Stochastic CRNs

Lecture 2
Asymptotic Reductions

Grzegorz Rempala
The Ohio State University

Statystyka Matematyczna XLV
Będlewo (POL)
Dec 5th 2019

1 Asymptotics for CRNs

- $X(t)$ number of molecules of each species in the system at time t
- ν_k number of molec's of each chemical species consumed in the kth reaction
- ν'_k number of molecules of each species created by the kth reaction
- the rate λ_k^N is

$$\lambda_k^N(x) = \kappa_k \frac{\prod_i \nu_{ik}!}{N^{|\nu_k|-1}} \prod_i \binom{x_i}{\nu_{ik}},$$

where $|\nu_k| = \sum_i \nu_{ik}$ and N is a *scaling parameter*

$$X(t) \;=\; X(0) + \sum_k Y_k\Big(\int_0^t \lambda_k^N(X(s))ds\Big)(\nu'_k - \nu_k)$$

Y_k's are independent unit PPs

- N is often (but not always) the volume of the system times Avogadro's number
- Previous lecture: deterministic approximation given by RRE (mean-field approximation, LLN)

1.1 Diffusion approximation

An appropriately renormalized Poisson process can be approximated by a standard Brownian motion

$$\frac{Y(Nu) - Nu}{\sqrt{N}} \approx W(u),$$

Setting $C^N(t) = X(t)/N$ and replacing $Y_k(Nu)$ by $\sqrt{N}W_k(u) + Nu$ we obtain the *Langevin equation*

$$\begin{aligned} C^N(t) &= C^N(0) + \sum_k N^{-1} Y_k\Big(\int_0^t \lambda_k(X^N(s))ds\Big)(\nu'_k - \nu_k) \\ &\approx C^N(0) + \sum_k N^{-1/2}\, W_k\Big(\int_0^t \tilde{\lambda}_k(C^N(s))ds\Big)(\nu'_k - \nu_k) \\ &\qquad + \int_0^t F(C^N(s))ds, \end{aligned}$$

where $F(c) = \sum_k \tilde{\lambda}_k(c)(\nu'_k - \nu_k)$.

1.2 Equivalent form – Itô formulation

The diffusion approximation is given by the equation

$$\tilde{C}^N(t) = \tilde{C}^N(0) + \sum_k N^{-1/2} W_k \left(\int_0^t \tilde{\lambda}_k(\tilde{C}^N(s))ds \right) (\nu'_k - \nu_k) + \int_0^t F(\tilde{C}^N(s))ds,$$

which is *distributionally equivalent* to the Itô equation

$$\begin{aligned} \tilde{C}^N(t) &= \tilde{C}^N(0) + \sum_k N^{-1/2} \int_0^t \sqrt{\tilde{\lambda}_k(\tilde{C}^N(s))} d\tilde{W}_k(s)(\nu'_k - \nu_k) \\ &\qquad + \int_0^t F(\tilde{C}^N(s))ds \\ &= \tilde{C}^N(0) + N^{-1/2} \int_0^t \sigma(\tilde{C}^N(s))d\tilde{W} + \int_0^t F(\tilde{C}^N(s))ds, \end{aligned}$$

where $\sigma(c)$ is the matrix with columns $\sqrt{\tilde{\lambda}_k(c)}(\nu'_k - \nu_k)$.

1.3 Multiple Scales: Cellular Systems

- The number of molecules involved, at least for some of the species, may be small
- Law of large numbers may not apply
- The deterministic model may does not provide a good representation of the behavior of the system
- Some species may be present in much greater abundance than others.
- The rate constants κ_k may vary over several orders of magnitude.
- Idea: use κ_k as guidance for the scaling N;
- NOTE: in what follows N is no longer volume ..

