

# Elemi szelekciós elmélet

Meszéna Géza

2020. febr. 25.

## 1. Exponenciális növekedés, szelekció és reguláció

Állandó körülmények között egy populáció létszáma exponenciálisan változik, hisz úgy a születések, mint a halálozások száma arányos a populáció létszámával. Két populáció közül az, amelyiknek az exponenciális növekedési rátája nagyobb, exponenciálisan túlnövi a másikat. Ez nevezzük szelekciónak: a nagyobb fitnessű populáció legyőzi a kisebb fitnessűt – ahol a „fitness” fogalma nem jelenthet mást, mint a növekedési rátát.

Exponenciális növekedéssel azonban a populáció létszáma igen hamar irreálisan nagyra nőne – hacsak nem pontosan nulla a növekedés rátája. A valóságban tehát hosszú távon sohasem teljesülhet az „állandó körülmények” feltétele. A populációk növekedése egészen biztosan visszahat a körülményekre, s ezzel a növekedési rátákra – méghozzá úgy, hogy a hosszútávú növekedési ráták nullák legyenek. Ezt a visszacsatolást nevezzük populáció-szabályozásnak/regulációnak.

A szelekció elméletét először mégis rögzített körülmények között érdemes tárgyalni: ezt tesszük az alábbiakban. Az természetesen nem bizonyos, hogy a pillanatnyi szelekciós/fitness előny végleges előny, nem bizonyos, hogy a pillanatnyilag előnyösebb változat végül is kiszorítja a vetélytársát. Sok olyan eset lehet, amikor ez a probléma nem lép fel, mert a visszacsatolás a szelekciós viszonyokat nem befolyásolja. Máskor pedig pont esetleg pont a visszacsatolás a lényeg – de ekkor is a rögzített körülmények esetéből érdemes kiindulni.

Az alábbiakban a rögzített körülményekre vonatkozó elemi szelekciós esetet tárgyaljuk először klonális, utána szexuális szaporodás esetére – az utóbbi esetben egyetlen lókuszt esetére szorítkozva. Jellemzően először felírom az állítást, azután indoklom.

## 2. Klonális reprodukció

### 2.1. Változatok

$L$  számú változat (replikátor, szaporodási egység) versenyez egymással. Felteesszük, hogy a változatok hibamentesen replikálódnak (azaz nincs mutáció) és az egyes változatok populációi homogének. Legyen az  $i$ -edik változat létszáma  $n_i$ . Ekkor az összlétszám

$$n = \sum_{i=1}^L n_i, \quad (1)$$

az egyes változatok gyakorisága pedig

$$p_i = \frac{n_i}{n}. \quad (2)$$

Természetesen,

$$\sum_{i=1}^L p_i = 1. \quad (3)$$

A gyakoriságokkal súlyozott átlagokat felülvonással fogjuk jelölni. Jelölje az  $i$ -edik változat egy tetszőleges jellemzőjét  $x_i$ . Ekkor az  $x$  értékek populációs átlaga

$$\bar{x} = \sum_i p_i x_i \quad (4)$$

## 2.2. Szelekció folytonos időben

### 2.2.1. Populációnövekedés

A szelekció folyamatát külön diszkutáljuk folytonos és diszkrét idő esetére. Folytonos időben az egyes változatok növekedését leíró egyenletek:

$$\frac{dn_i}{dt} = r_i n_i, \quad (5)$$

ahol  $r_i$  az  $i$ -edik változat növekedési rátája, azaz fitnessze.

Ha az  $r_i$  növekedési ráták állandóak, akkor a (5) differenciálegyenlet megoldása

$$n_i(t) = n_i(0)e^{r_i t}, \quad (6)$$

ami  $r_i > 0$  esetén a populáció exponenciális növekedését,  $r_i < 0$  esetén pedig exponenciális csökkenését – végső soron kihalását – írja le. Magától értetődő, hogy az  $r_i > 0$  helyzet nem tud korlátlan ideig fennmaradni, előbb-utóbb a populáció növekedésének valamilyen okból (például a tápanyag fogytával) le kell csökkentenie a növekedési rátát.

Azt kell tehát értenünk, hogy noha az  $r_i$  növekedési ráták jellemzően nem állandóak, s így a (6) exponenciális növekedés tartósan nem valósul meg, a (5) differenciálegyenletet általánosan érvényesnek tekinthetjük azzal, hogy  $r_i$  a *mindenkori* növekedési rátát jelenti.

