Eötvös Loránd Tudományegyetem

Természettudományi Kar

Demonstrációs célú szimulációk egyszerű populációdinamikai folyamatok modellezésére

Szakdolgozat

Sáfár Rebeka

Matematika B.Sc., elemző szakirány

Témavezető: **Izsák János**, egyetemi tanár

Alkalmazott Analízis Tanszék

Budapest

2011

Tartalomjegyzék

1. Biológiai bevezetés **2**

2. A Hardy-Weinberg törvény - szakirodalmi feldolgozás **6**

2.1. Panmiktikus szaporodás 6

2.2. Az egyensúly feltételei 7

2.3. Hardy-Weinberg egyensúly három allél esetén 8

2.4. Hardy-Weinberg egyensúly k számú allél esetén 9

3. Egy, az előzőektől eltérő szaporodási modell **11**

3.1. Leírás, feltételek 11

3.2. Eredmények 12

4. Ivari kromoszómához kötött lókuszok **16**

5. A szomatikus genotípusok előfordulási valószínűségeinek sorozatai az r_0, s_0 és p_0 értékek függvényében **23**

5.1. A p_n értékek változása 23

5.2. Az r_n értékek változása 26

5.3. Az s_n értékek változása 31

6. Összefoglalás **34**

1. fejezet

Biológiai bevezetés

Az alapvető genetikai összefüggéseket és törvényeket a legegyszerűbb párosodási sémán keresztül mutatom be, ezért felteszem, hogy az alábbi populációdinamikai sémát nem befolyásolják külső tényezők, például minden hím és női egyed utódainak száma a szülők genotípusától független és azonos (nincs szelekció), minden hím és női egyed egymással való párosodásának azonos a valószínűsége, nincs mutáció és nincs a populációba való be- és kivándorlás sem. Ez az úgynevezett *panmiktikus szaporodás*. Ilyen ideális populációk a természetben nincsenek.

A populáció egyedeit egyetlen *szomatikus sejttel* reprezentáljuk, ez megtehető, mert egy szervezet szomatikus sejtjei azonos genotípusúak. A szomatikus sejt genetikai anyaga egymáshoz tartozó apai, illetve anyai eredetű DNS láncokból áll, rajtuk a lókuszhoz tartozó génváltozatok, az úgynevezett *allélok* valamelyikével. A számfelező osztódás (*meiozis*) során a szomatikus sejtekből két utódsejt (*ivarsejt*) keletkezik. Így a lehetséges esetek, ha az adott lókuszhoz az A és az a allél tartozik:

- $AA \longrightarrow A,\ A$
- $Aa \longrightarrow A,\ a$
- $aa \longrightarrow a,\ a$

(Az első oszlopban a szülő szomatikus genotípusa, a másodikban az ivarsejtek genotípusa látható.) Egy hím és egy női ivarsejt egyesülésével egy utód (szomatikus) sejt, úgynevezett zigóta keletkezik, esetünkben az alábbi lehetséges módokon:

- $a + a \longrightarrow aa$
- $A + a \longrightarrow Aa$
- $A + A \longrightarrow AA$,

ahol a második oszlop mutatja a zigóta (és utód egyed)szomatikus genotípusát. A zigótákból hosszú fejlődés után alakul ki maga az egyed.

Tekintsük az egyedeket reprezentáló sejtek halmazában egy nem ivari kromoszómához kötött lókuszhoz tartozó allélok halmazát. Feltesszük, hogy két allél típus létezik (a és A) és minden szomatikus sejtből két ivarsejt keletkezik. Legyen ekkor az A allélok relatív gyakorisága a reprezentáns szomatikus sejtek összes A-a lókuszainak körében p. Ekkor az a allél relatív gyakorisága ugyanitt $q = 1 - p$. A reprezentáns ivarsejtek száma így kétszer annyi lesz, mint a reprezentáns testi sejteké. Az A allél relatív gyakorisága az ivarsejtek körében szintén p lesz. Összegezve tehát, egy párosodási sémában résztvevő hím ivarsejt p valószínűséggel A, $q = 1 - p$ valószínűséggel a genotípusú. Ugyanezek a valószínűségek vonatkoznak a petesejtre is a kezdeti feltételek miatt. Így az utódgenereció szomatikus sejtjeinek genotípus előfordulási valószínűségei könnyen áttekinthetők, a *Punnett-tábla* segítségével:

			hím	
			p	q
			A	a
	p	A	p^2	pq
nőstény	q	a	pq	q^2

1.1. ábra. Punett-tábla

Tehát elemi valószínűségszámítási meggondolások alapján az utódgenerációban a szomatikus genotípusok előfordulási valószínűségei:

$$P[(A,A)] = p^2, P[(A,a)] = 2pq, P[(a,a)] = q^2 \tag{1.1}$$

Az utódgeneráció lehetséges szomatikus genotípusainak előfordulási valószínűségei tehát levezethetők a reprezentáns hím, illetve női ivarsejtek körében való előfordulási valószínűségekből. Észrevehetjük, hogy a fenti három valószínűséget előállíthajuk $(p+q)^2$ kifejtésével. Ezt általánosítva egy tételhez jutunk, aminek kimondása előtt bevezetek néhány fogalmat. Legyenek egy populáció reprezentáns szomatikus sejtjeinek genotípusai $G_1, \ldots, G_g$, ezekből a sejtekből származó ivarsejtek genotípusai pedig $a_1, \ldots, a_s$ (könnyen belátható a $g = \frac{s(s+1)}{2}$ összefüggés). Ekkor a $P(a_1)(a_1) + \ldots + P(a_s)(a_s)$ eseménypolinomot *ivarsejt genotípus valószínűségeloszlási polinomnak* nevezzük. Hasonlóan, a $P(G_1)(G_1) + \ldots + P(G_g)(G_g)$ eseménypolinomot *szomatikus genotípus valószínűségeloszlási polinomnak* nevezzük. (Figyelembe vesszük, hogy a diszjunkt $[a_1], \ldots, [a_s]$ események teljes eseményrendszert alkotnak.)

1.1. Tétel. *Az utódgenerációban a szomatikus genotípus valószínűségeloszlási polinom együtthatói előállnak az előző generációbeli hím (és női) ivarsejt genotípus valószínűségeloszlási polinom négyzetének polinom-együtthatóiként.*

1.1. Bizonyítás. *Egy (a_i, a_j) genotípusú zigóta (utódsejt) $i = 1, \ldots, s$, $j = 1, \ldots, s$ a következőképpen jöhet létre:*

- *a_i genotípusú hím és a_j genotípusú női ivarsejt egyesül*
- *a_j genotípusú hím és a_i genotípusú női ivarsejt egyesül*

Az (a_i, a_j) genotípusú zigóta létrejötte mint esemény a fenti két diszjunkt esemény összege. Az (a_i genotípusú hímivarsejt) és (a_j genotípusú női ivarsejt) szereplése pedig két független esemény (innentől (a_i hím) és (a_j női)). Ennek megfelelően a zigóta (a_i, a_j) genotípusának keletkezési valószínűsége:

$$P((a_i, a_j)\ \textit{zigóta}) = P(a_i\ \textit{hím})P(a_j\ \textit{női}) + P(a_j\ \textit{hím})P(a_i\ \textit{női}),$$
$$i = 1, \ldots, s,\ j = 1, \ldots, s.$$

A kezdeti feltételek miatt $P(a_k\ \textit{hím}) = P(a_k\ \textit{női})$, $k = 1, \ldots, s$, *ezért az utóbbi valószínűségekre bevezethetjük a közös* $P(a_k)$ *jelölést. Ekkor a fenti összefüggés:*

$$P((a_i, a_j)\ \textit{zigóta}) = 2P(a_i)P(a_j)$$

alakba írható, ahol $i = 1, \ldots, s$ *és* $j = 1, \ldots, s$. *Figyelembe véve, hogy az utódgeneráció egyedinek szomatikus genotípusa megegyezik azon zigóták genotípusával, melyek az egyedek kifejlődésére vezettek, ezért annak szomatikus genotípus valószínűségeloszlási polinomja:*

$$P^2(a_1)(a_1, a_1) + 2P(a_1)P(a_2)(a_1, a_2) + 2P(a_1)(a_3)(a_1, a_3) + \ldots +$$
$$2P(a_1)(a_s)(a_1, a_s) + P^2(a_2)(a_2, a_2) + \ldots + 2P(a_{s-1})P(a_s)(a_{s-1}, a_s) + P^2(a_s)(a_s, a_s).$$

Észrevehetjük, hogy ez az összeg egyenlő $(P(a_1)(a_1) + \ldots + P(a_s)(a_s))^2$*-nel, amennyiben az ivarsejtekre vonatkozó* $(a_i)(a_j)$ *eseményszorzatokat a szomatikus sejtekre vonatkozó* (a_i, a_j) *eseményekkel azonosítjuk.*

2. fejezet

A Hardy-Weinberg törvény - szakirodalmi feldolgozás

2.1. Panmiktikus szaporodás

Az 1.1 összefüggés szerint a három szomatikus genotípus előfordulási valószínűsége az utódgenerációban a fent leírt panmiktikus szaporodást követően mindig olyan, hogy alkalmas p-re a megfelelő arányok rendre $P(A,A) = p^2$, $P(A,a) = 2pq$ és $P(a,a) = q^2$, ahol $q = 1 - p$.

A következőkben ezen utódgeneráció utódgenerációjának szomatikus genotípus valószínűségeloszlását vizsgálom. Azt állítjuk,hogy ez is $p^2 : 2pq : q^2$ lesz. Tekintsük az első utódgeneráció ivarsejtjeinek genotípus eloszlását. A teljes valószínűség tétele szerint

$$P(A) = P(A|A,A)P(A,A) + P(A|A,a)P(A,a) + P(A|a,a)P(a,a) =$$
$$1 \cdot P(A,A) + 0,5 \cdot P(A,a) = p^2 + 0,5 \cdot 2pq = p(p+q) = p. \tag{2.1}$$

Az állítást ezzel be is láttuk, hiszen p már ugyanezen generációban q-t és így az utódgeneráció utódgenerációjának szomatikus genotípus eloszlását is egyértelművé

teszi.

