Bistability in Boolean network models

Matthew Macauley

Department of Mathematical Sciences Clemson University http://www.math.clemson.edu/~macaule/

Math 4500, Spring 2016

Motivation

Weaknesses of previous Boolean models

- All processes take a single timestep.
- Only assumed high and low levels in intracellular lactose. Medium levels are needed for bistability to exist.
- No time-delays incorporated.

In this section, we'll see how to add these types of features to Boolean models.

In particular, we'll see how to incorporate features such as:

- loss of concentration due to dilution and degradation;
- time-delays due to cellular processes

Dilution and degradation

Suppose Y regulates the production of X.

Assume Y(t) = 1 implies X(t + 1) = 1. (activation takes 1 step).

Generally, the loss of X due to dilution and degradation takes several steps.

Introduce new variables $X_{\text{old}(1)}, X_{\text{old}(2)}, \dots, X_{\text{old}(n)}$

Properties

- (i) If Y(t)=0 and X(t)=1, then $X_{{\sf old}(1)}(t+1)=1$. ("X has been reduced once by dilution & degradation.")
- (ii) If Y(t) = 0 and $X_{\text{old}(i-1)}(t) = 1$, then $X_{\text{old}(i)}(t+1) = 1$. ("X has been reduced i times by dilution & degradation.")
- (iii) The number of "old" variables is determined by the number of timesteps required to reduced [X] below the discretation threshold.

Thus, X(t+1) = 1 when either of the following holds:

- Y(t) = 1 (new amount will be produced by t + 1),
- $X(t) \wedge \overline{X_{\text{old}(n)}(t)} = 1$ (previous amounts of X still available).

$$X(t+1) = Y(t) \vee \left(X(t) \wedge \overline{X_{\mathsf{old}(n)}(t)}\right)$$

Other features

Medium levels of lactose

Introduce a new variable L_{high} so that $L_{high} = 1$ implies L = 1.

- High lactose: L = 1, $L_{high} = 1$.
- Medium lactose: L = 1, $L_{high} = 0$.
- Low lactose: L = 0, $L_{high} = 0$.

We can ignore any state for which L=0, $L_{high}=1$.

Previously, we introduced a variable L_ℓ to denote "at least low levels of lactose," so $(L,L_\ell)=(1,0)$ described "medium lactose." This is an equally valid way to acheive the same goal.

Time delays

Say R regulates production of X, delayed by time τ (n steps).

Introduce new variables R_1, R_2, \ldots, R_n , with transition functions:

$$R_1(t+1) = R(t)$$

 $R_2(t+1) = R_1(t)$
 \vdots
 $R_n(t+1) = R_{n-1}(t)$
 $X(t+1) = R_n(t)$

Estimating constants for our Boolean model

3-variable ODE model of the lac operon (Yildirim and Mackey, 2004)

Let M(t)= mRNA, $B(t)=\beta$ -galactosidase, and A(t)= allolactose (concentrations), respectively.

$$\begin{split} \frac{dM}{dt} &= \alpha_M \frac{1 + K_1 (e^{-\mu \tau_M} A_{\tau_M})^n}{K + K_1 (e^{-\mu \tau_M} A_{\tau_M})^n} - \widetilde{\gamma_M} M \\ \frac{dB}{dt} &= \alpha_B e^{-\mu \tau_B} M_{\tau_B} - \widetilde{\gamma_B} B \\ \frac{dA}{dt} &= \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{A}{K_A + A} - \widetilde{\gamma_A} A \end{split}$$

We need to estimate these rate constants and time delays from the literature.