Az összlétszám változását így írhatjuk le:

$$\frac{dn}{dt} = \sum_i \frac{dn_i}{dt} = \sum_i r_i n_i = n \sum_i p_i r_i = \bar{r} n. \quad (7)$$

Azaz: az összlétszám növekedését az átlagos növekedési ráta szabja meg. Itt fontos érteni, hogy a  $p_i$  relatív gyakoriságok időben változnak, s így  $\bar{r}$  még akkor sem lenne egy konstans érték, ha az  $r_i$  növekedési ráták azok lennének.

### 2.2.2. A legegyszerűbb szelekciós egyenlet

Két populáció létszamarányának változását így írhatjuk fel:

$$\frac{d}{dt} \left( \frac{n_i}{n_j} \right) = (r_i - r_j) \frac{n_i}{n_j}, \quad (8)$$

ahol az (5) egyenletet és a hányados deriválási szabályát alkalmaztuk. (Ellenőrizd!) Azaz: a létszámok aránya a nagyobb növekedési rátájú (fitnessű) populáció javára változik.

Amennyiben az  $r_i - r_j$  fitness-különbség állandó (ami akkor is megtörténhet, ha a fitnesszok külön-külön nem állandóak) az egyenlet megoldása

$$\frac{n_i(t)}{n_j(t)} = \frac{n_i(0)}{n_j(0)} e^{(r_i - r_j)t}, \quad (9)$$

ami természetesen az (6) egyenletből is következik annak érvényessége esetén.

Azaz, állandó fitness-különbség esetén a létszámok aránya exponenciálisan változik a nagyobb fitnessű változat javára. A magasabb fitnessű változat így exponenciálisan kiszorítja az alacsonyabb fitnessűt. Relatív értelemben ez akkor is így van, ha az alacsonyabb fitnessű változat növekedési rátája is pozitív, s így mindkét változat növekedésben van.

### 2.2.3. Replikátor egyenlet

A gyakoriságok változását a relatív létszámok változásával is leírhatjuk:

$$\frac{dp_i}{dt} = (r_i - \bar{r})p_i, \quad (10)$$

amely összefüggést replikátor-egyenletnek nevezzük. Jelentése világos: egy változat gyakorisága nő, vagy csökken, a szerint, hogy fitnessze nagyobb, vagy kisebb az átlagos fitnessnél. Az összefüggés érvényességét ismét a hányados-deriválási szabállyal ellenőrizhetjük, ha az időderiváltakba az (5-7) összefüggéseket helyettesítjük:

$$\frac{dp_i}{dt} = \frac{d}{dt} \frac{n_i}{n} = \frac{r_i n_i n - n_i \bar{r} n}{n^2} = r_i p_i - \bar{r} p_i, \quad (11)$$

ami pont a replikátor-egyenlet.

Történetileg a (10) replikátor-egyenletet játékelméleti kontextusban szokás használni a következő értelmezéssel. Feltételezzük, hogy miközben replikátoraink szaporodnak, összlétszámukat egy „kényszer”-rel állandónak tartjuk. Ehhez az szükséges, hogy az összlétszám egyébkénti változását a populációt  $\bar{r}$  rátával való hígításával ellensúlyozzuk. Ez a hígítás lesz a replikátor-egyenlet második tagja. Valójában a „hígítás” matematikailag egyenértékű azzal, amikor csak a relatív gyakoriságokat vizsgáljuk - ami a populációgenetikában szokásos szemléletmód.

### 2.2.4. Szelekció két változat esetében

Mivel a populációgenetikába nagyon gyakran használjuk, érdemes külön kiírni a gyakoriság változását két változat, azaz  $L = 2$  esetére. Erre az esetre be szokás vezetni a  $p = p_1$ ,  $q = p_2 = 1 - p$ ,  $s = r_1 - r_2$  jelöléseket. Ezekkel a replikátor-egyenlet az alábbi alakba írható:

$$\frac{dp}{dt} = (r_1 - r_2)p(1 - p) = spq. \quad (12)$$

(A  $p = p_1$ -re felírt replikátor-egyenletbe  $\bar{r} = pr_1 + (1 - p)r_2$ -t helyettesítettünk.) A formula érthető: a szelekciót a fitness különbség hajtja, de leáll amikor akár  $p$ , akár  $q$  nullává válik. Az utóbbi esetekben ugyanis már csak egyetlen változat van jelen.