Multiple Scales (cont)

- Take N to be of the order of magnitude of the abundance of the most abundant species in the system.
- For each species i, $0 \leq \alpha_i \leq 1$ and

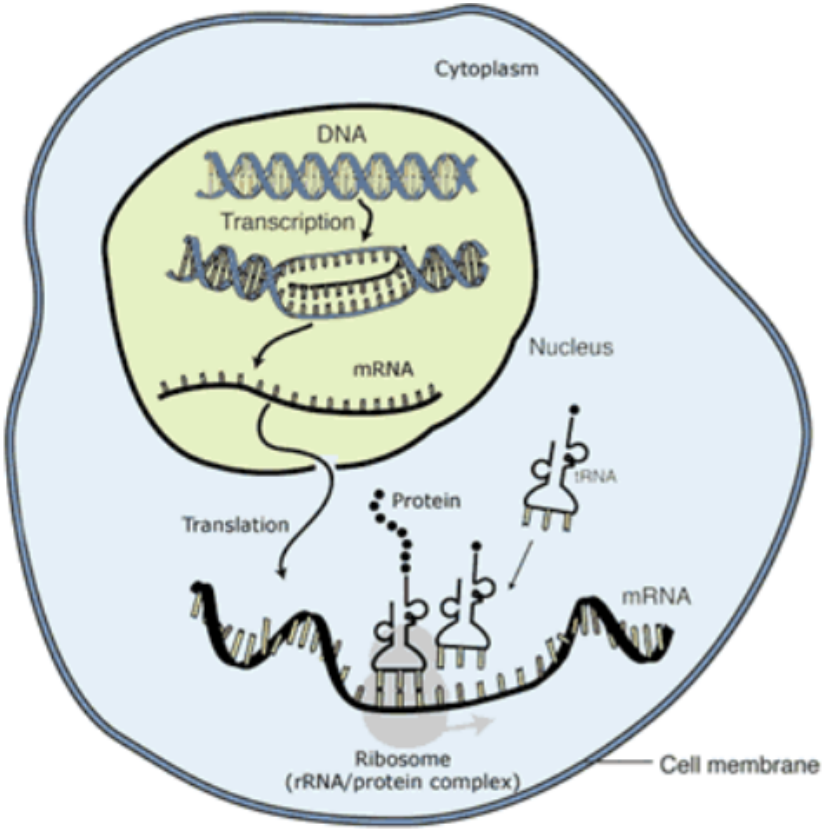
$$Z_i(t) = N^{-\alpha_i} X_i(t).$$

  α_i should be selected so that $Z_i = O(1)$.

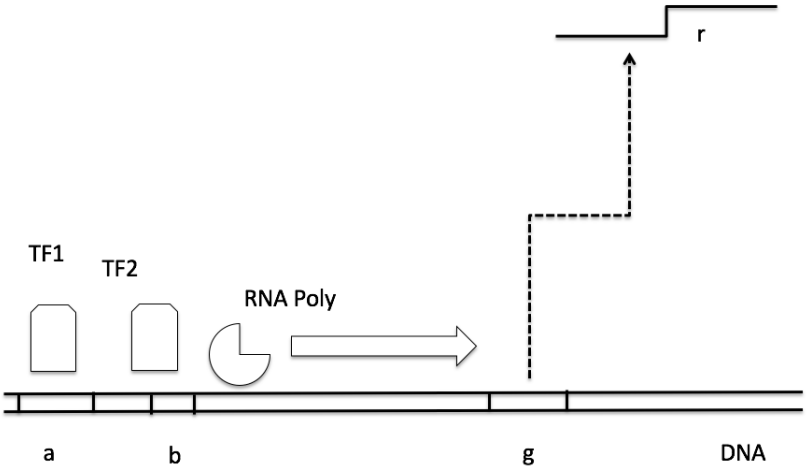
- Express the reaction rates in terms of Z rather than X and also take into account large variation in the reaction rates.
- Select β_k so that the reaction rates can be written as $N^{\beta_k}\lambda_k(z)$, where $\lambda_k(z) = O(1)$ for all relevant values of z.
- The model becomes

$$Z_i(t) = Z_i(0) + \sum_k N^{-\alpha_i} Y_k\Big(\int_0^t N^{\beta_k}\lambda_k^N(Z(s))ds\Big)(\nu'_{ik} - \nu_{ik})$$

