Towards Bayesian inference in multiple-input motifs

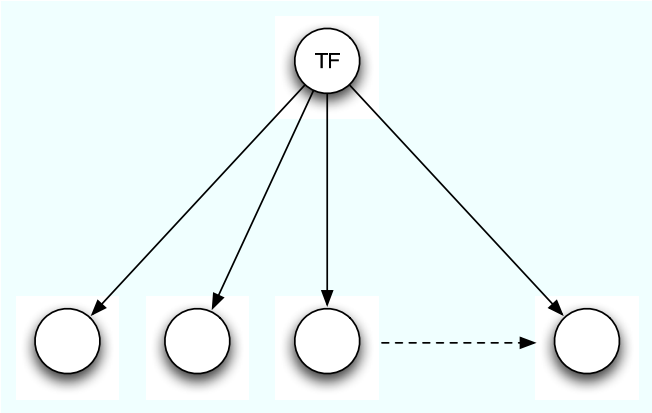
Simon Rogers and Mark Girolami

Bioinformatics Research Centre
Department of Computing Science
University of Glasgow

Overview

- TFA Inference in SIMs
- The full network
- Competitive transcription factors
- An MM-type expression for mRNA production
- Examples
- Conclusions

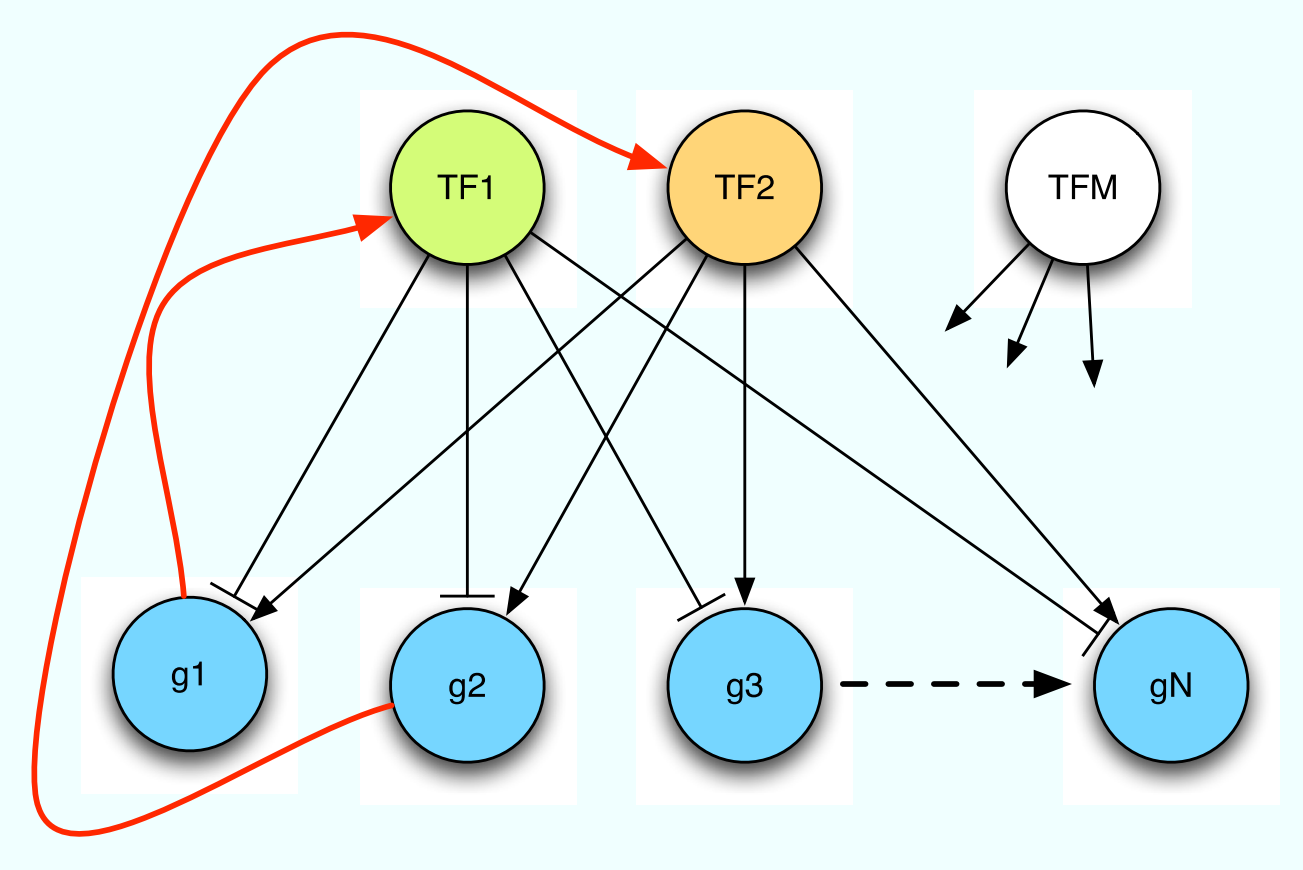
Previously.....TFA inference in SIMs



- For modeling purposes, mRNA expression of TF is **not** reliable proxy of its true activity
- Use expression of target genes and nonlinear model of transcription to *infer* activity over time

