

Baktériumok növekedésének  
modellezése és paramétereinek  
becslése véletlen lappangási idők  
mellett

Doktori értekezés

írta

Baranyi József

Informatikai Doktori Iskola  
Szegedi Tudományegyetem  
2010

### Legyakrabban használt jelölések

- $\alpha(t)$  Baktériumkultúra fiziológiai állapotára jellemző mennyiség a  $t$  időpontban ( $0 \leq \alpha(t) \leq 1$ ).  
Legegyszerűbb formában  $\alpha(t) = \mu(t) / \mu_{max}$
- $\alpha_0$  Az  $\alpha(t)$  értéke a 0 időpontban (amikor a baktériumok egy új környezetbe kerülnek).  $\alpha_0 = e^{-\mu \cdot \lambda}$
- $\alpha_N$  Az inokuláció pillanatában  $N$  sejtből álló populáció (kezdeti) fiziológiai állapota.
- $h_0 = \mu \cdot \lambda = -\ln \alpha_0$
- $\lambda$  Általános jelölés a (különbéle, empirikus vagy mechanisztikus módon definiált) lappangási időre.
- $L, L_g$  Az exponenciális fázisban, a sejtpopuláció úgy nő a log-skálán mint az  $L$ -paraméterrel késleltetett lineáris  $y(t) = \mu \cdot (t - L)$  függvény. A  $-L_g$  index a definíció "geometriai" voltára utal (szemben a lappangás fiziológiai definíciójával).
- $L^{(i)}$  Az  $i$ -dik sejt  $L$ -paramétere. Bár szigorú értelemben ez nem ugyanaz, mint  $\tau_i$ , a fiziológiai lappangási idő, a különbséget elhanyagoljuk, hacsak éppen nem az a vizsgálat tárgya. Numerikus szimulációval kimutatható hogy  $E(L^{(i)}) = E(\tau_i)$  és a varianciájuk is hasonló, feltéve hogy a generációs idők a lappangás után nem nagyon szórnak.
- $L_N$  "Geometriai" lappangási idő, hangsúlyozva hogy  $N$  számú kezdeti sejt generálta.
- $\mu(t)$  Pillanatnyi specifikus növekedési sebesség a  $t$  időpontban:  
 $\mu(t) = dx/dt / x(t) = d(\ln x(t)) / dt$
- $\mu_{max}, \mu$  (Maximális) specifikus növekedési sebesség

$m$	Szigmoid függvény görbületi paramétere, amely az exponenciálisból a stacionális fázisba való átmenetet jellemzi.
$v_b$	Specifikus felhalmozódási sebessége egy olyan intracelluláris anyagnak ami a lappangás leküzdéséhez elengedhetetlen.
$n_C$	Szigmoid függvény görbületi paramétere, amely a lappangásból az exponenciális fázisba való átmenetet jellemzi. Nálunk: $n_C = v_b / \mu_{max}$
$N, N_0$	“Batch” kultúra inokulációs szintje; $x_0$ -lal is jelöljük.
$\tau_i$	Véletlen változó, az $i$ -dik sejt “fiziológiai” lappangási ideje. Az első osztódásig terjedő idő $\tau_i$ és az első (szintén véletlen) generációs idő összege.
$t$	Idő
$T_{det}$	Detektációs idő. $x(T_{det}) = X_{det}$
$x(t)$	Sejtkoncentráció a $t$ időpontban
$x_0$	Kezdeti sejtkoncentráció, $x(0)$
$x_{max}$	Maximális populáció, a környezet eltartó képessége.
$X_{det}$	Detektációs szint (konstans): $x(T_{det}) = X_{det}$
$y(t)$	A sejtkoncentráció természetes alapú logaritmusa a $t$ időpontban: $y(t) = \ln x(t)$
$y_0$	A kezdeti sejtkoncentráció természetes alapú logaritmusa: $y_0 = y(0)$
$y_{max}$	A maximális sejtkoncentráció természetes alapú logaritmusa: $y_{max} = \ln x_{max}$
$y_{det}$	A detektációs szint természetes alapú logaritmusa: $y_{det} = \ln X_{det}$ .