2 Multi-scaling Approximation for Transcription and Translation



Transcription and translation (cont)



$$
\begin{aligned}
G + TF1 &\rightleftharpoons TF1 \cdot G \quad \text{(frame opening initiated)} \\
TF1 \cdot G + TF2 &\rightleftharpoons TF2 \cdot TF1 \cdot G \quad \text{(length negotiated)} \\
RNAP + TF2 \cdot TF1 \cdot G &\rightleftharpoons RNAP \cdot TF2 \cdot TF1 \cdot G \quad \text{(gene unlocking)} \\
RNAP \cdot TF2 \cdot TF1 \cdot G &\rightharpoonup RNAP \cdot TF2 \cdot TF1 \cdot G(1) \quad \text{(transcr. starts)} \\
&\vdots \quad \vdots \quad \vdots \\
RNAP \cdot TF2 \cdot TF1 \cdot G(k) &\rightharpoonup RNAP + TF2 \cdot TF1 \cdot G + R \quad \text{(transcr. ends)}
\end{aligned}
$$

2.1 Prokaryotic Model: Simplified TT Process

- Consider prokaryotic-type simplified system
- Interested in the amount of protein P instead of transcript R
- Averaging over multiple mol's of polymerase and rybosome ..
- Will used delay to skip over multiple production steps..
- Unlocking is on faster time scale then the rest of the system..

$$
\begin{array}{rcll}
G & \xrightarrow{\kappa_1} & P+G & \text{protein production (with delay, fast } \kappa_1 = Nk_1) \\
P+G & \xrightarrow{\kappa_2} & C & \text{gene frame locking, slow } \kappa_2 = k_2 \\
C & \xrightarrow{\kappa_3} & P+G & \text{gene frame unlocking (fast } \kappa_3 = Nk_3) \\
P & \xrightarrow{\kappa_4} & \emptyset & \text{protein deletion, slow } \kappa_4 = k_4
\end{array}
$$

Here $C = P \cdot G$ is a PG complex.

- Let $X(t) = (X_P, X_G, X_C)(t)$ with $\{X_C(0), X_G(0)\} = \{0, 1\}$ then

$$X_C(t) + X_G(t) = 1$$

TT process equations

- Define $X_G(t) \equiv 0$ for $t < 0$, $X_G(0) = 1, X_C(0) = 0$
- Then for $t \geq 0$

$$
\begin{aligned}
X_P(t) &= X_P(0) + Y_1\left(\int_0^{t-\tau} \kappa_1 X_G(s)\,ds\right) + Y_3\left(\int_0^t \kappa_3 X_C(s)\,ds\right) \\
&\quad - Y_2\left(\int_0^t \kappa_2 X_P X_G(s)\,ds\right) - Y_4\left(\int_0^t \kappa_4 X_P(s)\,ds\right) \\
X_G(t) &= X_G(0) - Y_2\left(\int_0^t \kappa_2 X_P X_G(s)\,ds\right) + Y_3\left(\int_0^t \kappa_3 X_C(s)\,ds\right) \\
X_C(t) &= X_C(0) + Y_2\left(\int_0^t \kappa_2 X_P X_G(s)\,ds\right) - Y_3\left(\int_0^t \kappa_3 X_C(s)\,ds\right)
\end{aligned}
$$

  where Y_1, Y_2, Y_3 and Y_4 are independent unit Poisson processes.

- Last equation simply states that $X_C(t) = 1 - X_G(t)$ for $t \geq 0$.

Multiscale equations

- Recall $\kappa_1 = Nk_1, \kappa_2 = k_2, \kappa_3 = Nk_3, \kappa_4 = k_4$
- Set $Z_P(t) = X_P(t)/N$ for $t \geq 0$
- Assume $Z_P(0) \to z_P(0)$ as $N \to \infty$

$$Z_P(t) = Z_P(0) + N^{-1}Y_1\left(N\int_0^{t-\tau} k_1 X_G(s)\,ds\right) + N^{-1}Y_3\left(N\int_0^t k_3 X_C(s)\,ds\right)$$
$$- N^{-1}Y_2\left(N\int_0^t k_2 Z_P(s)X_G(s)\,ds\right) - N^{-1}Y_4\left(N\int_0^t k_4 Z_P(s)\,ds\right)$$

$$N^{-1}X_G(t) = N^{-1}X_G(0) - N^{-1}Y_2\left(N\int_0^t k_2 Z_P(s)X_G(s)\,ds\right)$$
$$+ N^{-1}Y_3\left(N\int_0^t k_3 X_C(s)\,ds\right) \to 0$$

(Since $N^{-1}X_G(t) \to 0$ as $N \to \infty$ for any $t \geq 0$.)