### 2.2.5. Price egyenlet

Ha az egyes változatokhoz az  $x_i$  állandó mennyiségeket rendeljük (pl. az  $i$ -edik változat testmérete), akkor az átlagos  $x$  változását az

$$\frac{d\bar{x}}{dt} = \text{Cov}(r, x) \quad (13)$$

Price egyenlet adja meg. Itt

$$\text{Cov}(r, x) = \overline{(r - \bar{r})(x - \bar{x})} \quad (14)$$

az  $r$  és  $x$  mennyiségek kovarianciája (lásd Függelék, a felülvonás továbbra is a populációra való átlagolást jelöli). Ez a mennyiség pozitív (negatív), ha az átlagosnál nagyobb  $x_i$ -hez jellemzően az átlagosnál nagyobb (kisebb)  $r_i$  tartozik. Azaz: a szelekció akkor növeli az átlagos  $x$  értéket, ha  $x$  pozitívan korrelál a fitnesszel. Az egyenlet levezetése:

$$\frac{d\bar{x}}{dt} = \sum_i \frac{dp_i}{dt} x_i = \sum_i p_i (r_i - \bar{r}) x_i = \quad (15)$$

$$= \sum_i p_i (r_i - \bar{r}) (x_i - \bar{x}) = \quad (16)$$

$$= \overline{(r - \bar{r})(x - \bar{x})}. \quad (17)$$

Itt az utolsó tényezőt azért bővíthettük ki az  $\bar{x}$  taggal, mert  $\sum_i p_i (r_i - \bar{r}) = 0$  (azaz az átlagtól való eltérés átlaga zérus). A kapott kifejezés pont  $r$  és  $x$  kovarianciájának definíciója.

### 2.2.6. Lande egyenlet és a direkcionális evolúció sebessége

Tegyük most fel, hogy az  $r$  növekedési ráta lineáris függvénye az  $x$  tulajdonságnak:

$$r_i = A + Dx_i, \quad (18)$$

ahol a  $D$  együttható a „fitnessz tájkép” (az  $r$  fitnessz az  $x$  tulajdonság függvényében) meredeksége. Ebből  $r$  és  $x$  kovarianciája:

$$\text{Cov}(r, x) = \text{Var}(x)D. \quad (19)$$

(Miért?) Ily módon az  $x$  tulajdonság evolúciójára a Price egyenletből a következő adódik:

$$\frac{d\bar{x}}{dt} = \text{Var}(x)D. \quad (20)$$

Ezt az egyenletet Lande-egyenletnek nevezzük. Lande eredetileg soklókuszos populációgenetikai kontextusban vezette le – de nagyon általánosan érvényes. Jelentése nagyon világos: egy tulajdonság a növekvő fitnessz irányába evolválódik olyan sebességgel, amelyet a fitnessz meredeksége és a populáció  $x$ -tulajdonságra vonatkozó varianciája szab meg.

A Lande-egyenlet természetesen csak egy más felírása annak a körülménynek, hogy a nagyobb fitnesszű változatok gyakoriságának növekedésével az  $x$  változó átlagértéke a növekvő fitnessz irányába mozdul el. Ha új változatok nem keletkeznek (pl. mutációval), akkor előbb utóbb a legnagyobb fitnesszű változat

győzelmével a  $\text{Var}(x)$  variancia nullává válik, és a szelekció leáll. Valós esetben sokszor az a helyzet, hogy a variancia sokáig fennmarad és az  $x$  tulajdonság zavartalanul evolvál a növekvő fitness irányába, amíg egy maximumot el nem ér.

Az ilyesfajta szelekciót irányítottak, direkcionálisnak nevezzük. Noha az evolúció a közhely a közhely szerint egy nagyon lassú folyamat, a direkcionális szelekcióról – amelynek a sebességét épp kiszámoltuk, – ezt nem mondhatjuk el. Ha a variancia és a fitness-meredekség nem túl kicsi, akkor az evolúció sebessége sem az.