De közvetlen számolással is belátható (szintén csak az utódgenerációra vonatkozóan)

$$P(a) = P(a|A,A)P(A,A) + P(a|A,a)P(A,a) + P(a|a,a)P(a,a) =$$
$$0,5 \cdot P(A,a) + 1 \cdot P(a,a) = 0,5 \cdot 2pq + q^2 = q(p+q) = q. \tag{2.2}$$

Ezt, tehát már az első szabad párosodással létrejött generáció egyensúlyi voltát mondja ki a következő tétel, melyet már 1908-ban leírtak.

2.1. Tétel (Hardy-Weinberg). *Pánmiktikus körülmények között a második, harmadik, ... utódgeneráció szomatikus genotípus vektora nem változik.*

2.2. Az egyensúly feltételei

A következőkben azt vizsgáljuk, hogy milyen feltételek mellett lesz egyensúly egy nem feltétlenül panmiktikus (szabad) szaporodással létrejött populáció és az utódgenerációja között. A nem ivari kromoszómán lévő lókuszhoz itt is két allél, a és A tartozik.

2.1. Állítás. *Annak, hogy egy populáció és az utódgenerációja között egyensúly legyen, szükséges és elégséges feltétele, hogy fennálljanak a következő összefüggések:*

$$P((a,a)) = \left(1 - (P((A,A)))^{\frac{1}{2}}\right)^2 \tag{2.3}$$
$$P((A,a)) = 2\left(P((A,A))^{\frac{1}{2}}\left(1 - P((A,A))^{\frac{1}{2}}\right)\right). \tag{2.4}$$

2.1. Bizonyítás. $P(A,A) + P(A,a) + P(a,a) = 1$ *miatt elegendő csak az egyik, például a* $P(A,A) : P(a,a)$ *aránnyal foglalkozni. Ahogy az előzőekben láthattuk, az ivarsejtek genotípus valószínűségei:*

$$P(A) = P(A|A,A)P(A,A) + P(A|A,a)P(A,a) = P(A,A) + 0,5P(A,a).$$

Az 1.1 összefüggés szerint az utódnemzedék (A, A) *genotípusa ekkor* $(P(A, A) + 0,5P(A, a))^2$. *Egyensúly esetén ennek meg kell egyeznie a szülői nemzedék* (A, A) *genotípusának előfordulási valószínűségével,* $P(A, A)$*-val*

$$(P(A, A) + 0,5P(A, a))^2 = P(A, A).$$

Innen

$$P(A, a) = 2(P(A, A))^{\frac{1}{2}}(1 - (P(A, A))^{\frac{1}{2}}).$$

Ezért

$$P(a, a) = 1 - (P(A, A) + P(A, a)) = 1 - P(A, A) - 2(P(A, A))^{\frac{1}{2}}(1 - (P(A, A))^{\frac{1}{2}}),$$

ami elemi meggondolások miatt átírható a következő alakba

$$P(a, a) = (1 - (P(A, A))^{\frac{1}{2}})^2. \tag{2.5}$$

Az egyensúly teljesülésének szükséges és elégséges feltétele tehát az 2.5 összefüggés teljesülése

2.3. Hardy-Weinberg egyensúly három allél esetén

Ebben a fejezetben azt vizsgálom, hogy három allél esetén mi a feltétele az egyensúlynak. Tartozzon így a szintén nem ivari kromoszómán lévő lókuszhoz három allél, a_1, a_2 és a_3.

2.2. Állítás. *Adott generációbeli egyensúlynak elégséges feltétele, hogy a szomatikus genotípus vektor a következő alakú legyen:*

$$(r^2(a_1, a_1), p^2(a_2, a_2), q^2(a_3, a_3), 2pr(a_1, a_2), 2qr(a_1, a_3), 2pq(a_2, a_3)), \tag{2.6}$$

ahol r*,* p *és* q *az* a_1*,* a_2 *és* a_3 *allélok előfordulási valószínűségei az adott generációban. Fennáll a* $p + q + r = 1$ *egyenlőség és* $r^2 = P(a_1, a_1)$

2.1. Megjegyzés. *Ha az a_3 allél előfordulási valószínűsége 0, akkor a kétallélos esetet kapjuk a 2.6 vektor $(r^2(a_1,a_1), p^2(a_2,a_2), 2pr(a_1,a_2))$ alakú lesz, ami $r = P(A,A)^{\frac{1}{2}}, p = 1 - r = 1 - P(A,A)^{\frac{1}{2}}$ megfeleltetéssel visszavezet a kétallélos esetre, melyre az összefüggést beláttuk. Tehát tényleg az előző pontban tárgyalt összefüggés általánosításáról van szó. Továbbá az összefüggés értelmében az egyik feltétel az, hogy az (a_i, a_i) (úgynevezett homozigóta) genotípus előfordulási valószínűségek négyzetgyökeinek összege 1 legyen.*

2.2. Bizonyítás. *Az a_1 allélra vonatkozó ivarsejt genotípus valószínűségeloszlási vektor, mely mindig előáll a következőképpen:*

$$P(a_1) = P(a_1|a_1,a_1)P(a_1,a_1) + P(a_1|a_1,a_2)P(a_1,1_2) + P(a_1|a_1,a_3)P(a_1,a_3) + 0,$$

esetünkben így írható:

$$P(a_1) = 1 \cdot r^2 + 0,5 \cdot 2pr + 0,5 \cdot 2qr = r^2 + pr + qr = r(r+p+q) = r.$$

Hasonlóan,

$$P(a_2) = p^2 + pr + pq = p(p+r+q) = p$$

és

$$P(a_3) = q^2 + qr + qp = q$$

Ezek szerint az utódgeneráció szomatikus genotípus valószínűségeloszlási vektora, mint az ivarsejt genotípus valószínűségeloszlási vektor négyzete:

$$(r(a_1)+p(a_2)+q(a_3))^2 = r^2(a_1,a_1)+p^2(a_2,a_2)+2pq(a_1,a_2)+2rq(a_1,a_3)+2pq(a_2,a_3),$$

ami megegyezik a szülőgeneráció szomatikus genotípus valószínűségi vektorával, ezzel pedig beláttuk az előbbi állítást

2.4. Hardy-Weinberg egyensúly k számú allél esetén

Ebben a fejezetben általánosítom az előző feltételeket és k számú allélra vizsgálom az egyensúly feltételét. Az allélok $a_1, a_2, \ldots, a_k$, $P(a_i)$ az a_i allél valószínűsége a

populáció reprezentáns ivarsejtjeinek körében, $i = 1, \ldots k$ és így $P(a_1) + \ldots P(a_k) = 1$.

2.3. Állítás. *Az egyensúly elégséges feltétele, hogy a szülőgenerációban a szomatikus genotípus valószínűségeloszlási polinomja a következő alakú legyen:*

$$(r_1^2(a_1, a_1), \ldots, r_k^2(a_k, a_k), \ldots, 2r_1r_2(a_1, a_2), \ldots, 2r_1, r_k(a_1, a_k), \ldots, r_{k-1}r_k(a_{k-1}, a_k)),$$

ahol $r_i^2 = P(a_i, a_i)$

2.3. Bizonyítás. *Az ivarsejt genotípusok valószínűségei:*

$$\begin{aligned} P(a_1) &= P(a_1|a_1, a_1)P(a_1, a_1) + P(a_1|a_1, a_2)P(a_1, a_2) + \ldots + P(a_1|a_1, a_k)P(a_1, a_k) + 0 \\ &= 1 \cdot r_1^2 + 0,5 \cdot 2r_1r_2 + \ldots + 0,5 \cdot 2r_1r_n = r_1^2 + \sum_{i=2}^{k} r_1r_i = r_1(r_1 + \sum_{i=2}^{k} r_i) = r_1. \end{aligned}$$

Hasonlóan látható be a többi ivarsejtgenotípusra is, így $P(a_i) = r_i$, *ahol* $i = 2, \ldots, k$. *Így az utódgeneráció szoamtikus genotípus valószínűségeloszlás vektora:*

$$(r_1(a_1) + \ldots + r_k(a_k))^2.$$

Négyzetre emeléssel visszajutunk a szülőgeneráció szomatikus genotípus valószínűségeloszlási vektorához. Ezzel az állítást bizonyítottuk.

3. fejezet

Egy, az előzőektől eltérő szaporodási modell

3.1. Leírás, feltételek

A következő modellben egy olyan (nem ideális) ivarsejtpopulációt vizsgálok, melyben összesen 40 darab, 8 - 8 a és 12 - 12 A genotípusú hím és női ivarsejt van. A hím ivarsejtek közül véletlenül és azonos valószínűséggel kiválasztunk egyet. Ha ez az ivarsejt A genotípusú, akkor $p = 0,2$ valószínűséggel elpusztul, viszont $q = 0,8$ valószínűséggel tovább él és megtermékenyít egy női ivarsejtet. Ha a hím ivarsejt a genotípusú, akkor $p = 0,3$ valószínűséggel pusztul el és $q = 0,7$ valószínűséggel megtermékenyít egy női ivarsejtet. Abban az esetben, ha a kiválasztott ivarsejt elpusztul, úgy újra választunk véletlen módon a megmaradt 19 ivarsejtből, mindaddig, amíg egy, a megtermékenyítésben részt vevő hím ivarsejt nem adódik, majd ezt feljegyezzük. A kiválasztott hím ivarsejthez ugyanilyen módon és ugyanilyen túlélési valószínűségekkel választunk egy női ivarsejtet. (Így háromféle genotípusú utódsejt jöhet létre: AA, Aa és aa.) Ezt a kiválasztást értelemszerűen addig folytatjuk, amíg el nem fogynak a hím vagy a női ivarsejtek. Feltesszük, hogy a megmaradó hím és

női ivarsejtekből biztosan keletkezik egy zigóta. A modell egyszerű és realisztikus, azonban az utódnemzedékbeli szomatikus genotípus eloszlás megadása elméleti síkon meglehetősen nehézkes lenne

A szimulációhoz először véletlen sorrendbe rendeztem a 20 hím ivarsejtet, majd ebben a sorrendben vettem egyesével őket, generáltam hozzájuk egy-egy véletlen számot, ami eldöntötte, hogy az ivarsejt elpusztul vagy részt vesz-e a megtermékenyítésben. Ugyanígy a női ivarsejteknél is. Az így kialakult hím-női ivarsejtpárok genotípusát jegyeztem fel. Majd az egész metódust megismételtem 40-szer, melynek során ehhez hasonló sorozatokat kaptam:

$$aA, Aa, aa, Aa, AA, aa, aA, aA, aA, AA, Aa, AA, AA, aa, aA,$$

ahol az első tag mindig a hímivarsejt, a második a női genotípusa. Ahogy láthatjuk, itt 15 zigóta keletkezett, míg, ha az egyes ivarsejtek biztosan résztvennének a megtermékenyítésben, akkor minden ilyen sorozatban 20 zigóta keletkezne. A továbbiakban aA és Aa között nem teszek különbséget. A lenti táblázat összefoglalja, hogy a különböző sorozatokban hány AA, Aa és aa genotípusú utódsejt keletkezett.