Multiscale equations (cont)

- From the last equation it follows that as $N \to \infty$

$$\int_0^t k_2 Z_P(s) X_G(s)\, ds - \int_0^t k_3 X_C(s)\, ds \to 0 \qquad \text{for any } t > 0$$

- So $Z_P(t)$ must be uniformly bounded on any finite interval $[0, T]$
- It follows that $Z_P(t)$ has a convergent subsequence with limit $z_P(t)$ and along that subsequence

$$\int_0^t X_G(s) ds \to \int_0^t \frac{k_3}{k_3 + k_2 z_P(s)} ds$$

- Here we used the fact that $X_C(t) = 1 - X_G(t)$
- Will now return to the first equation, considering separately cases $t < \tau$ and $t > \tau$

ODE Approximation

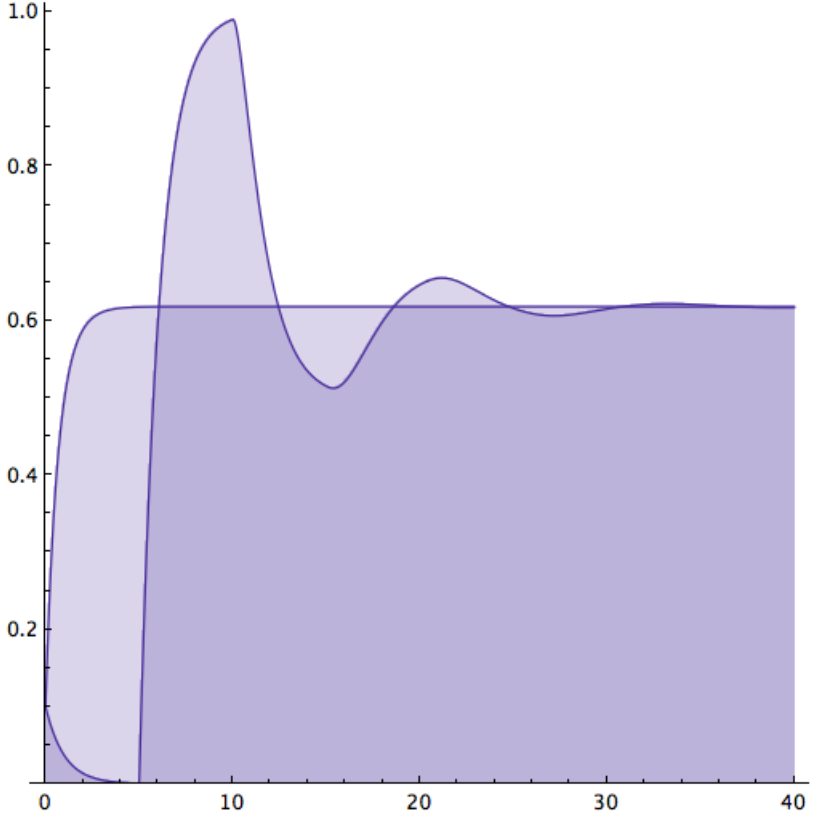
- Applying PLLN to the first equation along the same subsequence:
  - For $0 < t \leq \tau$
$$z_P(t) = z_P(0) - \int_0^t k_4 z_P(s) ds$$
  - For $t > \tau$
$$z_P(t) = z_P(0) + \int_0^{t-\tau} \frac{k_1 k_3}{k_3 + k_2 z_P(s)} ds - \int_0^t k_4 z_P(s) ds$$
- Uniqueness of the above solution gives the existence of the limit

Proposition 2.1 (Multiscale ODE approximation to TT process)**.** *As $N \to \infty$, the relative amt of protein in the system X_P/N converges UIP on bounded intervals to the deterministic function z_P which satisfies the delay differential equation*

$$\dot{z}_P(t) = \frac{k_1 k_3}{k_3 + k_2 z_P(t-\tau)} \mathbb{1}[t > \tau] - k_4 z_P(t)$$

$$z_P(0) \geq 0$$

Effect of delay



BD approximation

- Let $X(t) = (X_G(t), X_P(t), X_C(t))$ be the original TT process
- Then X_P may be approximated by BD process $\tilde{X}_P$