### 2.2.7. Fisher-féle alaptörvény

Elsősorban történetileg fontos a Fisher-féle alaptörvény (Fundamental Theorem). Tegyük fel, hogy az  $x_i$  tulajdonság nem más, mint az  $r_i$  fitness érték, amit most minden  $i$ -re rögzítettnek tételezünk fel. Ekkor a Price egyenlet az

$$\frac{d\bar{r}}{dt} = V_r \quad (21)$$

alakot veszi fel, amelyet pedig Fischer „fundamentális törvény”-ének nevezünk. Itt

$$V(r) = \sum_i p_i (r_i - \bar{r})^2 \geq 0 \quad (22)$$

az  $r_i$  értékek varianciája (szórásnégyzete).

Vegyük észre, hogy az átlagos növekedési ráta, vagy átlagfitness soha nem csökken: mindaddig nő, amíg a genetikai variancia jelen van. Ami persze egy természetes következménye annak a folyamatnak, amiben a magasabb fitnessű változatok fokozatosan többségbe kerülnek az alacsonyabb fitnessűekhez képest. A genetikai variancia azonban – mutációk hiányában – végeredményben nullává válik a legnagyobb fitnessű változat végső győzelmével.

### 2.3. És akkor az éjig nő a fitness?

Tételezzük fel, hogy a szelekció folyamatában újabb és újabb mutációk folyamatosan fenntartják a genetikai változatosságot, a genetikai varianciát. Ebben az esetben a szelekció tartósan működik. Naivan szemlélve a Fisher-törvény ekkor azt látszana jelenteni, hogy egy populáció átlagfitnessze állandóan nő a szelekció következtében. Ebből arra következtethetnénk, hogy a legalább 3,5 milliárd éves evolúció nyomán mára hatalmasak lennének a fitness értékek – ami nyilvánvalóan nincs így. Egy magasabbrenű élőlény növekedési rátája természetesen lényegesen alacsonyabb egy koli bacilusénál.

A helyzet megértéséhez érdemes végiggondolni a következőket. Az (5) egyenlet, s így a belőle levezetett (10) replikátor-egyenlet és a (13) Price egyenlet is, érvényes független attól, hogy az  $r_i$  állandó-e, vagy sem: a *pillanatnyi*  $r_i$  értékekre vonatkoznak. E formulák érvényét tehát nem befolyásolja őket sem a környezet esetleges változékonysága, sem pedig a populáció önmagára való visszahatása. A (13) Price-egyenletnél viszont természetesen számít, hogy a változatok  $x_i$  értékei állandóak, hisz ezért igaz, hogy az  $\bar{x}$  csak a  $p_i$  gyakoriságok változása miatt változik a (17) levezetésben. Ez viszont azt jelenti, hogy a (21) törvény (amelyet az  $x_i = r_i$  helyettesítéssel vezetünk le) már feltételezi az  $r_i$  értékek állandóságát.

Ezen a ponton válik szükségessé figyelembe venni a populációreguláció tényét. Valójában a nagyobb fitnessű változat, ha elterjed, jobban leterheli a saját tápanyagforrását, s végső soron neki is 0 lesz a növekedési rátája, fitnessze, ahogy az eredeti típusnak is annyi volt. Végső soron tehát a fitness nem nő az evolúció hosszú története során. A fentiekben azonban a növekedési rátának a tápanyag-leterhelés következtében való megváltozását nem vettük figyelembe, s ez vezetett az abszurdnak látszó eredményre.

Fisher magyarázata szerint törvénye az átlagfitnessnek csakis a *szelekció következtében* beállt változását írja le. Ha mindezt jó értjük, akkor a törvény használható a pillanatnyi szelekció (mikroevolúció) jellemzésére: a szelekció mindig úgy hat, hogy növelje az átlagfitnesszt. Mondhatjuk ezt úgy is, hogy a Fisher-törvény nem regulált, exponenciálisan növekvő populációkra érvényes. Ilyen persze csak a mesében van, viszont a *pillanatnyi* szelekciós viszonyok szempontjából mindegy, hogy a populáció regulált-e? Az viszont bizonyos, hogy a hosszú távú (makro) evolúciót nem jellemezhetjük a fitness folyamatos növekedésével.