- Time delays: $\tau_M = .10 \text{ min}$, $\tau_B = 2.00 \text{ min}$.
- Degradtion rates are harder to determine experimentally, and they vary widely in the literaure. Sample values:

$$\left\{ \begin{array}{l} \gamma_A = .52 \; \mathrm{min}^{-1}, \quad .0135 \; \mathrm{min}^{-1}, \quad .00018 \; \mathrm{min}^{-1} \\ \gamma_B = .00083 \; \mathrm{min}^{-1}, \\ \gamma_M = .411 \; \mathrm{min}^{-1}, \\ \mu \in (.0045, \; .0347) \end{array} \right.$$

Estimating constants for our Boolean model

Approach

We'll select "middle of range" estimates for the rate constants:

$$\mu = .03 \text{ min}^{-1}$$
,

$$Arr \gamma_B = .001 \ {
m min}^{-1} \qquad \Longrightarrow \qquad \widetilde{\gamma_B} = \gamma_B + \mu = .031,$$

$$\mathbf{P} \gamma_{M} = .411 \text{ min}^{-1} \Longrightarrow \widetilde{\gamma_{M}} = \gamma_{M} + \mu = .441.$$

Degradation is assumed to be exponential decay: x' = -kx implies $x(t) = Ce^{-kt}$.

The half-life is the time t such that:

$$x(t) = Ce^{-kt} = .5C \implies e^{-kt} = .5 \implies -kt = \ln \frac{1}{2} \implies t = \frac{\ln 2}{k}$$

Half-lives

$$\widetilde{h}_A = \frac{\ln 2}{\widetilde{\gamma}_A} = 15.753$$
 (approx. 1 time-step to decay)

$$\widetilde{h}_B = \frac{\ln 2}{\widetilde{\gamma}_B} = 22.360$$
 (approx. 2 time-steps to decay)

$$\widetilde{h}_M = \frac{\ln 2}{\widetilde{\gamma}_M} = 1.5$$
 (approx. 0 time-steps to decay)

Model assumptions

- Variables are M, B, A.
- Glucose absent. Intracellular lactose present, two parameters: L and L_{high} .
- Time-step ≈ 10 min.
- Ignore (all \ll 10): $\tau_M = .10$ min, $\tau_B = 2$ min, $\widetilde{h_M} = 1.572$ min.
- Introduce variables for dilution and degradation:
 - A_{old} (since $\widetilde{h_A} \approx 15.8 \approx 1 \text{ timestep}$)
 - B_{old} , $B_{\text{old}(2)}$ (since $\widetilde{h_B} \approx 22.4 \approx 2$ timesteps)

Proposed model

$$\begin{split} f_{M} &= A & f_{B} &= M \vee \left(B \wedge \overline{B_{\text{old}(2)}} \right) \\ f_{A} &= \left(B \wedge L \right) \vee L_{\text{high}} \vee \left(A \wedge \overline{A_{\text{old}}} \wedge \overline{B} \right) & f_{B_{\text{old}(1)}} &= \overline{M} \wedge B \\ f_{A_{\text{old}}} &= \left((\overline{B} \vee \overline{L}) \wedge \overline{L_{\text{high}}} \right) \wedge A & f_{B_{\text{old}(2)}} &= \overline{M} \wedge B_{\text{old}(1)} \end{split}$$

Most of the functions should be self-explanatory.

Justification for f_A

$$f_A = (B \wedge L) \vee L_{high} \vee (A \wedge \overline{A_{old}} \wedge \overline{B})$$

There are 3 ways for allolactose to be available at t + 1:

- (i) β -galactosidase and lactose are present;
- (ii) high levels of lactose (assume basal concentrations of β -galactosidase);
- (iii) Enough allolactose is present so that it's not degraded below the threshold, and no β -galactosidase is present.

Let's write our model into polynomials form, with parameters (L, L_h) and variables $(x_1, x_2, x_3, x_4, x_5, x_6) = (M, A, A_{\text{old}}, B, B_{\text{old}(1)}, B_{\text{old}(2)})$:

$$\begin{array}{ll} f_{M} = A & f_{1} = x_{2} \\ f_{A} = (B \wedge L) \vee L_{high} \vee \left(A \wedge \overline{A_{old}} \wedge \overline{B}\right) & f_{2} = x_{2}(1 + x_{3})(1 + x_{4}) + (Lx_{4} + L_{h} + x_{4}LL_{h}) \\ & + x_{2}(1 + x_{3})(1 + x_{4})(Lx_{4} + L_{h} + x_{4}LL_{h}) \\ & + x_{2}(1 + x_{3})(1 + x_{4})(Lx_{4} + L_{h} + x_{4}LL_{h}) \\ f_{A_{old}} = \left((\overline{B} \vee \overline{L}) \wedge \overline{L_{high}}\right) \wedge A & f_{3} = (1 + x_{4}L)(1 + L_{h})x_{2} \\ f_{B} = M \vee \left(B \wedge \overline{B_{old(2)}}\right) & f_{4} = x_{1} + x_{4}(1 + x_{6}) + x_{1}x_{4}(1 + x_{6}) \\ f_{B_{old(1)}} = \overline{M} \wedge B & f_{5} = (1 + x_{1})x_{4} \\ f_{B_{old(2)}} = \overline{M} \wedge B_{old(1)} & f_{6} = (1 + x_{1})x_{5} \end{array}$$

Using Sage to compute the fixed points (high lactose)

```
1
2     P.<x1,x2,x3,x4,x5,x6> = PolynomialRing(GF(2), 6, order ='lex'); P
3     Multivariate Polynomial Ring in x1, x2, x3, x4, x5, x6 over Finite Field of size 2
4     L=1;
     Lh=1;
     print "L =", L;
     print "L h =", Lh;
9     L = 1
     L h = 1
1     L h = 1
1     I = ideal(x1+x2, x2+(L*x4+Lh*x4*L*Lh)+(x2*(1+x3)*(1+x4))+(L*x4+Lh*x4*L*Lh)*(x2*(1+x3)*(1+x4)), x3+(1+x4*L)*(1+Lh)*x2, x4*x1*x4*(14*x6)*x1*x4*(14*x6), x5+(1+x1)*x4, x6+(1+x1)*x5); I
1     Ideal (x1 + x2, x2 + 1, x3, x1*x4*x6 + x1*x4 + x1 + x4*x6, x1*x4 + x4 + x5, x1*x5 + x6) of Multivar iate Polynomial Ring in x1, x2, x3, x4, x5, x6 over Finite Field of size 2
1     B = I.groebner_basis(); B
1     [x1 + 1, x2 + 1, x3, x4 + 1, x5, x6]
```

Conclusion: There is a unique fixed point,

$$(M, A, A_{\text{old}}, B, B_{\text{old}(1)}, B_{\text{old}(2)}) = (x_1, x_2, x_3, x_4, x_5, x_6) = (1, 1, 0, 1, 0, 0)$$

This is exactly what we expected: the *lac* operon is ON.

Using Sage to compute the fixed points (low lactose)

```
P_{\bullet}<x1,x2,x3,x4,x5,x6> = PolynomialRing(GF(2), 6, order = 'lex'); P
 3
       Multivariate Polynomial Ring in x1, x2, x3, x4, x5, x6 over Finite Field of size 2
    L=0:
    Lh=0:
    print "L =", L;
    print "L h =" . Lh:
       L h = 0
    I = ideal(x1+x2, x2+(L*x4+Lh+x4*L+Lh)+(x2*(1+x3)*(1+x4))+(L*x4+Lh+x4*L*Lh)*(x2*(1+x3)*(1+x4)), x3+(1+x4*L)*(x2*(1+x3)*(1+x4))
    (1+Lh)*x2, x4+x1+x4*(1+x6)+x1*x4*(1+x6), x5+(1+x1)*x4, x6+(1+x1)*x5); I
       Ideal (x1 + x2, x2*x3*x4 + x2*x3 + x2*x4, x2 + x3, x1*x4*x6 + x1*x4 + x1 + x4*x6, x1*x4 + x4 + x5, x1*x5)
12
        + x5 + x6) of Multivariate Polynomial Ring in x1, x2, x3, x4, x5, x6 over Finite Field of size 2
14
    B = I.groebner basis(); B
       [x1 + x6^2, x2 + x6^2, x3 + x6^2, x4 + x6^5 + x6^4 + x6, x5 + x6^4 + x6, x6^6 + x6^4 + x6^3]
```

We need to backsubstitute. Recall that $x_i^k = x_i$ for all k.

The last equation: $x_6^6 + x_6^4 + x_6^3 = 0$ implies $x_6 = 0$.

