

A génhelyettesítés valószínűségi elmélete

Meszéna Géza – Mizera Ferenc – Szilágyi András

2005. október 19.

1. Bevezetés: a kérdés

Mi egy mutáció sorsa?

Egy N (állandó) létszámú populáció egy lokuszán mutáció történik. Egyrészt, kíváncsiak vagyunk arra, mi a valószínűsége annak, hogy a mutáns allél fixálódik (azaz elterjed és kiszorítja az eredeti allélt), illetve annak, hogy kiesik. Biztosak lehetünk abban, hogy a két esemény közül az egyik megtörténik. Nulla ugyanis annak a valószínűsége annak, hogy a populáció heterogén állapota végtelen ideig fentmaradjon.

Másrészt, kérdezhetjük, mi a mutáns fixálódásának várható ideje, feltéve, hogy fixálódik.

Természetesen, a válasz mindkét kérdés esetében függeni fog attól, hogy a mutáns előnyös, hátrányos, vagy neutrális. Másrésztől lényeges a véletlen szerepe is, vagyis hogy a mutáns allélnak szerencséje van-e. Nyilvánvaló, hogy minél kisebb a populáció, annál nagyobb a sztochaszticitás szerepe. De még az $N \rightarrow \infty$ határeset sem teszi a problémátt determinisztikussá: a mutáns nagy populációban is egyetlen példányban jelenik meg, sorsa tehát nagy mértékben a véletlenen múlik.

Lényeges, hogy *egyetlen* mutáció további sorsáról beszélünk. Tárgyalásunk szempontjából érdektelen, hogy a vizsgált időtartam alatt bekövetkeznek-e további mutációk. Csak az a kérdésünk, hogy a távoli jövőben a populációt az általunk vizsgált mutációs eseményben létrejött mutáns leszármazottjai alkotják-e, vagy sem.

Feltételezzük, hogy egyetlen generációváltás alatt a populáció összetétele nem sokat változik. Ez egyrészt jelenti azt, hogy a populáció nem nagyon kicsi, másrészt jelenti azt, hogy a szelekció nem nagyon erős. Ekkor a mutáns gén részaránya a populációban (amelyet röviden génearánynak fogunk nevezni) olyasfajtképpen változik, mint egy bolyongó részecske koordinátája. A génearány-változás véletlen komponensének megfelelője a részecske Brown-mozgása, amit egy diffúziós egyenlettel írhatunk le. A szelekció okozta determinisztikus komponens pedig annak felel meg, mintha a részecsként egy külső (pl. elektromos) erő hatására mozogna.

A terminológiával mindig baj van. Az, hogy a vonatkozó egyenlet másképpen nevezik a fizikai és a biológiai irodalomban, a legkevesebb. Ami rosszabb: a biológiai irodalom „genetikai sodródás”-nak, „drift”-nek nevezi a génearány véletlen, nem szelekciós ingadozását, míg a részecskemozgási képből a determinisztikus, adott irányban történő mozgást hívnánk sodródásnak, driftnek! A drift szó használatát elkerülöm, a sodródás a genetikai terminológia szerint a szelekció ellentétpárjaként szerepel.

A következő fejezet a matematikai alapokat vezeti be, az utána következő tartalmazza a tulajdonképpeni biológiai elméletet. A fogalmazás az egyszerűség kedvéért aszexuális populációra vonatkozik. A 4.3. fejezetben mondjuk el, mi változik egy diploid populációban. Mindezt természetesen azért csináljuk, mert az eredményeknek jelentősége van az evolúció elmélete szempontjából. Erről esik szó nagyon röviden az 5. fejezetben.

2. Fokker-Planck (Kolmogorov) egyenlet

Jelölje $\Phi(p, x, t)$ annak a valószínűség-sűrűségét, hogy a génearány a p értékről indulva t idő alatt x -be jut. Értelmezése folytán:

$$\int \Phi(p, x, t) dx = 1, \quad (1)$$

hiszen valahova biztosan elfejlődik a populáció. A $\Phi(p, x, t)$ időbeli változását két Fokker-Planck egyenlet írja le.