3.2. Eredmények

A következő táblázat összefoglalja az előző szimulált modell eredményeit.

	1	2	3	4	5	6	7	8	9	10	
Aa	8	7	5	7	8	8	8	3	7	8	69
aa	2	2	3	3	3	3	1	2	2	1	22
AA	6	3	6	4	7	5	4	6	4	7	52
$\sum$	16	12	14	14	18	16	13	11	13	16	

	11	12	13	14	15	16	17	18	19	20	
Aa	7	4	6	6	5	6	11	4	6	9	64
aa	3	3	3	2	1	3	2	4	3	2	26
AA	6	6	6	7	4	8	3	4	2	3	49
$\sum$	16	13	15	15	10	17	16	12	11	14	

	21	22	23	24	25	26	27	28	29	30	
Aa	6	5	6	7	5	6	8	7	8	6	64
aa	4	1	4	3	2	0	2	3	3	2	24
AA	4	6	7	4	7	6	4	5	6	4	53
$\sum$	14	12	17	14	14	12	14	15	17	12	

	31	32	33	34	35	36	37	38	39	40	
Aa	1	10	6	5	8	6	5	5	5	5	56
aa	4	1	0	4	4	3	4	3	2	1	26
AA	6	5	5	7	4	5	5	7	3	5	52
$\sum$	11	16	11	16	16	14	14	15	10	11	

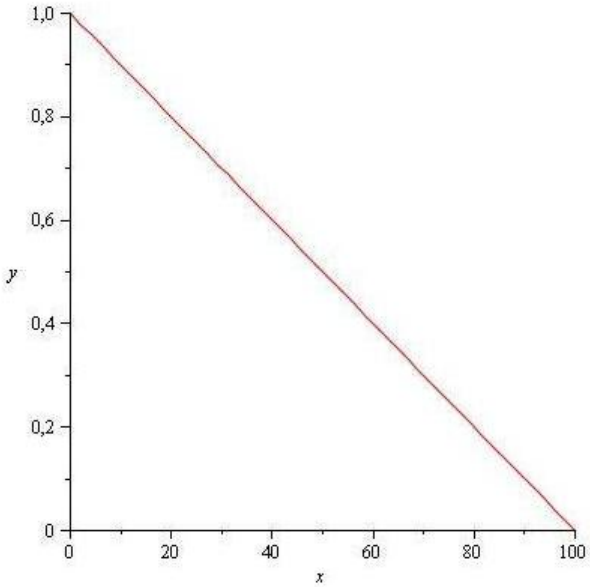
3.1. táblázat. Eredmények részletesen

Az első sorokban a sorozatok sorszáma, a második az egyes sorozatokban keletkezett Aa, a harmadik az aa, a negyedik az AA genotípusú utódsejtek száma, az utolsó sorokban pedig a sorozatban keletkezett összes utódsejtek, zigóták száma látható. Összesítve azt kapjuk, hogy az 557 keletkezett zigótából, 253 Aa ($45,42$ %), 98 aa ($17,60$ %) és 206 AA genotípusú ($36,98$ %). A kapott arányok tehát $0,3698 : 0,4542 : 0,1760$ ($AA : Aa : aa$). Visszatérve a Punett-táblára, kiszámíthatjuk, hogy mik lennének az elméleti arányok a panmiktikus modellben. A kiinduláskor az A és a genotípusok aránya $0,6 : 0,4$ így az utódgenerációban az $AA : Aa : aa$ arány $0,36 : 0,48 : 0,16$ lesz. Tehát észrevehetjük, hogy ez a szaporodási modell nem

térhet el sokban a panmiktikustól.

Valószínű, hogy a két modell azonos eredményre vezet. Erre a megállapításra tehát egyszerű, középiskolások által is használható szimulációs program segítségével jutottunk.

Előfordulhat, hogy az egyes genotípusok uralkodóvá válnak az adott populációban. Ezért fejlődésük során a populációk bizonyos mértékig önszabályozóvá váltak azért, hogy fenntartsák a genotípusok közötti megfelelő arányokat. A különböző genotípusú zigóták vizsgálata során észrevehetünk egy szabályt, miszerint ha egy bizonyos genotípusú zigóták túl nagy számban vannak jelen a populációban, úgy kisebb eséllyel válnak ivarérett egyeddé. Ezt nevezzük *denzitás-* vagy sűrűségfüggésnek. Ha a populáció egyedszáma n, az adott genotípusú (itt legyen például az AA genotípusú) zigóták száma pedig k, akkor az AA genotípusú zigótákból $p = 1 - \frac{k}{n}$ valószínűséggel fejlődik ki ivarérett egyed. Ezt mutatja a következő ábra is,



3.1. ábra. Denzitásfüggés

(kapcsolat adott genotípusú zigóták populációbeli aránya és azok ivarérett egyeddé fejlődésének valószínűsége között)

ahol az x tengely mutatja az AA genotípusú zigóták arányát a populációban lévő összes egyedhez viszonyítva (az egyszerűség kedvéért feltesszük, hogy a populációban 100 zigóta van, az AA genotípusú zigóták száma így 0 és 100 között változhat), az y tengely pedig annak a valószínűségét, hogy a zigótából ivarérett egyed fejlődik ki az adott arány mellett. Az ábra jól mutatja, hogy ha az AA genotípusú zigóták aránya nullához közeli, úgy majdnem biztosan ivarérett egyeddé fejlődnek, viszont ahogy nő a számuk és közelít a teljes populáció egyedszámához, úgy csökken kifejlődésüknek valószínűsége is. Vegyük például az ötödik sorozatot. Itt összesen 18 zigóta keletkezett, ebből 7 AA genotípusú, tehát itt az AA genotípus aránya $0,3889$ ($\frac{7}{18}$). Így a sűrűségfüggési modell szerint egy AA genotípusú zigótából $p = 0,6111$ valószínűséggel fejlődik ki ivarérett egyed, ami azt jelenti, hogy a hét AA genotípusú zigótából várhatóan csak négy ivarérett AA genotípusú egyed fejlődik ki.
Ezt minden sorozat minden genotípusára kiszámolva azt kapjuk, hogy várhatóan összesen csak 270 zigótából fejlődik ki ivarérett egyed, továbbá felírhatunk egy új $AA : Aa : aa$ arányt, ami most $0,393 : 0 : 414 : 0,193$ lesz (emlékeztetőül: a panmiktikus modellnél ez $0,36 : 0,48 : 0,16$ volt).

4. fejezet

Ivari kromoszómához kötött lókuszok

Az eddigiekben a nem ivari kromoszómákat vizsgáltuk, most azonban rátérünk az ivari kromoszómákon lévő lókusz alélljaira, melyeket szintén A és a jelöl. Az ivari kromoszómapárban az egyik ivar két egyforma kromoszómával vesz részt (XX) míg a másikban a két kromoszóma egymástól eltérő (XY). Az emberben és a legtöbb váltivarú állatban, és kétlaki növényben többnyire a hímivarú egyedekre jellemzők az eltérő alakú és többnyire különböző méretű X és Y ivari kromoszómák, míg a nőivarúakban az XX kromoszómapár található. Ahogy ezelőtt, most is felírható a hím és a női szomatikus genotípus valószínűségi vektor. Jelen esetben a női szomatikus genotípus valószínűségi vektor

$$(r_{AA}, 2s_{Aa}, t_{aa})$$

alakú, ahol $r_{AA} + 2s_{Aa} + t_{aa} = 1$. A továbbiakban ezt csak $(r, 2s, t)$-vel, a szülőgenerációt pedig $(r_0, 2s_0, t_0)$-val jelöljük.

A hím szomatikus genotípus valószínűségi vektor (itt egyetlen X kromoszóma van a szomatikus sejtben)

$$(p_A, q_a)$$

alakú, itt is fenáll a $p_A + q_a = 1$ egyenlőség, a továbbiakban ezt is csak (p, q)-val, a szülőgenerációt (p_0, q_0)-val jelöljük.

Így a női ivarsejt előfordulásának valószínűségei a következőképpen írhatók le (A_n női A-t jelöl):

$$\begin{aligned} P(A_n) &= P(A_n|AA_n)P(AA_n) + P(A_n|Aa_n)P(Aa_n) + 0 = r + 0,5 \cdot 2s = r + s \\ P(a_n) &= P(a_n|Aa_n)P(Aa_n) + P(a_n|aa_n)P(aa_n) + 0 = 0,5 \cdot 2s + t = s + t \end{aligned} \quad (4.1)$$

Vagyis a valószínűségi vektor, a női ivarsejtre nézve így írható:

$$((r+s)_A, (s+t)_a) = ((r+s), (s+t))$$

A hím ivarsejt előfordulására vonatkozó valószínűségek (A_h hím A-t jelöl):

$$\begin{aligned} P(A_h) &= P(A_h|A_h)P(A_h) = P(A_h) = p \\ P(a_h) &= P(a_h|a_h)P(a_h) = P(a_h) = q. \end{aligned} \quad (4.2)$$

Így a hím ivarsejtre felírt valószínűségi vektor

$$(p_A, q_a) = (p, q)$$

alakú, ugyanúgy mint a szomatikus sejtek esetén. A női utódgenerációban a szomatikus genotípus valószínűségek:

$$P(AA_n) = P(A_h)P(A_n) = p(r+s)$$

$$P(Aa_n) = P(A_n)P(a_h) + P(a_n)(A_h) = (r+s)q + (s+t)p$$

$$P(a, a_n) = P(a_h)P(a_n) = q(s+t).$$

Értelemszerűen, AA szomatikus genotípusú utódsejt hím A és női A egyesülésével keletkezik, ugyanilyen megfontolások alapján aa-ra hasonlóan, míg Aa kétféleképpen keletkezhet hím A, női a és hím a, női A egyesülésével, ezért itt ezek a valószínűségek összeadódnak. A fentieket összefoglalva felírhatjuk az utódgenerácóra vonatkozó női szomatikus genotípus valószínűségi vektort:

$$(p(r+s)_{AA}, (p(s+t) + q(r+s))_{Aa}, q(s+t)_{aa}),$$

ami a generációt is jelző írásmóddal $(r_1, 2s_1, t_1)$-nek felel meg.

