

## Bevezetés

A leszármazási (rokonsági) modellek a populációgenetika egy egzotikus témájából központi kérdéssé lépett elő mindenki számára, aki a DNS szekvenciákat vizsgál. Az a célunk, hogy egyszerűen és elemi eszközökkel bemutassuk a leszármazási (*coalescent*) modelleket. Ehhez elemi kombinatorikai ismeretekre és a valószínűségszámítás elemeire lesz szükségünk. A coalescens elméleté a jövő. Van olyan részterülete, ahol az elmélet nem működik (vagy elfogadhatatlanul lassú) valós adatokra. Ezért a modellezőknek és a szimulációs elemzéseknek is tág tere van.

## Urnamodellek a genetikában

A mikroevolúciós folyamatok jól modellezhetők különféle színű golyókat tartalmazó urnákkal (dobozokkal). Ezeket a modelleket urnamodellekként vagy Pólya-féle urnamodellekként szokás említeni. Ezek közül az egyik legfontosabb a *Wright-Fisher-féle modell* (vagy rövidebben Wright-Fisher modell). A Wright-Fisher-féle modell az allélgyakoriságok változásának matematikai modellje véges populációban. Hosszú távon a random mintavételezés miatt allélek veszhetnek el a populációból. Ez a populáció genetikai diverzitásának csökkenését jelenti. A populációgenetikában ezt a jelenséget *genetikai sodródásnak* vagy *driftnak* hívják. Genetikai sodródás az a folyamat, amikor a populáció genetikai összetételének időbeli változását kizárólag a random fluktuáció befolyásolja. Az evolúciós folyamatnak más összetevői is vannak, mint például a mutáció és a természetes szelekció. A populáció genetikai összetételét számos további folyamat is befolyásolja; példaként említhetjük a kivándorlást (emigráció) és a bevándorlást (imigráció).

### A Wright-Fisher modell

A Wright-Fisher modell számos idealizált feltétel teljesülése esetén érvényes.. Ezek a következők.

1. *Diszkrét, nem átfedő generációk.* Humán népesség esetére lefordítva ez azt jelenti, hogy mindenkinek ugyanaz a várható élettartalma a fogamzástól a reprodukcióig, és a reprodukciótól a halálig és a szaporodás egyszerre történik minden egyedre. Szerencsére ez nem egy szigorú megszorítás a gyakorlati követelményeket illetően. Az olyan modellek amelyek lehetővé teszik az átfedő generációkat (azaz nem minden gén születik és hal meg azonos időben) statisztikailag azonos leszármazási vonalat eredményez.
2. *Haploid egyedek vagy két rész-populáció (hímek és nőstények).* Olyan esetben, amikor nincs a heterózist érintő szelekció, akkor kevés kvantitatív következménye van annak a feltételnek, hogy  $2N$  mérete van a populációnak a haploid modell esetében használt  $N$  helyett.
3. *A populáció mérete konstans.* Ez lényeges biológiai feltétel és a modell lényeges mennyiségei változnak, ha a populációméret növekszik, elcsúszik vagy oszcillál vagy igen kis méretre való lecsökkenésen megy keresztül (ún. palacknyak hatás).
4. *Minden egyed egyforma.* Ez ésszerű feltételezés az alapmodell szempontjából, jóllehet minden bizonnyal nem reális minden lokusz esetén. Ennek

5. A populációnak nincs térbeli vagy szociális struktúrája.

6. A gének (vagy a szekvenciák) nem rekombinálódnak.

Egy gént (lokuszt) vizsgálunk és feltételezzük, hogy két eltérő változata, allélje van a génnek. Ezek az A és a B allél. Diploid populációt vizsgálunk, amelyben  $2N$  példánya van a géneknek minden egyes generációban. Az első generáció esetén  $i$  darab van az A allélből és  $(2N - i)$  darab B allél van. A következő generáció  $2N$  alléljét binomiális modell szerint generáljuk, amely szerint minden egyes allélt random módon választunk ki az előző generáció génkészletéből és a  $2N$  allélje a második generációnak egymástól teljesen függetlenül választjuk ki (azaz visszatevéses mintavétellel). Ez a folyamat tetszőlegesen sok generáción át folytatódik.

A modell alapján nyilvánvalóan lehetséges, hogy valamelyik generációban a  $2N$  allél egyforma lesz, mondjuk mind A típusú vagy B típusú. Ha az összes allél azonossá válik, akkor ez így is marad a további generációkban. A genetikában ezt *allél uniformitásnak* hívják.

Mindezt bemutathatjuk az alábbi urnamodell segítségével. Legyen az 1. urnában  $i$  darab piros golyó és  $(2N - 1)$  darab zöld golyó. Az 1. urnából  $2N$  golyót választunk random visszatevéses mintavétellel és ezeket tesszük a 2. urnába. Majd ebből az urnából választunk  $2N$  golyót random visszatevéses mintavétellel; ezek alkotják a 3. urna tartalmát, és így tovább. Az  $n$ -edik generáció genetikai összetétele annak felel meg, hogy hány piros (és zöld) golyó van az  $n$ -edik urnában.

Kérdések, amelyeket a Wright-Fisher modell segítségével megválaszolhatunk.

- Mekkora az esélye (valószínűsége), hogy az A allél rögzül?
- Hány generáció alatt rögzül az A allél (vész el a B allél)?
- Hány generáció alatt rögzül a B allél (vész el az A allél)?
- Hány generáció alatt alakul ki az allél uniformitás?

Vezessünk be néhány jelölést. Legyen

$p = i/(2N)$  = az A allélek aránya az első generációban;

$X_n$  = az A allélek száma az  $n$ -edik generációban;

$$p_{ij} = P(X_{n+1} = j \mid X_n = i)$$

$$p_\infty = P(X_n = 2N \text{ valamely véges } n - \text{re})$$

A binomiális mintavétel alapján adódik, hogy

$$P(X_{n+1} = j \mid X_n = i) = p_{ij} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}$$

Ez a Wright-Fisher egyenlet. A  $p_{ij}$ ,  $i=1, \dots, n$ ,  $j=1, \dots, n$  elemekből álló mátrixot („táblázatot”) *átmeneti mátrixnak* nevezzük. A Wright-Fisher modell révén előálló sztochasztikus folyamat egy *Markov-folyamat*. A Markov-folyamatokat egyértelműen meghatározza (leírja) az átmeneti mátrix.

## Allél uniformitás elérése

A valószínűségszámítás eszköztárának használatával megmutathatjuk, hogy a Wright-Fisher-féle modell esetében az allél uniformitás végül mindig bekövetkezik. Az urnamodell esetében ez azt jelenti, hogy előbb vagy utóbb az urnában csak azonos színű golyók lesznek. Minél nagyobb a kiinduláskor a piros labdák aránya annál nagyobb annak a valószínűsége, hogy egy ponttól (generációtól) kezdődően a labdák mind pirosak legyenek az urnában. A következő elegáns eredményt kapjuk.

**Tétel.**  $p(t \rightarrow \infty) = p$ .

Ez azt jelenti, hogy ha a B allél gyakorisága 40% (az A allél gyakorisága 60%), akkor 40% az esélye, hogy a B allél végső soron elvész a populációból.