## Bevezetés

Miközben a 80-as években egyre hatékonyabb számítógépek váltak mindennapos eszközzé, egy új fogalom terjedt el az élelmiszer-mikrobiológiában “prediktív mikrobiológia” névvel. Az új diszciplína legközelebbi társterülete a biotechnológiában található, amelynek kiterjedt prediktív matematikai modelljei a 60-as évek terméke.

A disszertáció elején megállapítottuk hogy az élelmiszer-mikrobiológia területére szükséges matematikai modellekkel szembeni elvárások sok tekintetben különböznek attól amit biotechnológiai problémák kezelése kíván. Az élelmiszerbeli baktériumok növekedésének és túlélésének modellezésére új megközelítési módokra van szükség.

### Baktériumpopulációk determinisztikus modellezése

Jelölje  $x(t)$  a baktérium koncentrációjának változását a  $t$  idő függvényében. Növekedésüket, állandó környezetben, az ismert konstans együtthatós lineáris differenciálegyenlettel írják le:

$$\frac{dx}{dt} = \mu x$$

A  $\mu$  paraméter a populáció specifikus növekedési sebességét jelöli. Értéke jellemző a fajra és a környezetre. Felfogható úgy mint egy sejt osztódásainak átlagos száma egységnyi idő alatt. A paraméter környezeti tényezők, pl. hőmérséklet, pH etc, függvénye. Értéke megbecsülhető az “ $\ln x(t)$  v. idő” görbe lineáris fázisához illesztett egyenes meredekségével.

Élelmiszer-mikrobiológiai laboratóriumokban tipikusan “batch” kísérleteket végeznek amelyek a logaritmus skálán szigmoid görbét alkotnak. Az exponenciális (logaritmus skálán: lineáris) fázis előtti időszakot lappangási vagy

adaptációs / igazodási fázisnak hívjuk. Míg a biotechnológiai kutatások mellőzik ezt a periódust, az élelmiszer-mikrobiológiában azon belül is a kockázatelemzésben kifejezetten nagy jelentősége van.

Aritmetikai skálán szigmoid függvényeket generál a következő Kezdeti Érték Probléma:

$$\frac{dx}{dt} = \mu(x) x$$

$$x(0) = x_0 \quad (0 < x_0 < x_{max} ; \quad x_{max} \text{ rögzített})$$

ahol

- $\mu: (0, x_{max}] \rightarrow \mathbb{R}$
- $\mu(x)$  folytonosan differenciálható  $(0, x_{max})$ -on
- $d\mu/dx$  szigorúan negatív  $(0, x_{max})$ -on
- $\mu(x_0) > 0$  és  $\mu(x_{max}) = 0$

A fenti általános modellből származtatható szigmoid függvényeket rutinszerűen alkalmazták a 90-es évekig a baktérium populációk log-skálán való kinetikai leírására. Ez a tisztán empirikus megoldás tarthatatlanná vált ahogyan a matematikai modellek kifinomultsága iránti igény növekedett élelmiszer-mikrobiológiában is.

### Sztocasztikus modellezés egyéni sejtekre

Legyen  $N_0=N$  egy növekvő sejtpopuláció kezdeti nagysága. Mellőzzük a későbbi stacionárius fázist. Defináljuk a lappangási fázis végét az  $y(t)=\ln x(t)$  növekedési görbéhez illesztett kétfázisú lineáris függvény (ahol kezdeti fázisban nincs növekedés) töréspontjával ( $L_N$ ). Ezt az ún. *populációs lappangási időt*, az  $N$  számú kezdeti sejtszám generálta.  $L_N$  kapcsolata az egyedi sejtek lappangási idejének eloszlásával matematikailag sem egyszerű. Ráadásul igen nehéz nagy tömegben megfigyelni az egyedi sejteknek lappangás alatti viselkedését mert ebben a fázisban a sejtkoncentráció kicsi,

vizsgálata automatizált rendszerrel (ahol megfelelő nagyságú mintát lehet gyűjteni) meglehetősen körülményes.