Ugyanígy felírhatók a szomatikus genotípus valószínűségek a hím utódpopulációban is, itt azonban fontos megjegyezni, hogy mivel ivari kromoszómán lévő lókuszról van szó, a hímek ehhez tartozó A vagy a allélja csakis a női szomatikus sejtből származhat. A valószínűségek tehát:

$$P(A_h) = P(A_h|AA_n)P(AA_n) + P(A_h|Aa_n)P(Aa_n) + 0 = 1 \cdot r + 0,5 \cdot 2s = r + s$$

és

$$P(a_h) = P(a_h|Aa_n)P(Aa_n) + P(a_h|aa_n)P(aa_n) = 0,5 \cdot 2s + t.$$

Az eddigieket összefoglalva az első utódgeneráció hím és női szomatikus genoípus valószínűségi vektorának komponensei a következőképpen írhatók:

$$\begin{aligned} r_1 &= p_0(r_0 + s_0), \\ 2s_1 &= p_0(s_0 + t_0) + q_0(r_0 + s_0), \\ t_1 &= q_0(s_0 + t_0) \\ p_1 &= r_0 + s_0, \\ q_1 &= s_0 + t_0. \end{aligned} \tag{4.3}$$

Most, hogy ismerjük az első utódgeneráció szomatikus genotípus valószínűségi vektorának tagjait és látjuk, hogy ezen értékek csak a szülő generáció hasonló értékeitől függenek, kiszámíthatjuk a második, harmadik és így az összes ezt követő generációra is a valószínűségeket. Mindez tehát általánosan az n és $n+1$-edik generáció viszonylatában $(n = 0, 1, 2, \ldots)$:

$$r_{n+1} = p_n(r_n + s_n), \tag{4.4}$$

$$2s_{n+1} = p_n(s_n + t_n) + q_n(r_n + s_n), \tag{4.5}$$

$$t_{n+1} = q_n(s_n + t_n) \tag{4.6}$$

$$p_{n+1} = r_n + s_n, \tag{4.7}$$

$$q_{n+1} = s_n + t_n. \tag{4.8}$$

A következő fejezetben ezen értékek változását vesszük sorra a szülő generáció szomatikus genotípus valószínűségi vektorának függvényében, több példán keresztül.
Felírhatók még további összefüggések is, melyeket az eddigiekből kaphatunk meg. Így a 4.4 és a 4.7 összefüggések alapján:

$$r_{n+1} = p_n p_{n+1}, \tag{4.9}$$

4.5, 4.7 és 4.8 összefüggésekből

$$2s_{n+1} = p_n q_{n+1} + q_n p_{n+1}, \tag{4.10}$$

4.6 és 4.8 összefüggésekből pedig

$$t_{n+1} = q_n q_{n+1} \tag{4.11}$$

A szomatikus genotípus valószínűségi vektor tart egy határeloszláshoz a következők szerint. A 4.4 és 4.5 összefüggésekből megkapjuk a

$$\begin{aligned} 2r_{n+1} + 2s_{n+1} &= 2p_n(r_n + s_n) + p_n(s_n + t_n) + q_n(r_n + s_n) = \\ &= 2p_n(r_n + s_n) + p_n(1 - r_n - s_n) + (1 - p_n)(r_n + s_n) \end{aligned} \tag{4.12}$$

egyenlőséget, mely azért áll fenn, mert a megfelelő valószínűségek összegére $2s_n + r_n + t_n = 1$. A 4.12 formula átalakítható a következővé:

$$2r_{n+1} + 2s_{n+1} = p_n + r_n + s_n, \tag{4.13}$$

ami a 4.7 számú összefüggés miatt

$$2p_{n+2} = p_n + p_{n+1}$$

alakba is átírható. Továbbá, mivel feljebb láthattuk, itt a baloldalon $2r_{n+1}+2s_{n+1} = 2p_{n+2}$, a jobboldalon pedig $r_n + s_n = p_{n+1}$ áll, ezért a 4.13 összefüggés

$$2p_{n+2} = p_n + p_{n+1}$$

alakra hozható, melyet átrendezve a

$$2p_{n+2} - p_{n+1} = p_n$$

összefüggéshez jutunk. Innen

$$\begin{aligned} 2p_{n+2} - 2p_{n+1} &= p_n - p_{n+1} \\ p_{n+2} - p_{n+1} &= -\frac{1}{2}(p_{n+1} - p_n). \end{aligned} \tag{4.14}$$

Így egy rekurziós formulát kaptunk a p értékek különbségeire.

A továbbiakban ennek alapján foglalkozunk az A genotípusú hím szomatikus és egyben ivarsejtek p_n előfordulási valószínűségek konvergenciájával (a hímek körében). Ehhez bevezetjük a

$$p_{n+1} - p_n = \Delta p_n = \Delta_n \qquad n = 0, 1, 2, \ldots$$

jelölést, amivel a 4.14 formula átírható

$$\Delta_{n+1} = -\frac{1}{2}\Delta_n$$

alakra és hasonlóan

$$\begin{aligned} \Delta_n &= -\frac{1}{2}\Delta_{n-1}, \\ &\vdots \\ \Delta_2 &= -\frac{1}{2}\Delta_1, \\ \Delta_1 &= -\frac{1}{2}\Delta_0, \\ (\Delta_0 &= p_1 - p_0) \end{aligned}$$

Innen

$$\Delta_{n+1} = -\frac{1}{2}\left(-\frac{1}{2}\Delta_{n-1}\right) = -\frac{1}{2}\left(-\frac{1}{2}\left(-\frac{1}{2}\Delta_{n-2}\right)\right) = \cdots,$$

ami végül a

$$\Delta_{n+1} = \left(-\frac{1}{2}\right)^{n+1}\Delta_0 = \left(-\frac{1}{2}\right)^{n+1}(p_1 - p_0) \tag{4.15}$$

alakra hozható.

Tekintsük most a

$$p_n = p_0 + (p_1 - p_0) + (p_2 - p_1) + \cdots + (p_n - p_{n-1})$$

azonosságot. A különbségeket 4.15 alapján átírva a

$$\begin{aligned} p_n &= p_0 + \Delta_0 + \left(-\frac{1}{2}\right)\Delta_0 + \left(-\frac{1}{2}\right)^2\Delta_0 + \cdots + \left(-\frac{1}{2}\right)\Delta_0 = \\ &= p_0 + \Delta_0\left(1 + \left(-\frac{1}{2}\right) + \left(-\frac{1}{2}\right)^2 + \cdots + \left(-\frac{1}{2}\right)^{n-1}\right) \end{aligned}$$

összefüggéshez jutunk, ahol az egyenlőség jobboldalán lévő, Δ_0 utáni zárójelben egy mértani sorozat eleminek összege áll ($q = -\frac{1}{2}$), így ez felírható a következő alakban:

$$\frac{1 - \left(-\frac{1}{2}\right)^n}{1 - \left(-\frac{1}{2}\right)} = \frac{1 - \left(-\frac{1}{2}\right)^n}{1 + \frac{1}{2}} = \frac{2\left(1 - \left(-\frac{1}{2}\right)^n\right)}{3}.$$

Ezt visszahelyettesítve a

$$p_n = p_0 + \Delta_0 \frac{2\left(1 - \left(-\frac{1}{2}\right)^n\right)}{3}$$

formulát kapjuk. Így

$$\begin{aligned} \lim_{x\to\infty} p_n (= p_\infty) &= p_0 + \Delta_0 \lim_{x\to\infty} \frac{2\left(1 - \left(-\frac{1}{2}\right)^n\right)}{3} = p_0 + \frac{2}{3}\Delta_0 = \\ &= p_0 + \frac{2}{3}(p_1 - p_0) = \frac{1}{3}(p_0 + 2p_1), \end{aligned} \tag{4.16}$$

és értelemszerűen $q_\infty = 1 - p_\infty$. Ekkor $r_{n+1} = p_n p_{n+1}$ figyelembevételével

$$\lim_{n\to\infty} r_{n+1} = r_\infty = \lim_{n\to\infty} p_n p_{n+1} = p_\infty^2 = \frac{1}{9}(p_0 + 2p_1^2).$$

A 4.11 formula alapján:

$$t_{n+1} = q_n q_{n+1} = (1 - p_n)(1 - p_{n+1}).$$

Ekkor

$$\lim_{n\to\infty} t_{n+1}(= t_\infty) = (1 - p_\infty)^2 = q_\infty^2 = (1 - \frac{1}{3}(p_0 + 2p_1))^2,$$

végül

$$s_\infty = p_\infty q_\infty = \frac{1}{3}(p_0 + 2p_1)(1 - \frac{1}{3}(p_0 + 2p_1)).$$

Összefoglalva a határeseti genotípus valószínűségi vektorok:

$$(p_{\infty AA}^2, 2p_\infty q_{\infty Aa}, q_{\infty aa}^2) \tag{4.17}$$

$$(p_{\infty A}, q_{\infty a}). \tag{4.18}$$

Így láthatjuk, hogy $r_\infty, s_\infty, t_\infty, p_\infty$ és q_∞ csupán p_0-tól és p_1-től, $p_1 = r_0 + s_0$ miatt végtére is p_0-tól és $r_0 + s_0$-tól függ. Ahogy azt az előbbiekben láttuk, $r_0 + s_0$ az A allél előfordulási valószínűsége a szülőgeneráció női ivarsejtjeinek körében, p_0 pedig ugyanezen allél előfordulási valószínűsége a hím ivarsejtek körében (szintén a szülőgenerációban).