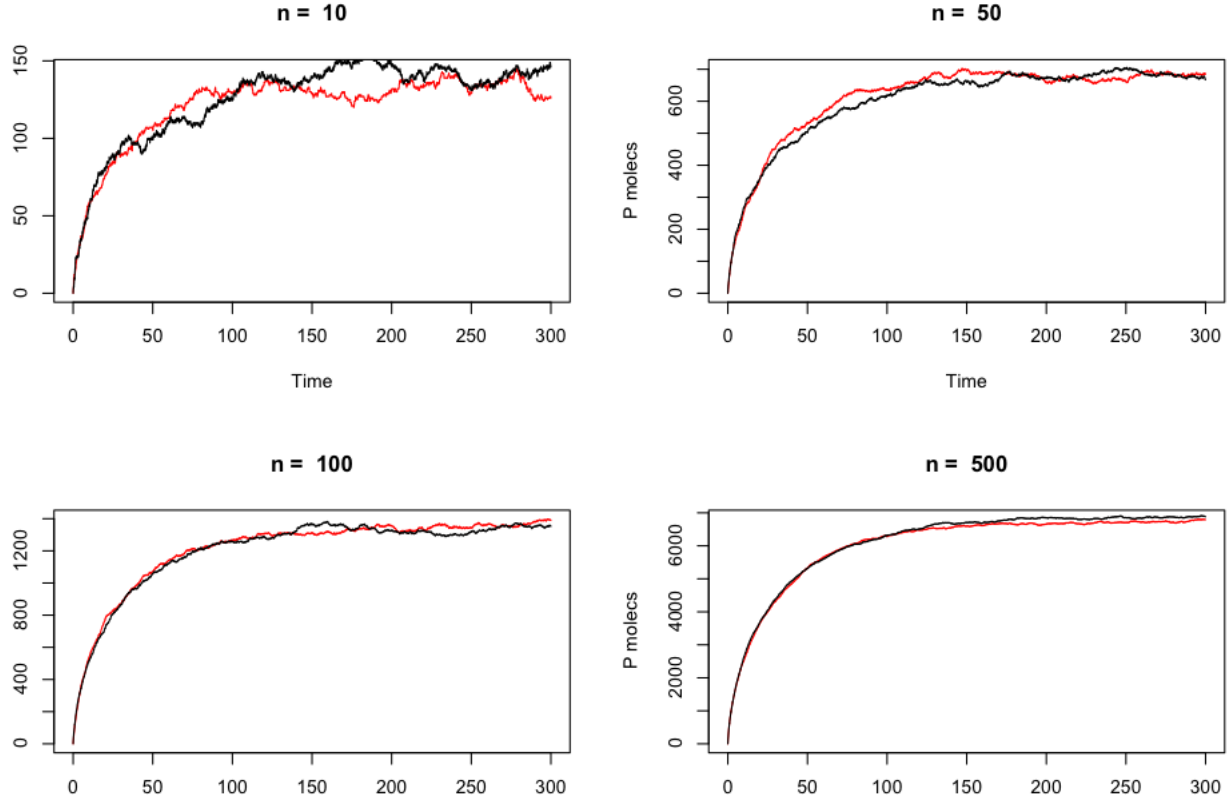
$$\tilde{X}_P(t) = \tilde{X}_P(0) + \tilde{Y}_1\left(\int_0^{t-\tau} \frac{N^2 k_1 k_3}{N k_3 + k_2 \tilde{X}_P(s)} ds\right) - \tilde{Y}_2\left(k_4 \int_0^t \tilde{X}_P(s) ds\right)$$

  where $\tilde{Y}_1$ and $\tilde{Y}_2$ are independent of X

- We can show that as $N \to \infty$

$$\sup_{0<t\leq T} N^{-1}|X_P(t) - \tilde{X}_P(t)| \xrightarrow{P} 0$$

$X(t) \approx \tilde{X}(t)$ **for large** N $(\tau = 0)$



2.2 TT Model: Fast Switching

- Locking/Unlocking is on faster time scale then the rest of the system..
- Interested in the *total* amount of protein P
- For the sake of simplicity: no delay (could handle delay as before)

$$
\begin{array}{rcll}
G & \xrightarrow{\kappa_1} & P+G & \text{protein production (SLOW } \kappa_1 = k_1) \\
P+G & \xrightarrow{\kappa_2} & C & \text{gene frame locking, FAST } \kappa_2 = Nk_2 \\
C & \xrightarrow{\kappa_3} & P+G & \text{gene frame unlocking (fast } \kappa_3 = Nk_3) \\
P & \xrightarrow{\kappa_4} & \emptyset & \text{protein deletion, slow } \kappa_4 = k_4
\end{array}
$$