## 2.4. Diszkrét generáció

A fenti egyenletek diszkrét generációkra vonatkozó változatai hasonlóképpen felírhatóak. Mindenki levezetheti őket magának, itt csak felsorolom őket. A replikátor-egyenlet analógja diszkrét időre:

$$\Delta p_i = p'_i - p_i = \frac{w_i}{\bar{w}} p_i - p_i = \frac{w_i - \bar{w}}{\bar{w}} p_i \quad (23)$$

ahol most  $w$  jelöli a diszkrét-idejű fitnesszt (amit más kontextusban  $\lambda$ -val jelölünk), a vessző a következő generációra vonatkozó értéket, a  $\Delta$  jelölés pedig a generációváltások között változást jelenti. Hasonló módon a Price egyenlet:

$$\Delta \bar{x} = \sum_i \Delta p_i x_i = \frac{1}{\bar{w}} \sum_i p_i (w_i - \bar{w})(x_i - \bar{x}) = \frac{1}{\bar{w}} \text{Cov}(w, x) \quad (24)$$

A Fisher egyenlet pedig:

$$\Delta \bar{w} = \frac{V_w}{\bar{w}} \quad (25)$$

Végül a szelekciós egyenlet két változatra:

$$\Delta p = \frac{spq}{\bar{w}}, \quad (26)$$

ahol ismét  $s = w_1 - w_2$  és  $q = 1 - p$ .

Vegyük észre, hogy a folytonos és a diszkrét idejű egyenletek csak az átlagfitnessszel való normálásban különböznek. Mivel a diszkrét idejű szelekcióban csak a növekedési ráták aránya számít, természetes, hogy egy ilyen – a lényegét nem érintő – normálásnak fel kell lépnie. (A folytonos idejű esetben a növekedési ráták különbsége számított.)

## 3. Diploid egy lókuszos szelekció

Azt az esetet vizsgáljuk, amikor egy diploid populáció szomatikus (nem ivari) kromoszómájának egyetlen lókuszan két allél (génváltozat) versenyez egymással. Ilyenkor a replikátor, vagy szaporodási egység, szerepét a génkópia játssza.

A génszintű fitnesszt azonban a diploid egyed sikerességéből kell meghatároz-  
nunk. A diszkrét időre vonatkozó jelöléseket használjuk, de pontosan ugyanezt  
mondhatnánk el folytonos időben is.

Feltesszük, hogy az  $\{ij\}$  genotípus (rögzített) fitnessze  $w_{ij}$ , a két nemben  
ugyanaz és  $w_{ij} = w_{ji}$ . Véletlen párosodást tételezünk fel. Ekkor az  $i$ -edik allél  
fitnessze (marginális, gén vagy allél fitnessz):

$$w_i = \sum_j p_j w_{ij} \quad (27)$$

A két allél közötti szelekciós különbség ekkor a következő alakba írható:

$$\begin{aligned} s = w_1 - w_2 &= (pw_{11} + qw_{12}) - (pw_{12} + qw_{22}) = (pw_{11} - qw_{22}) + (q - p)w_{12} \\ &= \frac{p + q}{2}(w_{11} - w_{22}) + \frac{p - q}{2}(w_{11} + w_{22}) + (q - p)w_{12} \\ &= \frac{p + q}{2}(w_{11} - w_{22}) + (q - p) \left( w_{12} - \frac{w_{11} + w_{22}}{2} \right) \\ &= \frac{w_{11} - w_{22}}{2} + 2 \left( \frac{1}{2} - p \right) \left( w_{12} - \frac{w_{11} + w_{22}}{2} \right). \end{aligned} \quad (28)$$

Hiába rögzítettük tehát a heterozigóta fitnesszeket, az allél-fitnesszek általá-  
ban már függenek az allél gyakoriságától. Kivételt képez azt a speciális esetet,  
amikor a heterozigóta fitnessz pontosan megegyezik a homozigóta fitnesszek átlá-  
gával - ekkor ugyanis a gyakoriságfüggést okozó második tag eltűnik. Amennyi-  
ben a heterozigóta fitnessz nagyobb, mint a homozigóta fitnesszek átlaga, a gyá-  
koriságfüggés negatív: minél nagyobb az első génváltozat aránya, annál kisebb a  
fitnessze. A fordított helyzet, amikor a heterozigóta fitnessz a homozigóták átlaga  
alatt marad, természetesen pozitív gyakoriságfüggést okoz.