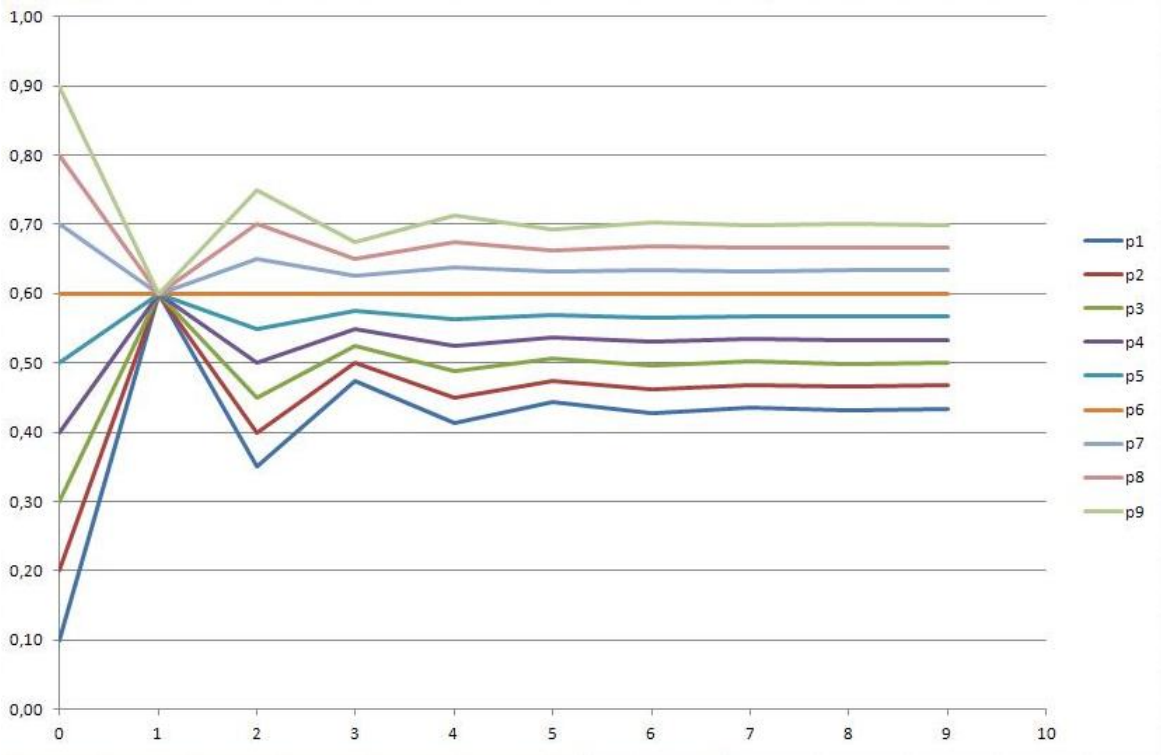
5. fejezet

A szomatikus genotípusok előfordulási valószínűségeinek sorozatai az r_0, s_0 és p_0 értékek függvényében

A következőkben az első 10 generáció szomatikus genotípus valószínűségi vektorának komponenseit vizsgáljuk. Megnézzük mi történik akkor, ha a p_0, r_0 és s_0 valószínűségek közül kettőt rögzítünk és csak a harmadikat változtatjuk. A grafikonok pontokból állnak, csak a jobb áttekinthetőség céljából kötöttem őket össze egy egyenessel.

5.1. A p_n értékek változása

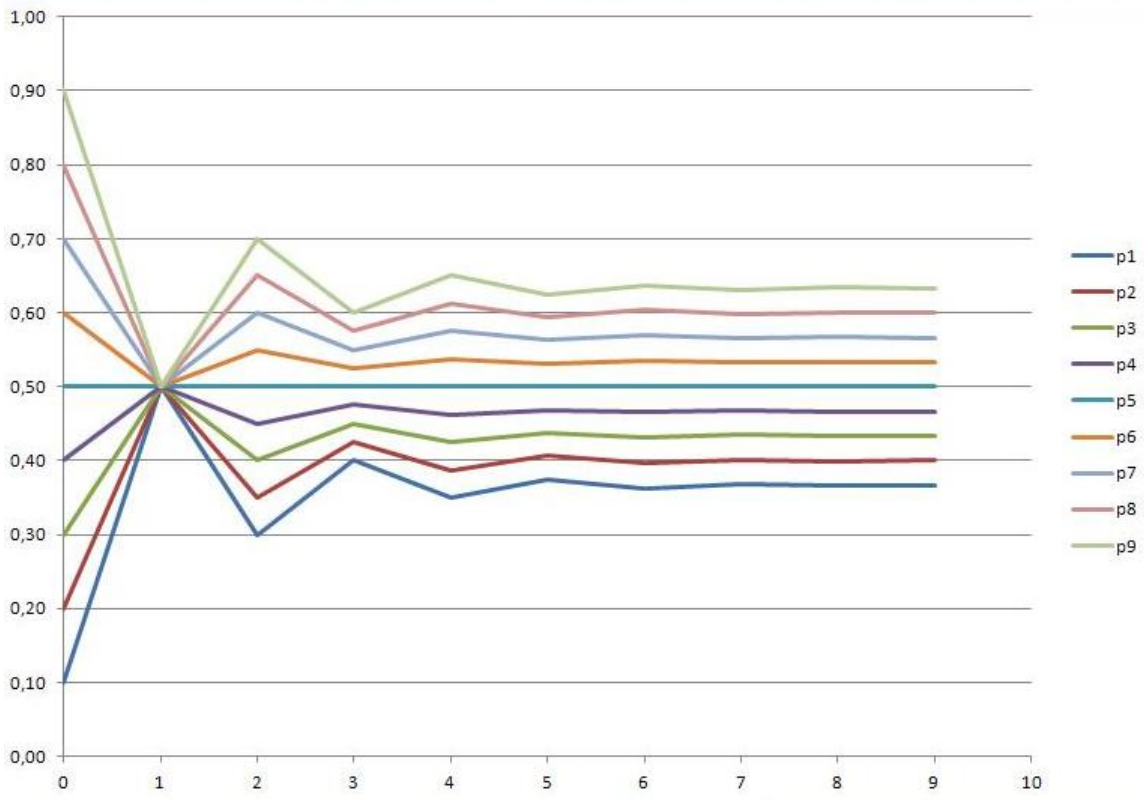
Ebben a részben a p_n értéket vizsgáljuk p_0 függvényében különböző r_0 és s_0 értékek mellett Az 5.1. ábra azt mutatja meg, hogyan változik p_n, ha $r_0 = 0,5$, $s_0 = 0,1$ és $t_0 = 0,3$ állandósága mellett p_0 $0,1$-től $0,9$-ig változik (persze így q_0 is változik, mivel $p_0 + q_0 = 1$). Így ezen az ábrán összesen kilenc grafikont láthatunk.



5.1. ábra. A p_n értékek alakulása p_0 függvényében, $r_0 = 0,5$, $s_0 = 0,1$

A legalsó (kék) vonal mutatja azt az esetet, ahol $p_0 = 0,1$, a felette lévőnél $p_0 = 0,2$, míg a legfelsőnél $p_0 = 0,9$. Láthatjuk, hogy alulról a hatodik grafikon konstans (ez a $p_0 = 0,6$-os esetet mutatja) és az egyes grafikonok annál simábbak és annál hamarabb "beállnak", minél közelebb vannak ehhez a grafikonhoz, tehát a $0,6$ közeli p_0 értékek esetén.

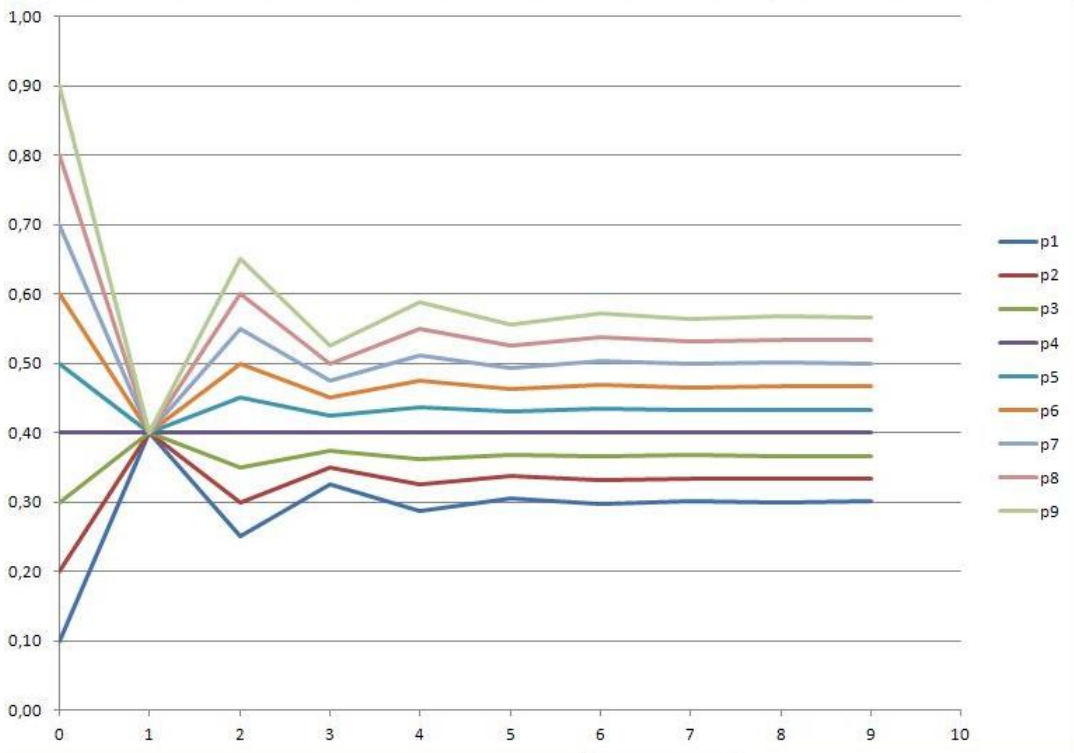
Az 5.2. ábra szintén a p_n értékeket mutatja, de most $r_0 = 0,3$ $s_0 = 0,2$ és $t_0 = 0,3$ esetén. Az előző ábrához hasonlóan itt is kilenc grafikont láthatunk, mivel p_0 itt is a $0,1$, $0,2$, $\ldots$, $0,9$ értékeket veszi fel.



5.2. ábra. A p_n értékek alakulása p_0 függvényében, $r_0 = 0,3$, $s_0 = 0,2$

Láthatjuk, hogy itt $p_0 = 0,5$-nél kapunk konstans egyenest, és az ehhez közeli grafikonok a simábbak.