Here as before $C = P \cdot G$ is a PG complex as we assume as before

$$X_C(0) + X_G(0) = 1$$

FS process equations

- For $t \geq 0$ as before

$$X_T(t) = X_P(0) + Y_1\left(\int_0^t \kappa_1 X_G(s)\,ds\right) + Y_3\left(\int_0^t \kappa_3 X_C(s)\,ds\right) - Y_2\left(\int_0^t \kappa_2 X_P X_G(s)\,ds\right) - Y_4\left(\int_0^t \kappa_4 X_P(s)\,ds\right)$$

$$X_C(t) = X_C(0) + Y_2\left(\int_0^t \kappa_2 X_P X_G(s)\,ds\right) - Y_3\left(\int_0^t \kappa_3 X_C(s)\,ds\right)$$

$$X_G(t) = X_G(0) - Y_2\left(\int_0^t \kappa_2 X_P X_G(s)\,ds\right) + Y_3\left(\int_0^t \kappa_3 X_C(s)\,ds\right)$$

- Define $X_T(t) = X_P(t) + X_C(t)$, then

$$X_T(t) = X_T(0) + Y_1\left(\int_0^t \kappa_1 X_G(s)\,ds\right) - Y_4\left(\int_0^t \kappa_4 X_P(s)\,ds\right)$$

Multiscale equations

- Recall $\kappa_1 = k_1, \kappa_2 = Nk_2, \kappa_3 = Nk_3, \kappa_4 = k_4$ and
- $X_T(t) = X_P(t) + X_C(t)$ for $t \geq 0$

$$X_T(t) = X_T(0) + Y_1\left(\int_0^t k_1 X_G(s)\,ds\right) - Y_4\left(\int_0^t k_4(X_T(s) - X_C(s))\,ds\right)$$

$$N^{-1}X_G(t) = N^{-1}X_G(0) - N^{-1}Y_2\left(N\int_0^t k_2 X_P(s)X_G(s)\,ds\right) + N^{-1}Y_3\left(N\int_0^t k_3 X_C(s)\,ds\right) \to 0$$

By PLLN, since $X_T(s)X_G(s) = X_P(s)X_G(s)$,

$$\int_0^t k_2 X_T(s)X_G(s)\,ds - \int_0^t k_3 X_C(s)\,ds \to 0$$

and consequently (at least UIP on finite intervals)

$$\int_0^t X_G(s)ds - \int_0^t \frac{k_3}{k_3 + k_2 X_T(s)}ds \to 0$$

Stochastic Approximation for FS

- Since $X_C = 1 - X_G$ then UIP

$$\int_0^t X_C(s)ds - \int_0^t \frac{k_2 X_T(s)}{k_3 + k_2 X_T(s)} ds \to 0$$

and

$$\int_0^t (X_t(s) - X_C(s))ds - \int_0^t X_T(s) \left(1 - \frac{k_2}{k_3 + k_2 X_T(s)}\right) ds \to 0$$

Proposition 2.2 (Multiscale approximation for FS process)**.** *As $N \to \infty$, the amt of total protein in the system X_T converges UIP on bounded intervals to the process Z_T which satisfies the stochastic equation*

$$Z_T(t) = Z_T(0) + Y_1 \left(\int_0^t \frac{k_1 k_3}{k_3 + k_2 Z_T(s)} \, ds \right) - Y_2 \left(\int_0^t Z_T(s) \left(1 - \frac{k_2}{k_3 + k_2 Z_T(s)}\right) ds \right)$$
$$Z_T(0) \geq 0$$

where Y_1, Y_2 are independent unit PP.