Plug this into the previous equation: $x_5 + x_6^4 + x_6 = 0$ (with $x_6 = 0$) implies $x_5 = 0$.

And so on. We get a unique fixed point:

$$(M, A, A_{\text{old}}, B, B_{\text{old}(1)}, B_{\text{old}(2)}) = (x_1, x_2, x_3, x_4, x_5, x_6) = (0, 0, 0, 0, 0, 0)$$

This is exactly what we expected: the *lac* operon is OFF.

Using Sage to compute the fixed points (medium lactose)

```
P.<x1,x2,x3,x4,x5,x6> = PolynomialRing(GF(2), 6, order = 'lex'); P
3
       Multivariate Polynomial Ring in x1, x2, x3, x4, x5, x6 over Finite Field of size 2
   L=1:
   Lh=0;
   print "L =", L;_
   print "L h =", Lh;
       L = 1
       L h = 0
   I = ideal(x1+x2, x2+(L*x4+Lh+x4*L*Lh)+(x2*(1+x3)*(1+x4))+(L*x4+Lh+x4*L*Lh)*(x2*(1+x3)*(1+x4)), x3+(1+x4*L)*
    (1+Lh)*x2, x4+x1+x4*(1+x6)+x1*x4*(1+x6), x5+(1+x1)*x4, x6+(1+x1)*x5); I
12
       Ideal (x1 + x2, x2*x3*x4^2 + x2*x3 + x2*x4^2 + x4, x2*x4 + x2 + x3, x1*x4*x6 + x1*x4 + x1 + x4*x6, x1*x4
       + x4 + x5, x1*x5 + x5 + x6) of Multivariate Polynomial Ring in x1, x2, x3, x4, x5, x6 over Finite Field
        of size 2
   B = I.groebner basis(); B
15
       [x1 + x4 + x6^9 + x6^8 + x6^5 + x6^4, x2 + x4 + x6^9 + x6^8 + x6^5 + x6^4, x3 + x6^9 + x6^5, x4^2 + x4 + x6^6]
       x6^{11} + x6^{10} + x6^{9} + x6^{8} + x6^{6}, x4*x6 + x6^{10} + x6^{9} + x6^{6} + x6^{2}, x5 + x6^{8} + x6^{4}, x6^{12} + x6^{9}
       + x6^5 + x6^4 + x6
```

The last (7th) equation implies $x_6 = 0$. The 6th one then implies $x_5 = 0$.

The 5th equation gives no information (x_4 can be anything), as does the 4th ($x_4^2 + x_4 = 0$).

The 3rd equation says $x_3 = 0$.

The 2nd equation says $x_2 = x_4$, and the 1st equation says $x_1 = x_4$.

We get two fixed points:

$$(M, A, A_{\mathsf{old}}, B, B_{\mathsf{old}(1)}, B_{\mathsf{old}(2)}) = (x_1, x_2, x_3, x_4, x_5, x_6) = (0, 0, 0, 0, 0, 0), \text{ or } (1, 1, 0, 1, 0, 0).$$

Fixed points of our model and bistability

Here is a table showing the fixed points of our model, depending on whether extracellular lactose levels are low, medium, or high.

Inducer level	L	L_{high}	M	В	$B_{old(1)}$	$B_{\text{old}(2)}$	Α	A_{old}	operon
Low lactose	0	0	0	0	0	0	0	0	OFF
High lactose	1	1	1	1	0	0	1	0	ON
Medium lactose	1	0	0	0	0	0	0	0	OFF
Medium lactose	1	0	1	1	0	0	1	0	ON

Suppose lactose concentration is low ($L=L_{\rm high}=0$), and so the operon is OFF. The current state is

$$(M, A, A_{\mathsf{old}}, B, B_{\mathsf{old}(1)}, B_{\mathsf{old}(2)}) = (x_1, x_2, x_3, x_4, x_5, x_6) = (0, 0, 0, 0, 0, 0),$$

Now, let's change L from 0 to 1, increasing the lactose level to medium. We are now in the 3rd fixed point above, and so the operon is still OFF.