Most megnézünk egy harmadik esetet is, itt is a p_n értékeket láthatjuk, itt azonban $r_0 = 0,1$ és $s_0 = 0,3$ feltételek esetén. Ezen kiindulási r_0 és s_0 valószínűségek mellett, a $p_0 = 0,4$-es esetben kapunk konstans grafikont p_n-re.



5.3. ábra. A p_n értékek alakulása p_0 függvényében, $r_0 = 0,1$, $s_0 = 0,3$

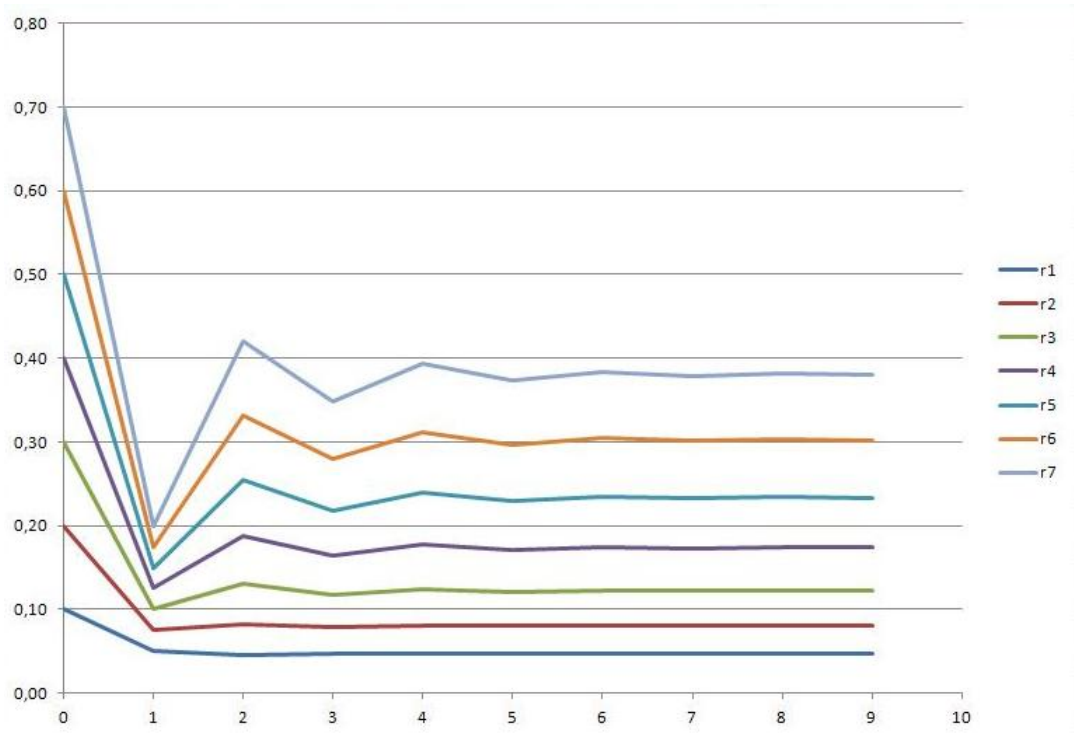
A három ábrát együtt vizsgálva, a konstans grafikon helyét leszámítva nem látunk köztük nagy eltéréseket.

5.2. Az r_n értékek változása

Az előző részhez hasonlóan most is a három valószínűség közül kettőt hagyunk fixen és csak egyet változtatunk, itt r_0-t. Először három olyan grafikonegyüttest láthatunk, melyeknél $s_0 = 0,1$ és a p_0 értékeket rendre a $0,25$-ra, $0,5$-re és $0,75$-ra állítottuk (így q_0 rendre $0,75$, $0,5$ és $0,25$ lesz). Mivel s_0-t fixen hagyjuk, ezért t_0 r_0-val együtt változik. Láthatjuk, hogy az $x = 0$ pontban, tehát a szülő generációban r_0 $0,1$-től $0,7$-ig vesz fel értékeket.

Az 5.4. ábránál megfigyelhetjük, hogy minél kisebb r_0, a grafikonok annál simábbak.

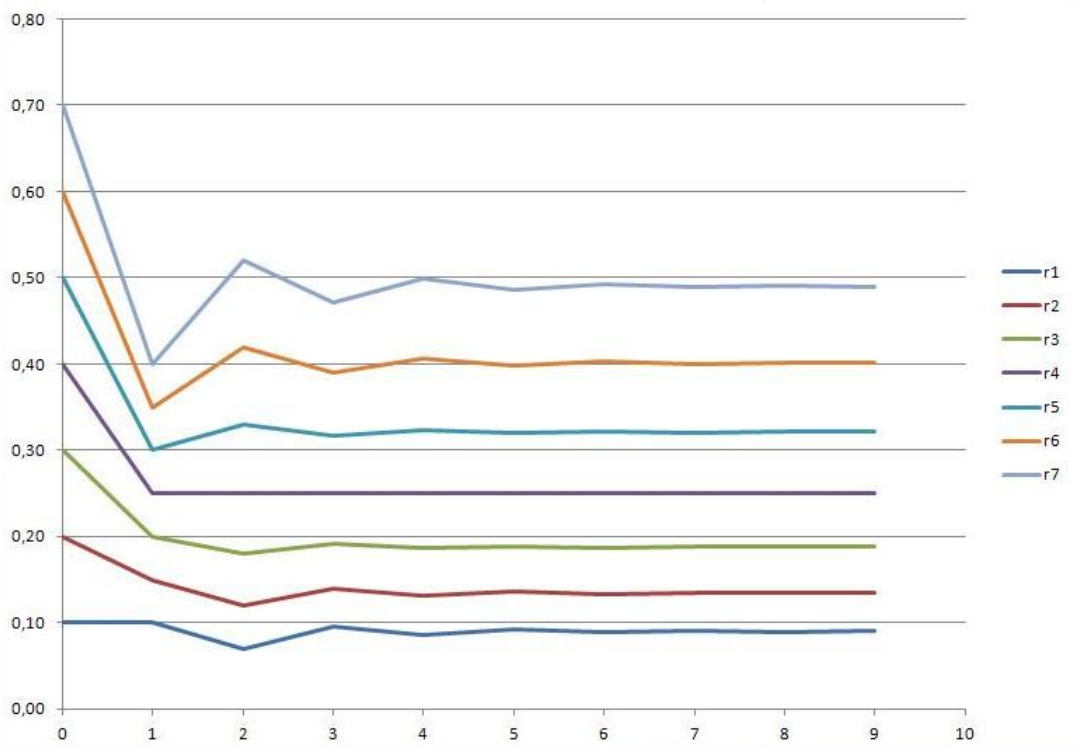
(Emlékeztetőül: itt $p_0 = 0,25$.)



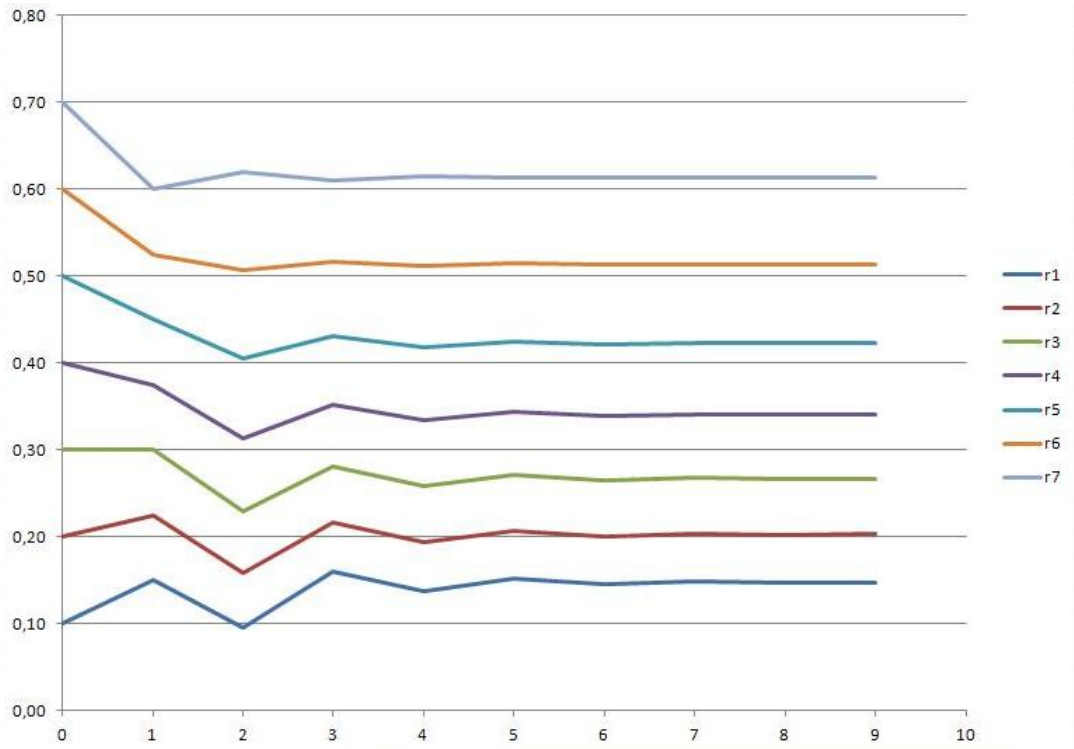
5.4. ábra. A r_n értékek alakulása r_0 függvényében, $p_0 = 0,25$, $s_0 = 0,1$

Az 5.5. ábra esetében, ahol $p_0 = 0,5$, észrevehetjük, hogy alulról a negyedik grafikon ($r_0 = 0,4$) $x = 1$-től konstans és alatta simább grafikonokat találhatunk.

Az 5.6. ábránál, amelynél $p_0 = 0,75$, azt tapasztaljuk, hogy az első esettől eltérően, itt a nagyobb r_0 értékek esetén kapunk simább grafikonokat.