### **Célkitűzések**

Baktériumok kinetikájára szolgáló új modellek bevezetését tűztük ki célul, a populációs szint determinisztikus modelljétől és numerikus alkalmazásuktól az egyéni sejtek kinetikájának sztochasztikus modelljéig, beleértve a paraméter becsléseket célzó numerikus algoritmusokat. A cél az volt hogy segítsük a baktériumoknak az élelmiszer-környezetre adott válaszainak jobb megértését és előrejelzését, a mikrobiológiai kockázatelemzés javítását.

A sejtpopulációs szinten bevezetett determinisztikus problémák kapcsán meg kívántuk mutatni, hogy az élelmiszer-mikrobiológia mely területein a legfontosabb empirikus modelleket mechanisztikus megfontolásokkal felváltani. Célunk könnyen implementálható modellek bevezetése valamint numerikus tulajdonságaiknak elemzése volt.

Az egyéni sejtek szintjére alkalmazott sztochasztikus modellekkel az volt a célunk hogy formulákat vezessünk le melyek kapcsolatot teremtenek egyéni sejtek lappangási idejének eloszlása és a populáció lappangási ideje között. Az elmélet arra volt hivatott hogy numerikus algoritmusokat találjunk amelyek segítségével populációs szintű automatizált mérésekből következtetéseket vonjunk le az egyéni sejtek variabilitására.

## Eredmények

### Determinisztikus, nem-autonóm modellek populáció növekedésére és implementációjuk

1. Alkalmaztunk egy adaptációs függvényt a Bevezetésben említett determinisztikus modell  $\mu(x)$  függvényének módosítására

$$\frac{d}{dt}x = \alpha(t)\mu(x) \cdot x \quad (0 \leq t < \infty; \quad 0 < x)$$
$$x(0) = x_0 \quad (0 < x_0 < x_{\max})$$

Az  $\alpha(t)$  adaptációs függvényről feltételeztük, hogy elegendően sima; esetleg függ az inokuláció előtti környezettől. Emellett  $0 \leq \alpha(t) \leq 1 \quad (0 \leq t < \infty)$

$\alpha(t) \rightarrow 1$  monoton növekedve amint  $t \rightarrow \infty$

A kapott nem-autonóm modell  $g_\alpha(t)$  megoldása kifejezhető az autonóm modell  $f$  megoldásával és az  $\alpha(t)$  adaptációs függvénnyel:

$$g_\alpha(t) = f(A(t)) \quad \text{ahol} \quad A(t) = \int_0^t \alpha(s) ds$$

Ezt Baranyi et al (1993b) publikálta és így megvetette az alapjait később konkretizált eseteknek, amelyek felhasználóbarát applikációkra koncentráltak.

2. A Hill-függvény egy lehetőség volt  $\alpha(t)$  választására – ld. Baranyi et al (1993a). A kapott formulák algebrai kifejezések voltak, de túl komplikáltak ahhoz hogy közönséges táblázatkezelő programokba, mindennapi használatra kerüljenek. Emellett  $\alpha(t)$  –nek nem volt mechanisztikus alapja, s nem volt könnyű értelmezni arra az esetre amikor a környezet dinamikusan változna a lappangási idő alatt.

**3.** Baranyi - Roberts (1994) egy Michaelis-Menten kinetikán alapuló adaptációs függvényt vezetett be, amely a mai napig is a legsikeresebb gyakorlati célokra. Ezzel a modell így írható fel:

$$\frac{d}{dt}x = \frac{q(t)}{1 + q(t)} \mu(x) \cdot x$$

$$\frac{d}{dt}q = v_b \cdot q$$

$$x(0)=x_0 \quad ; \quad q(0)=q_0$$

ahol  $v_b$  az adaptáció sebessége. Kiderült, hogy

$$\alpha(t) = \frac{q_0}{q_0 + e^{-v_b \cdot t}}$$

és

$$A(t) = t + \frac{1}{v} \ln \left( \frac{e^{-v_b t} + q_0}{1 + q_0} \right)$$

Valamint

$$\lambda = \frac{\ln(1 + 1/q_0)}{v_b}$$

a lappangási időnek egy jó tulajdonságokkal rendelkező definíciója:  $f(A(t))$  szigmoid görbe a log-skálán, és  $f(t-\lambda)$ -hoz konvergál. A  $0 \leq \alpha_0 \leq 1$  paraméterre javasoltuk hogy használjuk a kezdeti fiziológiai állapot jellemzésére.