Conversely, suppose lactose concentration is high ($L=L_{\rm high}=1$), and so the operon is ON. The current state is

$$(M, A, A_{\text{old}}, B, B_{\text{old}(1)}, B_{\text{old}(2)}) = (x_1, x_2, x_3, x_4, x_5, x_6) = (1, 1, 0, 1, 0, 0),$$

Now, let's change L_{high} from 1 to 0, reducing the lactose level to medium. This takes us to the 4th fixed point above, and so the operon is still ON.

A Boolean model incorporating dilution & degradation, and time-delays

Instead of the a "middle value" (.135 min⁻¹), let's choose the high estimate $\gamma_A = .52 \text{ min}^{-1}$.

This makes the half-life of A (which was $\widetilde{h_A} = 15.753$) much smaller:

$$\widetilde{h_A} = \frac{\ln 2}{\widetilde{\gamma_A}} = 1.260, \qquad \widetilde{h_B} = \frac{\ln 2}{\widetilde{\gamma_B}} = 22.360 \qquad \widetilde{h_M} = \frac{\ln 2}{\widetilde{\gamma_M}} = 1.5$$

In this case, let's choose a much smaller time-step (e.g., t = 1 min).

We can no longer ignore all of the time-delays, so we introduce the following new variables:

- M_1 , M_2 to model the delayed effect (by $\tau_B=2$ min) of mRNA on the production of β -galactosidase.
- A_1 to model the delayed action of A on the production of mRNA by $\tau_M = .1$ min.

We will use the following new variables to model dilution & degradation:

- M_{old} since $\widetilde{\gamma_M} = 1.5$ is approximately 1 time-step.
- A_{old} since $\widetilde{\gamma_A} = 1.26$ is approximately 1 time-step.
- $B_{\text{old}(1)}$, $B_{\text{old}(2)}$ since loss of β -galactosidase is slower.

Remark

We really should use more variables, e.g., $B_{\text{old}(1)}, B_{\text{old}(2)}, \dots, B_{\text{old}(22)}$ to accurately track the loss of β -galactosidase. However, we will argue shortly why this won't matter.

A Boolean model incorporating dilution & degradation, and time-delays

Proposed model

$$\begin{array}{lll} f_{M} = A_{1} \vee (M \wedge \overline{M_{\text{old}}}) & f_{A_{1}} = A \\ f_{M_{1}} = M & f_{A_{\text{old}}} = \left((\overline{B} \vee \overline{L}) \wedge \overline{L_{\text{high}}} \right) \wedge A \\ f_{M_{2}} = M_{1} & f_{B} = M_{2} \vee \left(B \wedge \overline{B_{\text{old}(2)}} \right) \\ f_{M_{\text{old}}} = \overline{A_{1}} \wedge M & f_{B_{\text{old}(1)}} = \overline{M_{2}} \wedge B \\ f_{A} = (B \wedge L) \vee L_{\text{high}} \vee (A \wedge \overline{A_{\text{old}}} \wedge \overline{B}) & f_{B_{\text{old}(2)}} = \overline{M_{2}} \wedge B_{\text{old}(1)} \end{array}$$

Analysis of the long-term behavior of this model leads to similar results as the previous one.

Lactose	L	Lhigh	М	M_1	M ₂	$M_{\rm old}$	В	$B_{old(1)}$	$B_{old(2)}$	Α	A_1	A_{old}
Low	0	0	0	0	0	0	0	0	0	0	0	0
High	1	1	1	1	1	0	1	0	0	1	1	0
Medium	1	0	0	0	0	0	0	0	0	0	0	0
Medium	1	0	1	1	1	0	1	0	0	1	1	0

A Boolean version of the 5-variable ODE model

5-variable ODE model (Yildirim and Mackey, 2004)

Let M(t)= mRNA, $B(t)=\beta$ -galactosidase, A(t)= allolactose, P(t)= lac permease, L(t)= lactose (concentrations). Extracellular lactose (L_e) is a parameter.