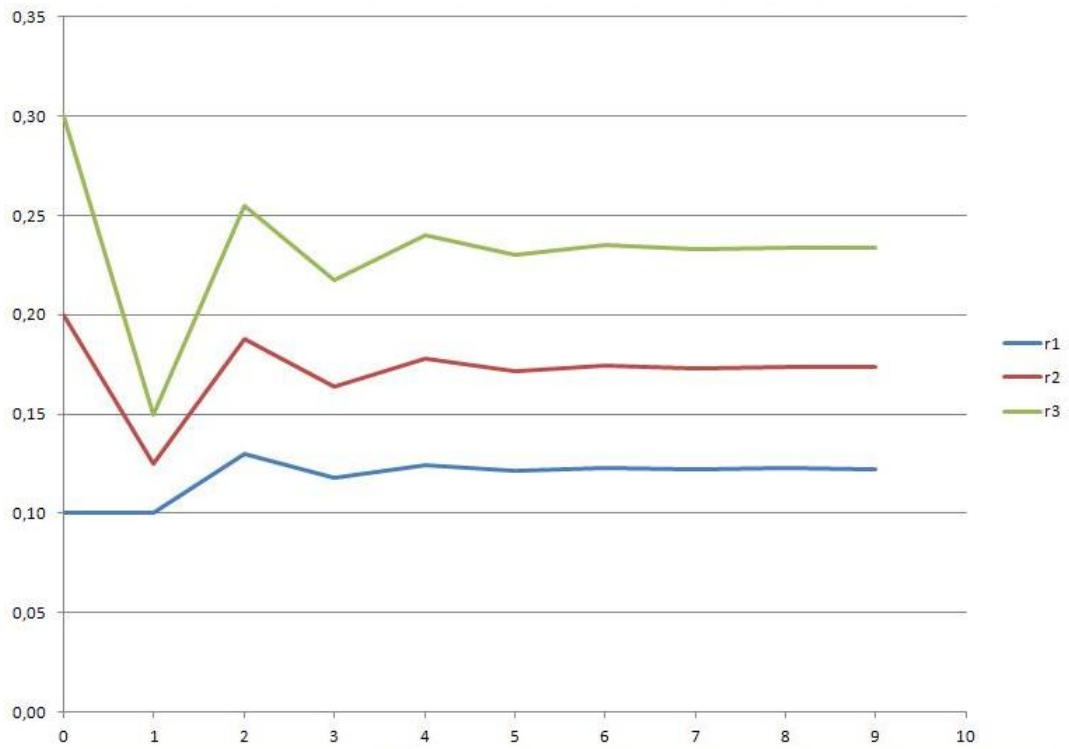
5.5. ábra. A r_n értékek alakulása r_0 függvényében, $p_0 = 0,5$, $s_0 = 0,1$



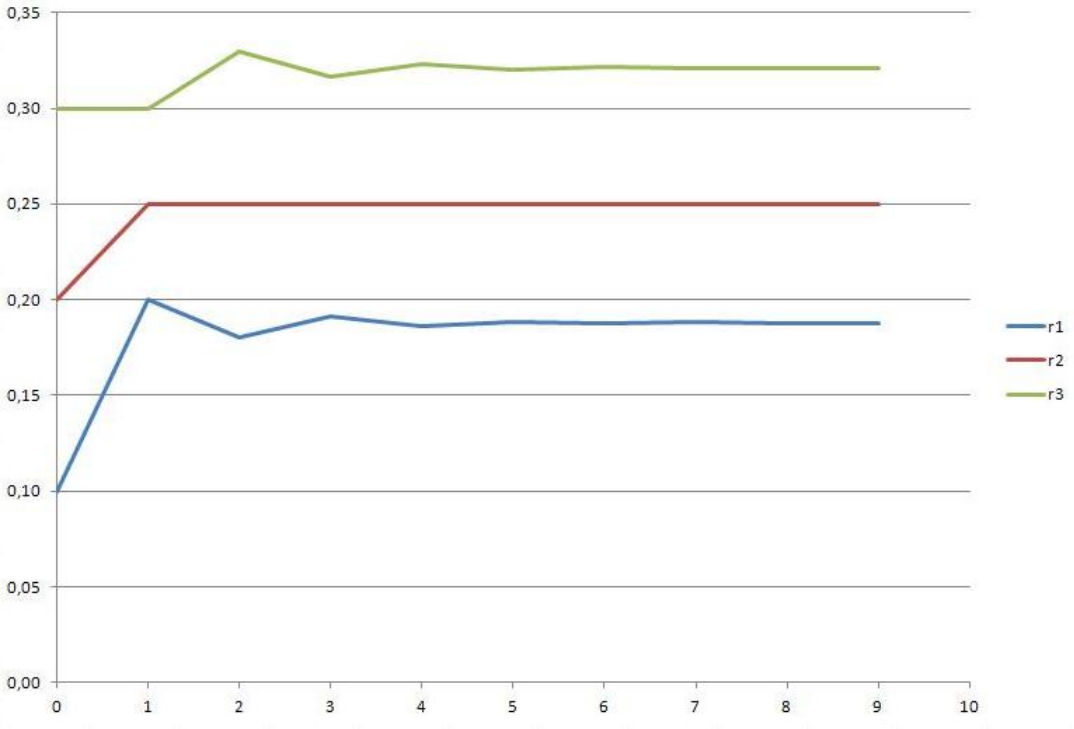
5.6. ábra. Az r_n értékek alakulása r_0 függvényében, $p_0 = 0,75$, $s_0 = 0,1$

Továbbá, az 5.4.-5.6. ábrák grafikonjait együtt tekintve azt láthatjuk, hogy minél nagyobb p_0 mellett vizsgáljuk r_n-t, az annál simább lesz.

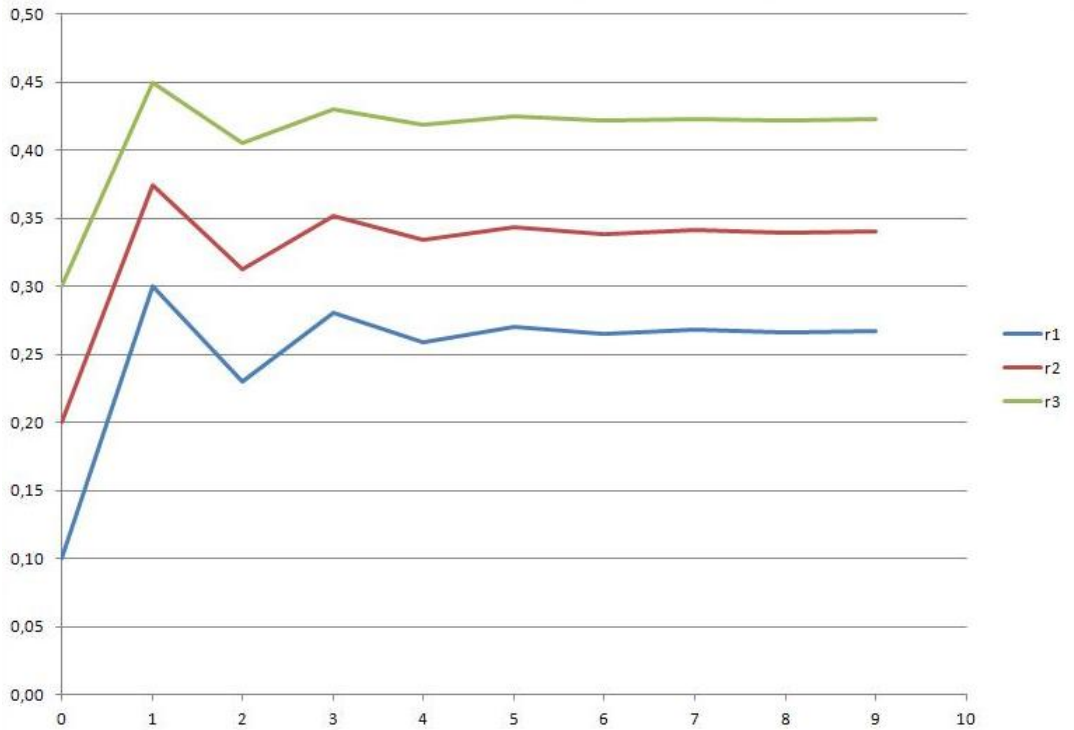
A következőkben három olyan ábrát láthatunk, amelyeknél szintén r_0 változik, $s_0 = 0,3$ és p_0 rendre a $0,25$, $0,5$ és $0,75$ értékeket veszi fel. Itt az 5.7. ábrán láthatjuk a legsimább grafikonokat, tehát $p_0 = 0,5$ esetén, sőt itt az $r_0 = 0,2$-es esetben az $x = 1$-től r_n konstans.
Az 5.8. ábrán, $p_0 = 0,25$ esetén a grafikonok annál simábbak, minél kisebb r_0.
Az 5.9. ábrán lévő grafikonok, itt $p_0 = 0,75$, pedig nagyjából egyformák.



5.7. ábra. Az r_n értékek alakulása r_0 függvényében, $p_0 = 0,25$, $s_0 = 0,3$



5.8. ábra. Az r_n értékek alakulása r_0 függvényében, $p_0 = 0,5$, $s_0 = 0,3$



5.9. ábra. Az r_n értékek alakulása r_0 függvényében, $p_0 = 0,75$, $s_0 = 0,3$

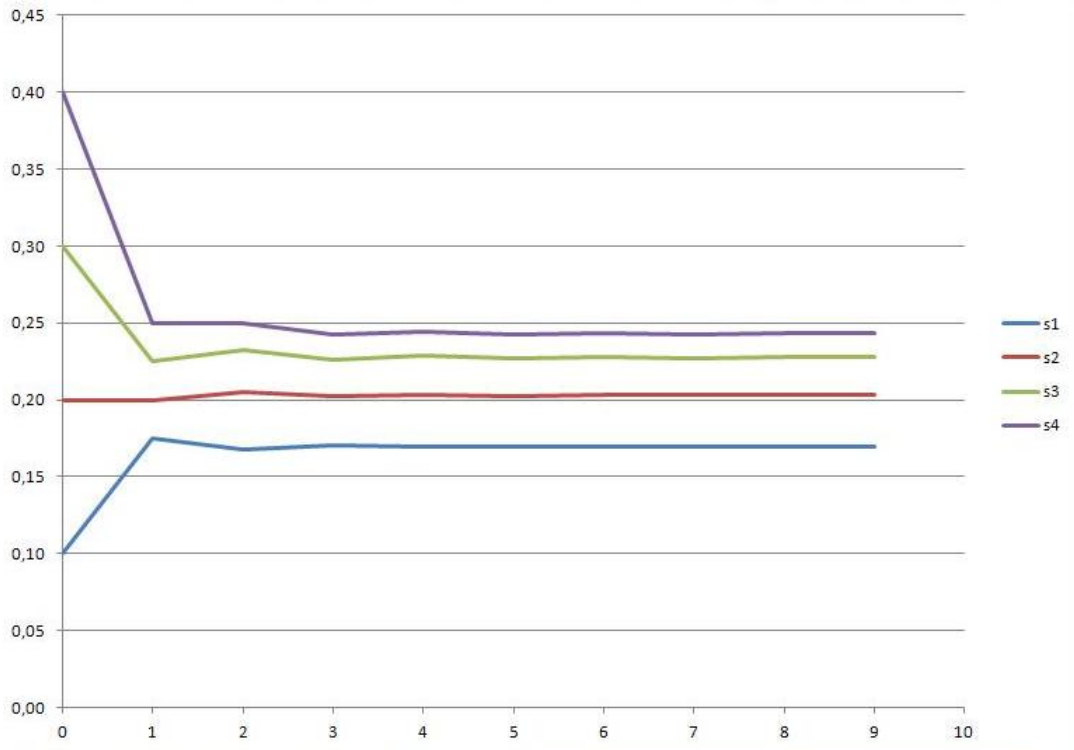
5.3. Az s_n értékek változása

A következőkben megnézzük a harmadik esetet is, amikor s_0-át változtatjuk fix p_0 és r_0 mellett. Itt is három ábra látható, melyeknél r_0 végig $0,1$ lesz és p_0 értéke sorban $0,25$, $0,5$ és $0,75$ lesz.