Direkt (forward) Fokker-Planck:

$$\frac{\partial \Phi(p, x, t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} (D(x) \Phi(p, x, t)) - \frac{\partial}{\partial x} (M(x) \Phi(p, x, t)), \quad (2)$$

Fordított (backward) Fokker-Planck:

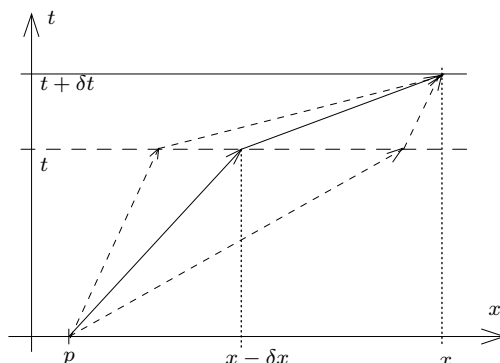
$$\frac{\partial \Phi(p, x, t)}{\partial t} = \frac{1}{2} D(p) \frac{\partial^2}{\partial p^2} \Phi(p, x, t) + M(p) \frac{\partial}{\partial p} \Phi(p, x, t). \quad (3)$$

A biológiai irodalomban a fenti egyenleteket Kolmogorov nevét viselik. A fizikusok ezeket az egyenleteket Fokker és Planck nevéhez fűzik, Kolmogorov egyenletként egy általánosabb egyenlőséget emlegetnek.

2.1. A Fokker-Planck egyenlet levezetése direkt esetben

A Fokker-Planck egyenlet levezetéséhez a $\Phi(p, x, t)$ eloszlásfüggvénynek azt a tulajdonságát használjuk ki, hogy ha különböző összetételeken keresztül fejlődik a populáció ugyanabba a végállapotba, akkor a különböző utakra vett átmenetek valószínűségeit felösszegezve kapjuk az átmenet valószínűségét.

Vizsgáljuk meg azt az esetet, amikor a populáció $t + \delta t$ idő alatt a p kezdeti géngyakoriságról x végső géngyakoriságba fejlődik. Bontsuk ezt az időbeli fejlődést egy t illetve egy δt időszakaszra. Az első szakaszban a populáció a p összetételből $x - \delta x$ összetételbe fejlődik, a maradék δt idő alatt pedig x -be fejlődik tovább, amint azt az alábbi ábrán a folytonos nyíl szemlélteti.



Mivel δx -et önkényesen választottuk, más úton (azaz más δx -szel) is eljuthat a populáció az x végállapotba, ezt szemléltetik a szaggatott vonalak. Az összes lehetséges fejlődési út valószínűségeinek összege adja meg a $p \rightarrow x$ átmenet valószínűségét. Matematikailag ezt így írhatjuk:

$$\Phi(p, x, t + \delta t) = \int \Phi(p, x - \delta x, t) \Phi(x - \delta x, x, \delta t) d(\delta x) = \int \Phi(p, x - \delta x, t) g(x - \delta x, \delta x, \delta t) d(\delta x),$$

ahol a könnyebb érthetőség kedvéért bevezettük a következő jelölést:

$$\Phi(x - \delta x, x, \delta t) = g(x - \delta x, \delta x, \delta t).$$

Kicsi δt -re csak a kis δx értékek adnak jelentős járulékot. Ennek megfelelően, az integrandust közelítsük úgy, hogy az $x - \delta x$ változójában másodrendig Taylor sorba fejtsük x körül:

$$\begin{aligned} \Phi(p, x, t + \delta t) = \int d(\delta x) \left[\Phi(p, x, t) g(x, \delta x, \delta t) - \delta x \frac{\partial}{\partial x} (\Phi(p, x, t) g(x, \delta x, \delta t)) + \right. \\ \left. + \frac{1}{2} (\delta x)^2 \frac{\partial^2}{\partial x^2} (\Phi(p, x, t) g(x, \delta x, \delta t)) \right]. \end{aligned}$$