**4.** Ha Richards modelljével azonosítjuk az autonóm differenciálegyenletet:

$$\frac{dx(t)}{dt} = \mu_{\max} x(t) \left( 1 - \left( \frac{x(t)}{x_{\max}} \right)^m \right)$$



akkor algebrai megoldást kapunk amelynek a természetes logaritmus felírható mint

$$y(t) = y_0 + \mu_{\max} t - \frac{1}{m} \ln \left( I + \frac{e^{m\mu_{\max}t} - 1}{e^{m(y_{\max} - y_0)}} \right)$$

A fent említett  $A(t)$  függvényt írva  $t$  helyett, ahogy az 1-es és 3-as pontban javasoltuk, egy jól használható 6-paraméteres szigmoid függvényt kapunk, amelynek két paramétere,  $m$  és  $n_C = v_b / \mu_{\max}$  a fázisok közötti két átmeneti állapot görbületét jellemzi.

A modellnek különféle átparaméterezését javasoltuk, hogy a numerikus stabilitás az illesztések során jól kezelhető legyen (Baranyi et al 1994, 1995a).

**5.** Feltételeztük, hogy az adaptációs sebesség,  $v_b$ , és a potenciális maximális növekedési sebesség késleltetés nélkül felveszi azt az aktuális  $\mathbf{E}$  környezetre jellemző értéket:

$$\frac{dx(t)}{dt} = \frac{q(t)}{1 + q(t)} \mu(\mathbf{E}(t)) x \cdot u(x)$$

$$\frac{dq(t)}{dt} = v_b(\mathbf{E}(t))$$

Ez a dinamikus modell akkor is alkalmazható baktériumnövekedés előrejelzésére ha a környezet a lappangási idő alatt változik. Fontos egyszerűsítés volt hogy feltételeztük:  $v_b = \mu_{\max}$  and  $q_0 = \text{const}$  különféle környezetekre. Ezek az egyszerűsítések nagy adatbázisokon végzett regressziós tapasztalatokon alapulnak. A feltételek biztosan elfogadhatóak, ha a dinamikus környezetet a hőmérséklet ingadozásával azonosítjuk, egy stabilan növekedést-garantáló tartományban, nem túl hirtelen változások mellett (Baranyi et al, 1995).

Sztoczasztikus modell az egyéni sejtek lappangási idejének leírására

6. Csak a lappangási és az exponenciális fázisra szorítkozva, definiáljuk az  $N$  kezdeti sejt szám által generált lappangási időt ( $L_N$ ) az illesztett kétfázisú lineáris függvényvel:

$$y(t) = y_0 + \max(\mu(t-L_N), 0)$$
$$y_0 = \ln N$$

A definíció miatt a kapott  $L$ -értéket "geometriai" lappangási időnek hívjuk. Ha  $N=1$ , az egyéni lappangási idő,  $L_1$  definíciójához jutunk. Ez nem ugyanaz mint a fiziológiai lappangási idő fogalma, de gyakorlati alkalmazásokra az  $L$ -definíció hasznosabb. Történelmi okok miatt a két lappangási idő közötti különbséget elhanyagoltuk.