Létezik-e olyan  $p = p_0 \in (0, 1)$  géngyakoriság érték, amikor az  $s = 0$  egyen-  
súlyi feltétel teljesül, azaz amikor a lokuszra nem hat szelekció? Az egyensúlyi  
pontra a

$$p_0 - \frac{1}{2} = \frac{(w_{11} - w_{22})/2}{2w_{12} - w_{11} - w_{22}} \quad (29)$$

kifejezés adódik, ami a

$$p_0 = \frac{w_{12} - w_{22}}{2w_{12} - w_{11} - w_{22}} \quad (30)$$

vagy az

$$1 - p_0 = \frac{w_{12} - w_{11}}{2w_{12} - w_{11} - w_{22}}. \quad (31)$$

alakba is írható. A  $p_0$  érték felhasználásával (28) a következő alakot ölti:

$$s = 2(p_0 - p) \left( w_{12} - \frac{w_{11} + w_{22}}{2} \right) = (2w_{12} - w_{11} - w_{22})(p_0 - p), \quad (32)$$

Világos, hogy úgy a (30), mint a (31) kifejezésnek pozitívnak kell lennie  
ahhoz, hogy  $p_0$  „belső”, azaz a  $(0, 1)$  intervallumba eső egyensúlyi pont legyen.  
Ez két esetben teljesül: akár akkor, ha a heterozigóta fitnessz nagyobb mindkét  
homozigóta fitnessznél, akár pedig akkor, ha kisebb mindkettőjüknél.

Három esetet különíthetünk el tehát:

1.  $w_{12} \in [w_{11}, w_{22}]$

Ez a legtermészetesebb eset. Ha a heterozigóta fitnessz a homozigóta fitnesszek közé esik, akkor nincs belső egyensúlyi pont. Ilyenkor az  $s$  szelekciós előny nem vált előjelet a  $p \in [0, 1]$  intervallumban. A  $p = 1/2$  érték helyettesítésével láthatjuk, hogy ezt az előjelet (28) első tagja határozza meg. Ekkor az a génváltozat, amelynek homozigóta fitnessze nagyobb a versenytárs homozigóta fitnesszénél, az minden körülmények között előnyben van, és a másikat kiszorítva fixálódik.

2.  $w_{12} > w_{11}, w_{22}$

Amennyiben a heterozigóta fitnessz nagyobb mindkét homozigótáénál, az  $s$  fitness-előny pozitív, amíg  $p < p_0$ , de negatívvá válik  $p > p_0$  esetén. Ilyenkor a szelekció hatására a  $p$  génarány tetszőleges értékről kiindulva bekonvergál a  $p = p_0$  stabil egyensúlyi pontba. A két allél tehát tartósan együttél egy meghatározott arányban.

Egy fontos részlet: a (32) egyenletet a (26) szelekciós dinamikába helyettesítve láthatjuk, hogy a konvergencia monoton, azaz a géngyakoriság sohasem lép át az egyensúlyi pont túloldalára. Ez a probléma a (12) egyenlet által leírt folytonos dinamika esetén nem merül fel.

3.  $w_{12} < w_{11}, w_{22}$

Amennyiben viszont a heterozigóta fitnessz a legalacsonyabb, az előző gondolatmenet megfordításával a belső fixpont instabil. A kezdőállapottól függően vagy az első (ha  $p < p_0$ ), vagy a második (ha  $p > p_0$ ) génváltozat kihal.

Az első eset a triviális eset: a két génverzió egyike egyértelműen jobb a másikinál, és győz. A második eset speciális okokból fordulhat elő, a sarlóssejtes vérszegénység a híres példa. A harmadik esetet

És mi a helyzet az átlagfitnessszel a diploid esetben? Könnyű belátni, hogy a  $w_{ij}$  diploid fitnesszek populációs átlaga ugyanaz, mint a (27) allél-fitnesszeknek az allépopulációra vett átlaga:

$$\bar{w} = pw_1 + (1-p)w_2 = \tag{33}$$

$$= p^2w_{11} + 2p(1-p)w_{12} + (1-p)^2w_{22} = \tag{34}$$

$$= p^2(w_{11} + w_{22} - 2w_{12}) + 2p(w_{12} - w_{22}) + w_{22} = \tag{35}$$

$$= (p - p_0)^2(w_{11} + w_{22} - 2w_{12}) + \dots \tag{36}$$

Érdemes észrevenni, hogy

$$s = \frac{1}{2} \frac{d\bar{w}}{dp} \tag{37}$$

Azaz: a szelekció az átlagfitnessz növelésének irányába hat, ahogy az a korábbiak szerint el is várható.