Most elvégezzük az integrálást δx szerint:

$$\Phi(p, x, t) \underbrace{\int d(\delta x) g(x, \delta x, \delta t)}_1 - \frac{\partial}{\partial x} \left[\Phi(p, x, t) \underbrace{\int d(\delta x) \cdot \delta x \cdot g(x, \delta x, \delta t)}_{\langle \delta x \rangle_{\delta t} = M(x) \delta t} \right] +$$

$$+\frac{1}{2}\frac{\partial^2}{\partial x^2}[\Phi(p, x, t) \underbrace{\int d(\delta x) \cdot (\delta x)^2 g(x, \delta x, \delta t)}_{\langle (\delta x)^2 \rangle_{\delta t=D(x)\delta t}],$$

közben kihasználtuk, a g függvény értelmezéséből adódó eredményeket.
Ezzel a (2) alatti egyenletet igazoltuk.

2.2. A Fokker-Planck egyenlet levezetése fordított esetben

Az egyenlet fordított alakjának levezetését hasonlóan végezzük, de most a p állapothoz közel vesszünk fel egy közbenső állapotot. A fenti levezetéshez hasonlóan a eljárva:

$$\begin{aligned} \Phi(p, x, t + \delta t) &= \int \Phi(p, p + \delta p, \delta t) \Phi(p + \delta p, x, t) d(\delta p) = \\ &= \int d(\delta p) g(p, \delta p, \delta t) \left[\Phi(p, x, t) + \delta p \frac{\partial}{\partial p} \Phi(p, x, t) + \frac{1}{2} (\delta p)^2 \frac{\partial^2}{\partial p^2} \Phi(p, x, t) \right] \end{aligned} \quad (4)$$

Átrendezve:

$$\begin{aligned} \Phi(p, x, t + \delta t) &= \Phi(p, x, t) \underbrace{\int d(\delta p) g(p, \delta p, \delta t)}_1 + \underbrace{\int d(\delta p) \cdot \delta p \cdot g(p, \delta p, \delta t)}_{\langle \delta p \rangle_{\delta t=M(p)\delta t}} \cdot \frac{\partial}{\partial p} \Phi(p, x, t) + \\ &+ \frac{1}{2} \underbrace{\int d(\delta p) \cdot (\delta p)^2 \cdot g(p, \delta p, \delta t)}_{\langle (\delta p)^2 \rangle_{\delta t=D(p)\delta t}} \cdot \frac{\partial^2}{\partial p^2} \Phi(p, x, t) \end{aligned} \quad (5)$$

Innen a (3) fordított Fokker-Planck egyenlet már adódik.

3. Szelekció és sodródás

A fent levezetett Fokker-Planck egyenletek természetesen bármilyen bolyongást leírhatnak. A genetikai alkalmazáshoz specifikálnunk kell az $M(x)$, $D(x)$ függvényeket. Az egyszerűség kedvéért fel fogjuk tenni, hogy a generációk elkülönülnek, mint például egy egynyári növény esetén. Ekkor célszerű a generációváltást választani időegységnek.

3.1. Szelekció

Jelölje s a mutáns szelekciós előnyét, amely negatív is lehet. Ez azt jelenti, hogy a populáció x hányadát képző mutáns szubpopuláció által hordozott gének fitnessz-értéke $1 + s$, szemben az eredeti gént hordozó szubpopuláció 1-es fitnesszével.

Mivel a génrány-változás determinisztikus komponensét keressük, úgy számolhatunk, mintha végtelen nagy lenne a populációnk. A génrány egy generáció múlva:

$$x' = \frac{x(1+s)}{x(1+s) + (1-x)} = \frac{x(1+s)}{1+xs} \quad (6)$$

$$\delta x = x' - x = \frac{xs(1-x)}{1+xs} \approx sx(1-x) \quad (7)$$

ahol az utolsó lépésben felhasználtuk, hogy $sx \ll 1$. A keresett $M(x)$ együttható éppen az egységnyi idő – vagyis a generációs idő – alatti génrány-változás:

$$M(x) = sx(1-x) \quad (8)$$

4. Genetikai sodródás

A diffúziós együttható a genetikai sodródást, vagyis a véletlen komponenst írja le. Itt már lényeges, hogy a populáció N , tehát véges sok, egyedből áll.