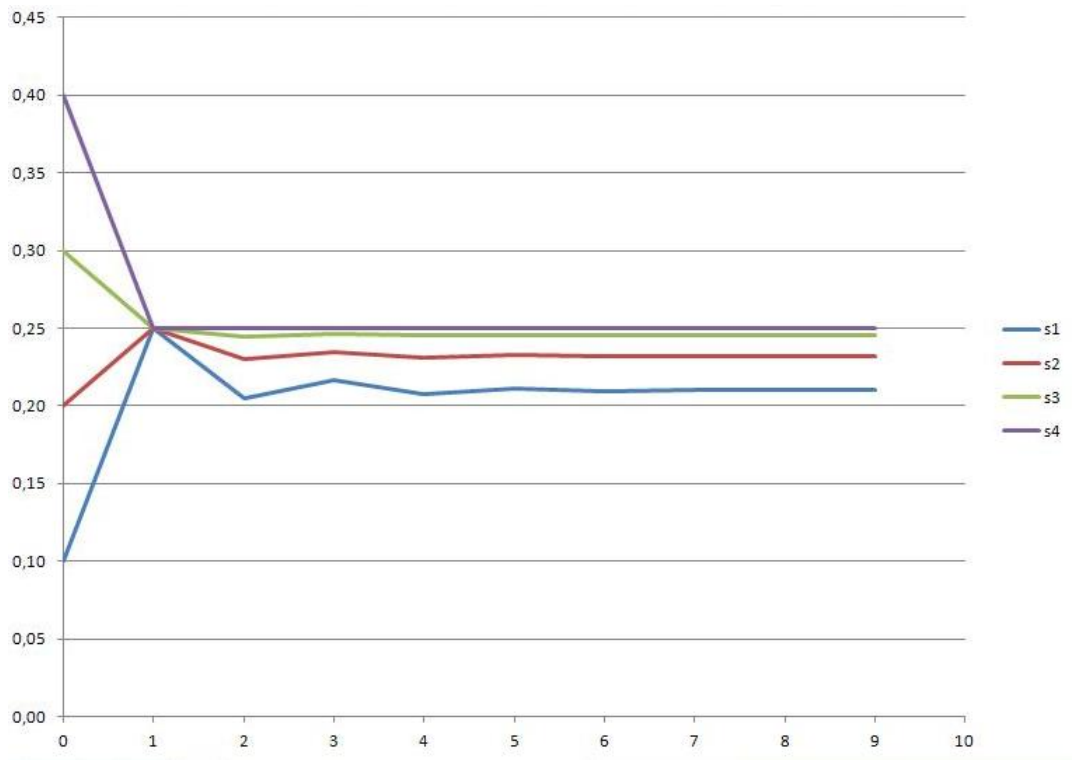
Az 5.10. ábra egy olyan esetet mutat, ahol $p_0 = 0,25$ és $r_0 = 0,1$ és s_0 különböző értékeket vesz fel $0,1$-től $0,4$-ig. Láthatjuk, hogy s_n képe akkor a legsimább mikor $s_0 = 0,2$, ez alulról a második grafikon.

Az 5.11. ábránál észrevehetjük, hogy a grafikonok annál simábbak, minél nagyobb s_0 értéke, sőt az $s_0 = 0,4$-es esetnél az első lépés után s_n már konstans.

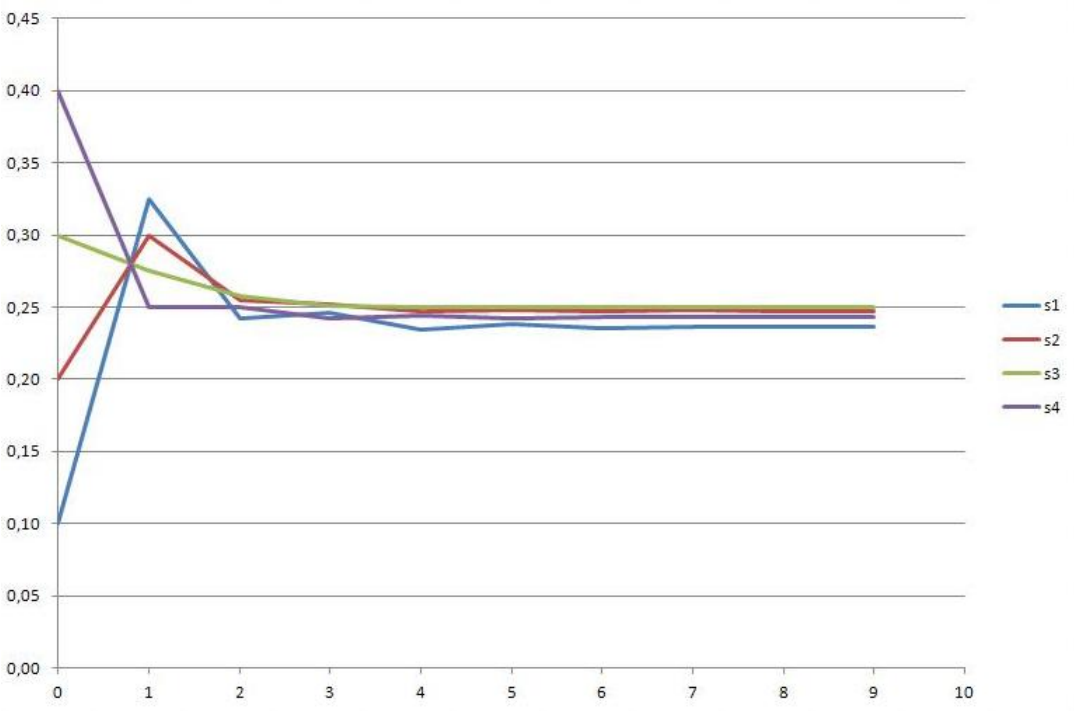
Az 5.12. ábránál, ahol $p_0 = 0,75$, az $s_0 = 0,3$-es grafikon a legsimább és az ettől legtávolabb lévő, ahol $s_0 = 0,1$ áll be a legkésőbb.



5.10. ábra. Az s_n értékek alakulása s_0 függvényében, $p_0 = 0,25$, $r_0 = 0,1$



5.11. ábra. Az s_n értékek alakulása s_0 függvényében, $p_0 = 0,5$, $r_0 = 0,1$



5.12. ábra. Az s_n értékek alakulása s_0 függvényében, $p_0 = 0,75$, $r_0 = 0,1$

A három ábrát együtt vizsgálva azt fedezhetjük fel, hogy a grafikonok annál közelebb vannak egymáshoz, minél nagyobb p_0 értéke. A végső pontok között, ahová konvergálnak, tehát egyre kisebb a különbség. Ha az r_0-nak nagyobb értékeket adunk, akkor értelemszerűen egyre kevesebb lehetőségünk van s_0 értékének megadásakor, mivel, ahogy a korábbiakban láthattuk $r_0 + 2s_0 + t_0 = 1$, ezért csak ezt az egy esetet mutattam be ebben a részben.

6. fejezet

Összefoglalás

Dolgozatom első fejezetében olyan alapvető biológiai és populációdinamika fogalmakat vezettem be, melyek elengedhetetlenek a további fejezetek értelmezéséhez.

A második fejezetben a Hardy-Weinberg egyensúlyról és annak feltételeiről írtam kettő, három és végül k számú allél esetén. A harmadik fejezetben bemutattam egy egyszerű, szimulált szaporodási modellt, majd elemeztem a kapott eredményeket. A fejezet végén egy új fogalmat is bevezettem, a populációk denzitás függését, majd ennek tükrében újra értékeltem a szimuláció által kapott eredményeket.

A dolgozatom eddigi fejezeteiben a nem ivari kromoszómákhoz kötött lókuszokhoz tartozó allélokról írtam, innentől, a negyedik fejezettől, rátértem az ivari kromoszómákhoz kötött lókuszokra. A negyedik fejezetben különböző egyszerű összefüggéseket mutattam be a szülő és utódgeneráció hím és női szomatikus genotípus valószínűségi vektorának komponensei között.

Ezeket az összefüggéseket felhasználva az ötödik fejezetben megvizsgáltam, hogyan változnak ezek a komponensek az első 10 generációban.

Köszönetnyilvánítás

Köszönettel tartozom a témavezetőmnek, Izsák Jánosnak a dolgozatom elkészítésében nyújtott segítségéért és végtelen türelméért.
Köszönöm továbbá Oláh Tibornak és Krasnyánszki Lászlónak a programozásban nyújtott segítségét.

Nyilatkozat

A szakdolgozat szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló munkám eredménye, saját szellemi termékem, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Irodalomjegyzék

[1] Ács Tamás, Csaba György, Kiszely György, Szabó Gábor: Biologia *Medicina*, 1970. A populációgenetika alapjai

[2] Günter Vogel, Harmut Angermann: Biológia *Springer Hungarica*, 1992. Mikroevolúció, Populációk

[3] Izsák János, Juhász-Nagy Pál, Varga Zoltán: Bevezetés a biomatematikába *Tankönyvkiadó*, 1981. Valószínűségszámítás, Példák