Tegyük fel, hogy  $\tau_i$  ( $i = 1 \dots N$ ) azonos eloszlású valószínűségi változók  $\tau$  várható értékkel. Vezessük be az  $\alpha_i = e^{-\mu\tau_i}$  változókat. Ezek aritmetikai átlaga az egész populáció megfelelő  $\alpha$  értékéhez közelít, ahogy  $N$  növekszik:

$$\alpha(N) = e^{-\mu L_N} = \frac{\sum_{i=1}^N e^{-\mu\tau_i}}{N}$$

Kiderült, hogy a populáció fiziológiai állapotát  $N$  nagy értékére vehetjük úgy mint az egyéni sejtek fiziológiai állapotának átlaga. Vagyis:

$$\alpha(N) \xrightarrow{N \rightarrow \infty} \alpha \quad \text{és} \quad L_N \xrightarrow{N \rightarrow \infty} \frac{-\ln(\alpha)}{\mu} = \lambda$$

Továbbá, a konvergencia sebessége ezzel becsülhető:

$$\text{Var}[\alpha(N)] = \frac{\text{Var}(\alpha_i)}{N}$$

A fenti eredmények (ld. Baranyi, 1998) megoldották a kitűzött problémát és alapot szolgáltattak további fejlesztésekhez.

7. Összehasonlítottunk egy determinisztikus compartment-modellt a mi sztochasztikus modellünkkel és bebizonyítottuk hogy az eredmények megegyeznek ha a baktériumok szaporodását Poisson-féle születési folyamatként kezeljük. PI:

$$L_N \xrightarrow{N \rightarrow \infty} \frac{\ln(1 + \mu \cdot \tau)}{\mu} = \lambda$$

megkapható a compartment-modellből is (Baranyi, 1998).

8. Párhuzamot vontunk a “vállasodó” túlélési görbék és a lappangási idővel kiegészített növekedési görbék között, mind populációs, mind pedig egyéni sejt szinten. Három különböző esetet vizsgáltunk az egyéni túlélési/lappangási idő eloszlására és formulákat származtattunk analitikai tulajdonságaikra.

A gamma eloszlás bizonyult a leghasználhatóbbnak. Növekedés esetén az előző pontban ismertetett formulák általánosítását kaptuk:

$$\alpha = E[\alpha_i] = \left(\frac{r}{r+1}\right)^p ; \quad \text{Var}[\alpha_i] = \left(\frac{r}{r+2}\right)^p - \left(\frac{r}{r+1}\right)^{2p}$$

ahol  $r = \kappa/\mu$  és  $\kappa$  az eloszlás skála-,  $p$  pedig az alakparamétere. Ezek az eredmények (Baranyi - Pin, 2001) alapját szolgálják azoknak a numerikus módszereknek amelyek az utolsó felyezetben ismertetünk az egyéni sejtek lappangási ideje eloszlásának mérésére.

Bebizonyítottunk egy, a populációs lappangási idő és az egyéni sejtek lappangási idejének átlaga közötti általánosítást:

$$\lambda = \frac{\ln\left(1 + \frac{\mu}{p} \tau\right)}{\mu / p}$$

ahol  $\tau=p/\kappa$  a gamma eloszlás várható értéke. Ezzel megmutattuk, hogy a populációs lappangási idő hogyan függ az egyéni sejtek lappangási ideje eloszlásának alakjától.

**9.** A fenti formulákat felhasználva, módszert fejlesztettünk ki a növekedési sebesség és az egyéni lappangási idők várható értékének és varianciájának becslésére. Különböző inokulációs szintről indulva, ahol a szintek közötti arány ismert volt, pl mert hígítással kaptuk őket), megköveteltük, hogy egy  $x_0$  szintről induló kultúra az exponenciális fázisban éri el a detektációs szintet. Ezután egy ANOVA eljárást vezettünk be:

Jelölje  $T_{det}$  a detektációs időt amikor  $x(T_{det}) = X_{det}$ . Használjuk a következő jelöléseket: sejt kultúrák  $m$  csoportját növesztjük  $m$  különböző inokulum szintről indulva. A  $k$  jelű ( $k=1 \dots m$ ) csoportban  $j=1 \dots n_k$  kultúra található. Itt, egy rögzített  $j$ -re, az  $X_{det}$  detektációs szintet a sejtek  $T^{(j)}$  idő alatt éri el, az  $x_0^{(j)}$  szintről indulva. Jelölje  $\alpha^{(j)} = \alpha(x_0^{(j)})$  e kultúra fiziológiai állapotát. Ekkor

$$\alpha^{(j)} = e^{-\mu T^{(j)}} / r^{(j)}$$

ahol  $r^{(j)} = x_0^{(j)} / X_{det}$ . Valamint  $Var(\alpha^{(j)}) = v / X_{det} r^{(j)}$

ahol  $v$  jelöli az  $\alpha_i = e^{-\mu T_i}$  egyéni fiziológiai állapotok közös varianciáját. Az összes  $\alpha_i$  közepe:

$$\bar{\alpha} = \frac{\sum_{j=1}^n r^{(j)} \alpha^{(j)}}{\sum_{j=1}^n r^{(j)}} = \frac{\sum_{j=1}^n e^{-\mu T^{(j)}}}{r_{sum}} \quad \text{ahol} \quad r_{sum} = \sum_{j=1}^n r^{(j)},$$

míg a csoportközepek: 
$$\bar{\alpha}^{(k)} = \frac{\sum_{j=1}^{n_k} \alpha_{k,j}}{n_k} = \frac{\sum_{j=1}^{n_k} e^{-\mu T_{k,j}}}{n_k r^{(k)}}$$

Vezessük be a variancia hányadost:

$$V = \frac{\sum_{k=1}^m r^{(k)} \left( \frac{-r^{(k)}}{\alpha} - \frac{-}{\alpha} \right)^2}{n_k \frac{1}{r_{sum}} + \frac{2}{n}}$$

$$V = \frac{\sum_{k=1}^m \frac{r^{(k)}}{1 - 1/n_k} \sum_{j=1}^{n_k} \left( \alpha_{k,j} - \frac{-r^{(k)}}{\alpha} \right)^2}{n_k}$$

V eloszlása közel F-eloszlású és minimalizálásával jó becslés kapható a  $\mu$ -re és az  $\alpha_i$  értékek eloszlására; ezzel az egyéni lappangási idők hasonló paramétereire.

A módszert egy könnyen használható táblázatkezelő programba építve (Baranyi and Pin, 1999) számos kísérleti adathalmazon teszteltük.

**10.** Formulákat származtattunk arra vonatkozólag is, hogyan becsüljük az egyéni lappangási idők eloszlásának alakját. A számítások azon alapultak, hogy a kezdeti sejtszámok Poisson eloszlást követnek (ami hígítással elérhető, ld Baranyi et al, 2009) és hogy véletlen tagszámú véletlen változók összegének momentumai explicit módon kiszámíthatóak, ha a tagszám szintén Poisson-eloszlású. Feltételeztük, hogy az egyéni lappangási idők “eltolt” gamma eloszlásúak tehát várható értékük  $L_g = T_{\text{shift}} + \beta\theta$  ahol  $\beta$  az eloszlás alakját,  $\theta$  a skáláját jellemzi. Az eméleti és a tapasztalati momentumokat egyenlővé téve, olyan módszert kaptunk ami az előző pontban leírtak kiterjesztése. Az eljárást maximum-likelihood elven alapuló módszerrel hasonlítottuk össze. Az eredmények között nem volt lényeges különbség, de a momentumokon alapuló módszer sokkal gyorsabb és számítógép-erőforrás igénye általában sokkal kisebb. Emiatt a módszert továbbra is egyszerű táblázatkezelő programba építve alkalmaztuk, ami nagy előny, lévén hogy a legtöbb élelmiszer-mikrobiológiai laboratórium ilyen formában gyűjti és tárolja mérési adatait.

## **Deklaráció**

A fentiek saját munkám eredményei. Cikkeim társszerzői nem matematikusok hanem mikrobiológusok. Ez alól kivétel Carmen Pin, akinek eredménye néhány formula a Baranyi and Pin (2001) cikkben; és Kutalik Zoltán, aki a 10-es pontban ismertetett maximum-likelihood programot írta és az eredményt összehasonlította a momentumok módszerével kapott eredményekkel (Baranyi et al, 2009).