Az egyszerűség kedvéért tegyük fel, hogy a születendő utódok egymástól függetlenek. Mivel bármelyikük x valószínűséggel számazik mutáns őstől, a mutánsok k száma az utódpopulációban egy binomiális eloszlás:

$$P_N(k) = \binom{N}{k} x^k (1-x)^{N-k}. \quad (9)$$

Ennek az eloszlásnak a várható értéke:

$$\sum_{k=0}^N k P_N(k) = Nx,$$

szórásnégyzete pedig:

$$\sum_{k=0}^N ((k - Nx)^2 P_N(k) = Nx(1-x).$$

Ez azt jelenti, hogy várhatóan Nx darab mutáns gén továbbítódik, a génearány várható értéke x lesz, varianciája pedig:

$$D(x) = \frac{x(1-x)}{N}. \quad (10)$$

4.1. A fixálódás valószínűsége

Legyen a fixálódási valószínűség

$$u(p) = \Phi(p, 1, \infty), \quad (11)$$

amelyre a modell feltevései miatt

$$u(0) = 0, \quad u(1) = 1 \quad (12)$$

teljesül. Értékét a fordított Fokker-Planck egyenletből számítjuk ki:

$$\frac{1}{2} D(p) \frac{d^2 u}{dp^2} + M(p) \frac{du}{dp} = 0, \quad (13)$$

hiszen $\frac{\partial \Phi}{\partial t}$ a $t = \infty$ -ben eltűnik. Ennek a másodrendű lineáris egyenletnek a megoldása:

$$u(p) = \frac{\int_0^p G(x) dx}{\int_0^1 G(x) dx}, \quad (14)$$

ahol

$$G(x) = e^{-2 \int_0^x \frac{M(x')}{D(x')} dx'}. \quad (15)$$

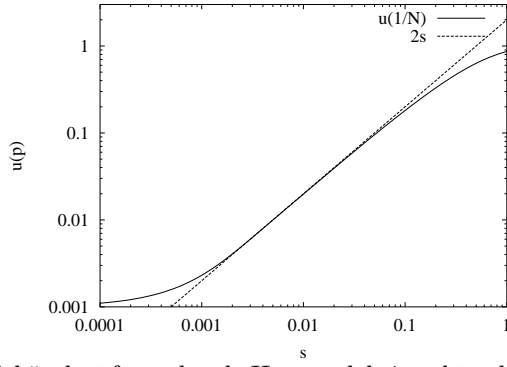
Az integrálásokat elvégezve:

$$u(p) = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}} = \begin{cases} 2Nsp, & \text{ha } Ns \gg 1 \\ p, & \text{ha } Ns \ll 1 \end{cases} \quad (16)$$

Ha $p = \frac{1}{N}$:

$$u(p) = \begin{cases} 2s, & \text{ha } Ns \gg 1 \\ \frac{1}{N}, & \text{ha } Ns \ll 1 \end{cases} \quad (17)$$

A alábbi grafikonon egy új mutáns fixálódás valószínűségét ábrázoljuk a szelekciós előny függvényében – logaritmikuskálán. A populáció létszáma $N = 1000$, s mivel csak egy mutáns van $p = 0.001$.



Látható az összhang a (17) közelítő formulával. Ha a szelekciós előny kicsi, csak a véletlenül múlik, hogy elterjed-e az új mutáns: mint mindenkinek, neki is csak $1/N = 0,001$ erre az esélye. Nagyobb szelekciós előnyöknél valószínűbbé válik a fixálódás valószínűsége, és az $u(s) = 2s$ formula viszonylag jó közelítés $s < 0,1$ -re.