3 Virus kinetics (Ball, Popovic, Kurtz, R 2006)

The model includes four time-varying species:

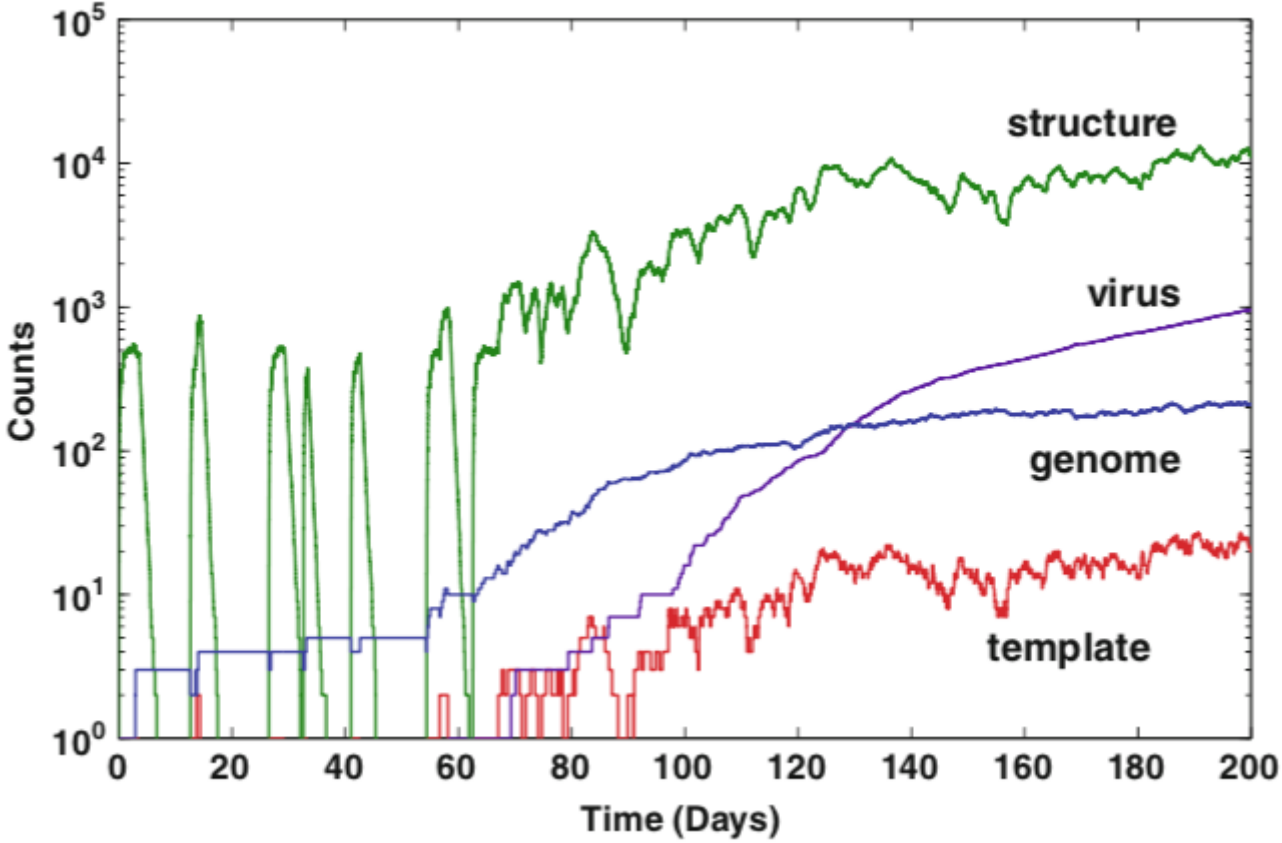
- the viral genome (G),
- the viral structural protein (S),
- the viral template (T), and
- the secreted virus itself (V).

The model has six reactions

$$R_1: \; T \xrightarrow{1} T + G, \quad R_2: \; G \xrightarrow{0.025} T, \quad R_3: \; T \xrightarrow{1000} T + S$$

$$R_4: \; T \xrightarrow{0.25} \emptyset, \qquad R_5: \; S \xrightarrow{2} \emptyset, \qquad R_6: \; G + S \xrightarrow{7.5\times10^{-6}} V$$

Trajectory simulation (Anderson & Kurtz 2015)



Model equations

Denote the species G, S, T, V by 1, 2, 3, and 4, respectively, and let $X_i(t)$ denote the number of molecules of species i at time t.

