tarkított „kamcsatkai” vidéket. Amikor az evolúció elkezdődik, a populáció a Kimura-elv szerint elkezd mászni fölfelé azon a hegyen, amelyen éppen van, vagyis gaméta összetétele ahhoz a ponthoz tart, amely fölött a hegy csúcsa helyezkedik el. Ha ezt a pontot elérte, nyugalomba kerül, hiszen innen már nem tud elmozdulni (úgy, hogy átlagfitnessze növekedjék). Innen vágyakozva tekinthet más közeli, vagy távoli, magasabb hegy-csúcsokra, ahol már nem lenne vakbele, nem kapna rendszeresen torokgyulladást és könnyebben megértené a matematikát. Más pontba csak akkor juthat el, ha a körülmények változása következtében, az időben a fitnessz táj változik, átalakul úgy, hogy lehetővé tegye a mozgást, illetve akkor, ha igen kis valószínűségű, lényeges, hasznos mutáció történik.

**Hivatkozások**

- [1] FARKAS, M.: *Dynamical Models in Biology*. Academic Press, San Diego, 2001
- [2] SHAHSHAHANI, S.: *A new mathematical framework for the study of linkage and selection*. Memoires Amer. Math. Soc. 211., 1979

(Beérkezett: 2007. február 21.)

FARKAS MIKLÓS

BP. Műszaki és Gazdaságtudományi Egyetem  
Matematikai Intézet  
1521 Budapest

## NATURAL SELECTION AND RIEMANNIAN GEOMETRY

MIKLÓS FARKAS

The Fisher differential equation of natural selection is presented and discussed. The Kimura maximum principle is stated and proved by introducing a new metric, the Shahshahani metric in the space of gameta genome types.