$$\begin{split} \frac{dM}{dt} &= \alpha_M \frac{1 + K_1 (e^{-\mu \tau_M} A_{\tau_M})^n}{K + K_1 (e^{-\mu \tau_M} A_{\tau_M})^n} + \Gamma_0 - \widetilde{\gamma_M} M \\ \frac{dB}{dt} &= \alpha_B e^{-\mu \tau_B} M_{\tau_B} - \widetilde{\gamma_B} B \\ \frac{dA}{dt} &= \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{A}{K_A + A} - \widetilde{\gamma_A} A \\ \frac{dP}{dt} &= \alpha_P e^{-\mu (\tau_B + \tau_P)} M_{\tau_B + \tau_P} - \widetilde{\gamma_P} P \\ \frac{dL}{dt} &= \alpha_L P \frac{L_e}{K_{L_e} + L_e} - \beta_{L_e} P \frac{L}{K_{L_e} + L} - \alpha_A B \frac{L}{K_L + L} - \widetilde{\gamma_L} L \end{split}$$

We'll use the same estimates for degradation and delay constants as in the 3-variable model:

$$\mu=.03~\mathrm{min}^{-1}, \qquad \widetilde{\gamma_{A}}=\gamma+\mu=.044, \qquad \widetilde{\gamma_{B}}=\gamma+\mu=.031, \qquad \widetilde{\gamma_{M}}=\gamma+\mu=.441.$$

New degradation constants estimated at $\gamma_L = 0.0 \text{ min}^{-1}$, and $\gamma_P = .65 \text{ min}^{-1}$. Delay constant estimate is $\tau_P = .83 \text{ min}$.

We need a new parameter to help distinguish high vs. medium extracellular lactose: Lehigh.

A Boolean version of the 5-variable ODE model

Model assumptions

- Variables are M, B, A, P, L.
- lacksquare Glucose absent. Extracellular lactose present, two parameters: L_e and $L_{e_{high}}$.
- Ignore time-delays (Yildirim and Mackey showed that they do not affect bistability).
- Time-step \approx 10 min.
- Ignore (all \ll 10): $\tau_M = .10$ min, $\tau_B = 2$ min, $\widetilde{h_M} = 1.572$ min.
- Introduce dilution & degradation variables: A_{old}, B_{old}, L_{old}, P_{old}.

Proposed model

Justification for f_A

$$f_A = (B \wedge L) \vee (L \wedge L_{e_{high}}) \vee \left(A \wedge \overline{A_{old}} \wedge \overline{B}\right)$$

There are 3 ways for allolactose to be available at t + 1:

- (i) β -galactosidase and at least medium levels of lactose are present.
- (ii) Internal lactose is present and the concentration of extracellular lacatose is high. This ensures that by time t+1, intracellular lactose concentration is high enough to find available trace amounts of β -galactosidase.
- (iii) The concentration of allolactose is high enough that it wont' be reduced below the threshold due to dilution & degradation, or to conversion (by β -galactosidase) to glucose & galctose.

Justification for f_L

$$\mathit{f}_{\mathit{L}} = \left((P \, \land \, \mathit{L}_{e}) \, \lor \, \mathit{L}_{e_{\mathsf{high}}} \right) \lor \left((\mathit{L} \, \land \, \overline{\mathit{L}_{\mathsf{old}}}) \, \land \, (\overline{\mathit{B}} \, \land \, \overline{\mathit{P}}) \right)$$

There are 3 ways for intracellular lactose to be available at t + 1:

- (i) Lac permease and extracellular lactose are available.
- (ii) There are high levels of extracellular lactose available (even if *lac* permease level is low).
- (iii) There is enough lactose in the cell that it won't be lost to dilution & degradaton, transport out, or conversion into allolactose (by β -galactosidase).

Model:

Fixed points:

Ext. Lactose	Le	$L_{e_{\mathrm{high}}}$	М	M _{old}	В	B_{old}	Α	$A_{\rm old}$	L	$L_{\rm old}$	Р	P_{old}
Low	0	0	0	0	0	0	0	0	0	0	0	0
High	1	1	1	0	1	0	1	0	1	0	1	0
Medium	1	0	0	0	0	0	0	0	0	0	0	0
Medium	1	0	1	0	1	0	1	0	1	0	1	0