$$
\begin{aligned}
X_1(t) &= X_1(0) + Y_1\left(\int_0^t X_3(s)ds\right) - Y_2\left(0.025\int_0^t X_1(s)ds\right) \\
&\quad - Y_6\left(7.5\times 10^{-6}\int_0^t X_1(s)X_2(s)ds\right) \\
X_2(t) &= X_2(0) + Y_3\left(1000\int_0^t X_3(s)ds\right) - Y_5\left(2\int_0^t X_2(s)ds\right) \\
&\quad - Y_6\left(7.5\times 10^{-6}\int_0^t X_1(s)X_2(s)ds\right) \\
X_3(t) &= X_3(0) + Y_2\left(0.025\int_0^t X_1(s)ds\right) - Y_4\left(0.25\int_0^t X_3(s)ds\right) \\
X_4(t) &= X_4(0) + Y_6\left(7.5\times 10^{-6}\int_0^t X_1(s)X_2(s)ds\right)
\end{aligned}
$$

Focus on the first three equations.

Rescaling

Taking $N = 1000$, we scale the rate constants so that

κ_1	1	1
κ_2	0.025	$2.5N^{-2/3}$
κ_3	1000	N
κ_4	0.25	0.25
κ_5	2	2
κ_6	7.5×10^{-6}	$0.75N^{-5/3}$

that is, we take $\beta_1 = \beta_4 = \beta_5 = 0, \beta_2 = -2/3, \beta_3 = 1$, and $\beta_6 = -5/3$. Scaling the species numbers, we take $\alpha_1 = 2/3$, $\alpha_2 = 1$, and $\alpha_3 = 0$.

$$
\begin{aligned}
Z_1^N(t) &= Z_1^N(0) + N^{-2/3}Y_1\left(\int_0^t Z_3^N(s)ds\right) - N^{-2/3}Y_2\left(2.5\int_0^t Z_1^N(s)ds\right) \\
&\quad - N^{-2/3}Y_6\left(0.75\int_0^t Z_1^N(s)Z_2^N(s)ds\right) \\
Z_2^N(t) &= Z_2^N(0) + N^{-1}Y_3\left(N\int_0^t Z_3^N(s)ds\right) - N^{-1}Y_5\left(N2\int_0^t Z_2^N(s)ds\right) \\
&\quad - N^{-1}Y_6\left(0.75\int_0^t Z_1^N(s)Z_2^N(s)ds\right)
\end{aligned}
$$

Hybryd model

Assuming $(Z_1^N(0), Z_2^N(0), Z_3^N(0)) \to (Z_1(0), Z_2(0), Z_3(0))$, we have

$$
\begin{aligned}
Z_1^0(t) &= Z_1(0) \\
Z_2^0(t) &= Z_2(0) + \int_0^t Z_3^0(s)ds - 2\int_0^t Z_2^0(s)ds \\
Z_3^0(t) &= Z_3(0) + Y_2\left(2.5\int_0^t Z_1^0(s)ds\right) - Y_4\left(0.25\int_0^t Z_3^0(s)ds\right)
\end{aligned}
$$

- This system is an example of a piecewise deterministic or *hybrid model,*
- one component (Z_3) is discrete and stochastic (here: BD process)
- the other (Z_2) is an ordinary differential equation with coefficients depending on the stochastic component
- As it turns out Z_1 operates on a different time scale ($N^{2/3}t$).

Summary

- Reaction networks LLN and CLT
- General method for multiscale approximation
- Multiscale: both in physical and temporal scales
- We only discussed the physical one here
- Example: prokaryotic transcription-transl. (TT/FS) model
- For TT ODE and BD approximation
- Delay introduces initial perturbation only
- For FS approximation is available as switching rate increases
- Example: Virus infection (hybryd model)

References

[1] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*, volume 1. Springer, 2015.

[2] Karen Ball, Thomas G Kurtz, Lea Popovic, and Grzegorz A Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *The Annals of Applied Probability*, 16(4):1925–1961, 2006.

[3] Hye-Won Kang, Wasiur R KhudaBukhsh, Heinz Koeppl, and Grzegorz A Rempała. Quasi-steady-state approximations derived from the stochastic model of enzyme kinetics. *Bulletin of mathematical biology*, pages 1–34, 2019.

[4] Darren J Wilkinson. *Stochastic modelling for systems biology*. CRC press, 2011.