4.2. Fixálódási idő

Határozzuk meg a fixálódás idejét! Legyen

$$v(p, t) = \Phi(p, 1, t) \quad (18)$$

és

$$T(p) = \int_0^\infty t \frac{\partial v(p, t)}{\partial t} dt. \quad (19)$$

Ekkor

$$\bar{t}(p) = \frac{T(p)}{u(p)}. \quad (20)$$

Felírjuk a Fokker-Planck egyenletet v -re:

$$\frac{\partial v(p, t)}{\partial t} = \frac{1}{2}D \frac{\partial^2 v(p, t)}{\partial p^2} + M \frac{\partial v(p, t)}{\partial p}, \quad (21)$$

majd deriváljuk t szerint, szorozzuk t -vel, és integráljuk t szerint 0-tól ∞ -ig:

$$\int_0^\infty \frac{\partial^2 v(p, t)}{\partial t^2} t dt = \frac{1}{2}D \frac{\partial^2 T(p)}{\partial p^2} + M \frac{\partial T(p)}{\partial p}. \quad (22)$$

Az egyenlet bal oldalát parciális integrálással számoljuk:

$$\int_0^\infty \frac{\partial^2 v(p, t)}{\partial t^2} t dt = \underbrace{\left[t \frac{\partial v(p, t)}{\partial t} \right]_0^\infty}_0 - \int_0^\infty \frac{\partial v(p, t)}{\partial t} dt = -u(p). \quad (23)$$

Ily módon

$$T''(p) + \frac{2M}{D}T'(p) + \frac{2u(p)}{D} = 0 \quad (24)$$

adódik, ahol vesszővel a p szerinti deriválást jelöljük. Neutrális esetben szelekció nélkül, amikor $M = 0$, $u(p) = p$ és $D = \frac{p(1-p)}{N}$:

$$T''(p) = -\frac{2p}{\frac{p(1-p)}{N}} = -\frac{2N}{1-p}. \quad (25)$$

Kétszer integrálva:

$$T' = 2N \ln(1-p) + c_1 \quad (26)$$

$$T = -2N((1-p)(\ln(1-p) - 1)) + c_1 p + c_2. \quad (27)$$

A $T(0) = T(1) = 0$ határfeltételekből:

$$T(p) = -2N(1-p)\ln(1-p) \quad (28)$$

és

$$\bar{t}(p) = \frac{T(p)}{u(p)} = \frac{-2N(1-p)\ln(1-p)}{p} \rightarrow 2N, \quad \text{ha } p \rightarrow 0 \quad (29)$$

adódik.

4.3. Diploid populációk

Eddig aszexuálisan szaporodó populációt vizsgáltunk. Egy véletlenszerűen párosodó diploid populáció esetében a különbség annyi, hogy az N létszámú populációban a szóbanforgó lokusz $2N$ példányban van jelen – minden egyedben egy apai és egy anyai verzió. Ennek megfelelően a diploid esetben mindenütt $2N$ -et kell írni az N helyett. Mivel csak egyetlen lokuszról beszélünk, a rekombináció, amely a szexuális szaporodást érdekessé tenné, itt számunkra nem játszik szerepet.

Nem neutrális mutáció esetén különbséget okozhat még a dominancia, vagyis az apai és anyai eredetű gének kölcsönhatása. Ha azonban feltételezzük, hogy az apai és anyai gének additívan járulnak hozzá az egyed fitnesszéhez, akkor ez leírható úgy, mintha a géneknek egymástól függetlenül lenne fitnessz előnye-hátránya.

Ezek szerint, részletesebb információ hiányában a fentiek alkalmazandók egy szexuális populációban is, azzal, hogy a neutrális mutáns fixálódásának várható ideje $2N$ helyett $4N$.

5. Genetikai óra

Legyen μ a nem-hátrányos mutáció valószínűsége a genom egy szakaszán. Egy N létszámú populációban generációnként μN számú nem-hátrányos mutáns jelenik meg. Ezek túlnyomó többsége neutrális, mivel az előnyös mutánsok ritkák. Tekintve, hogy a neutrális mutánsok $1/N$ hányada fixálódik, μ a valószínűsége egy olyan mutáns megjelenésének, amely a végül fixálódik.

A fixálódó mutánsok tehát egy állandó, a populációmérettől független rátával érkeznek. Ezek felhalmozódása révén változik a genom-szekvencia, illetve válnak a szekvenciák különbözővé a párhuzamos leszármazási vonalakon. A szekvencia-összehasonlítások ily módon alkalmasak a szétválás óta eltelt idő becslésére.

Fontos megjegyezni, hogy a nem-hátrányos mutánsok aránya a összes mutánsok között génről génre változik, nagyságrendet is.