## Függelék: egyenesillesztés és kovariancia

Arra vagyunk kíváncsiak hogy az  $x$  és az  $y$  mennyiség (mondjuk egy ország GDP-je és az ott élők – szociológusok által valahogy számszerűsített – jóléte) között találhatunk-e olyan összefüggést, hogy az  $x$  változó növekedése tendenciaszerűen

együttjár az  $y$  növekedésével/csökkenésével? Ehhez legegyszerűbben egyenest illesztünk az  $(x_i, y_i)$  értékpárokra. Számoljuk ki a legjobban illeszkedő egyenes  $b$  meredekségét!

Kezdetnek az egyszerűség kedvéért tegyük fel, hogy mindkét változó átlagértéke zérus, azaz  $\bar{x} = \bar{y} = 0$ . Ekkor az egyenes nyilván az origón megy keresztül, egyenletét az  $y = bx$  alakban keressük. Definiáljuk az egyenesillesztés hibáját az  $y$  értékek négyzetes eltéréseinek átlagával:

$$F = \overline{(y - bx)^2}. \quad (38)$$

Ezt az eltérést akarjuk minimalizálni  $b$  megválasztásával. A derivált:

$$\frac{dF}{db} = -2\overline{x(y - bx)} = 0, \quad (39)$$

ahonnan  $b$  optimális értéke:

$$b = \frac{\overline{xy}}{\overline{x^2}}. \quad (40)$$

Amennyiben nem tesszük fel, hogy az átlagértékek nullák, úgy a fenti megfontolás az átlagértéktől való eltérésekre érvényes. Ekkor

$$b = \frac{\overline{(y - \bar{y})(x - \bar{x})}}{\overline{(x - \bar{x})^2}} = \frac{\text{Cov}(y, x)}{\text{Var}(x)} \quad (41)$$

adódik az egyenes meredekségére. (Házi feladat: határozzuk meg a tengelymetszetet!) Itt  $\text{Var}(x)$ -et az  $x$  változó varianciájának (vagy szórásnégyzetének),  $\text{Cov}(y, x)$ -t pedig az  $x$  és az  $y$  változó kovarianciájának nevezzük.

Ami itt érdekel bennünket, az a kovariancia. A kovariancia (s ezzel együtt a legjobban illeszkedő egyenes  $b$  meredeksége) akkor pozitív, ha az  $x$  és az  $y$  változó jellemzően ugyanabba az irányba tér el a saját átlagától – vagyis ha nagyobb  $x$ -hez jellemzően nagyobb  $y$ , kisebb  $x$ -hez pedig jellemzően kisebb  $y$  tartozik. (Rajzold le!) Megfordítva, a kovariancia és meredekség negativitása azt jelenti, hogy nagyobb  $x$  jellemzően kisebb  $y$ -hoz vezet. Ha a kovariancia a nulla környékén van, akkor nem találtunk tendenciaszerű összefüggést a két változó között. (Statisztikában megtanítják, hogy milyen erős kovarianciát kell ahhoz találnunk, hogy ne tartsuk a kapott összefüggést a véletlen játéknak.)

Könnyen belátható, hogy ha a két változó (valószínűségelméleti értelemben) független egymástól, akkor  $\text{Cov}(y, x) = 0$ . (Gondold végig!) Az állítás megfordítása azonban nem igaz: zérus kovarianciából a függetlenség nem következik. Legyen például  $x$  és  $y$  egy kör pontjainak koordinátái. Ekkor természetesen nem függetlenek egymástól, hiszen kielégítik a kör

$$x^2 + y^2 = 1$$

egyenletét. A szimmetria okán viszont világos, hogy a kovariancia csak nulla lehet. A két változó között tehát – noha nem függetlenek egymástól – nincs meg az a fajta tendenciaszerű összefüggés, amit a kovariancia mér.

Végül egy házi feladat. Tessék belátni a következő nagyon hasznos összefüggéseket:

$$\text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) + 2\text{Cov}(x, y) \quad (42)$$

$$\overline{x \cdot y} = \bar{x} \cdot \bar{y} + \text{Cov}(x, y) \quad (43)$$