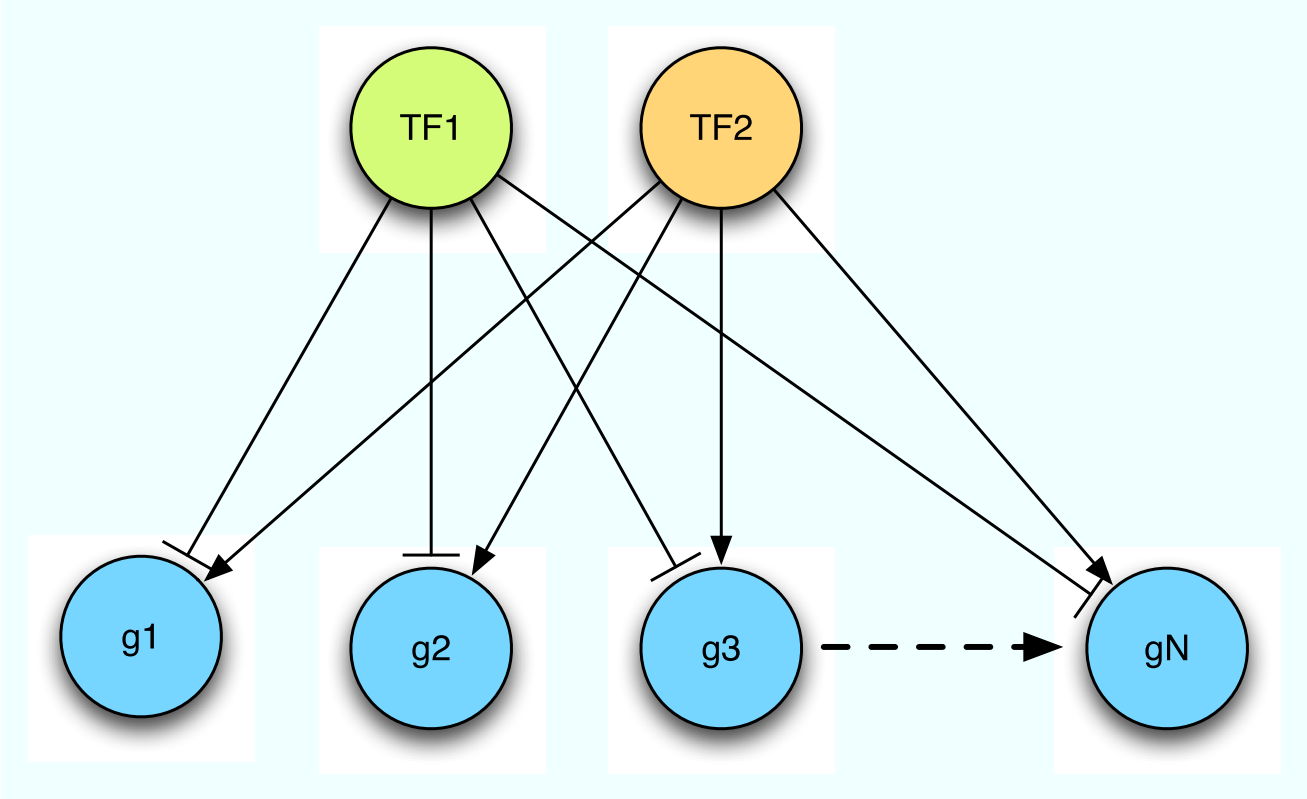
$$\dot{x}_g(t) = \alpha_g + \frac{\beta_g \eta(t)}{K_g + \eta(t)} - \delta_g x_g(t)$$

G/M2 Transition in fission yeast



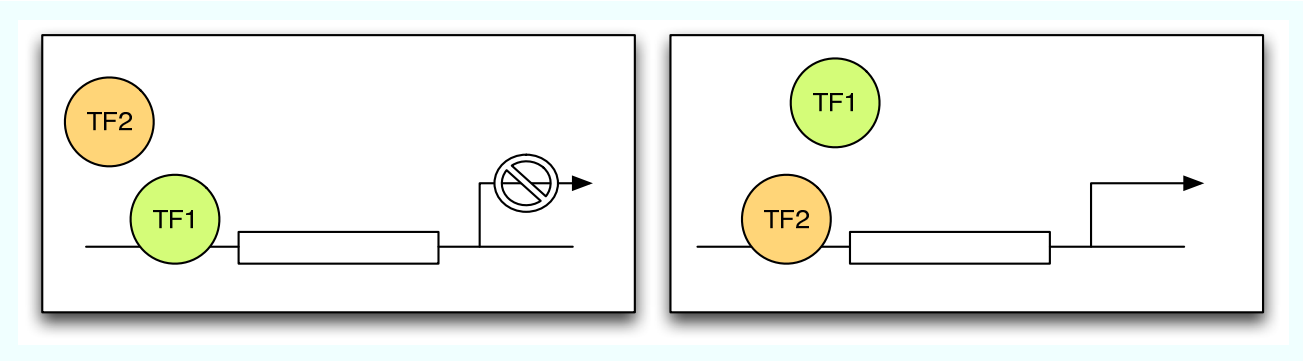
At least 2 TFs, both of which regulate themselves

G/M2 Transition in fission yeast



Starting point - remove feedback, and assume 2 TFs

Competitive transcription factors



- It is thought that the two TFs operate in a competitive manner - both binding to same site on promoters.
- We have continuous data (target mRNA), so would like model of the form

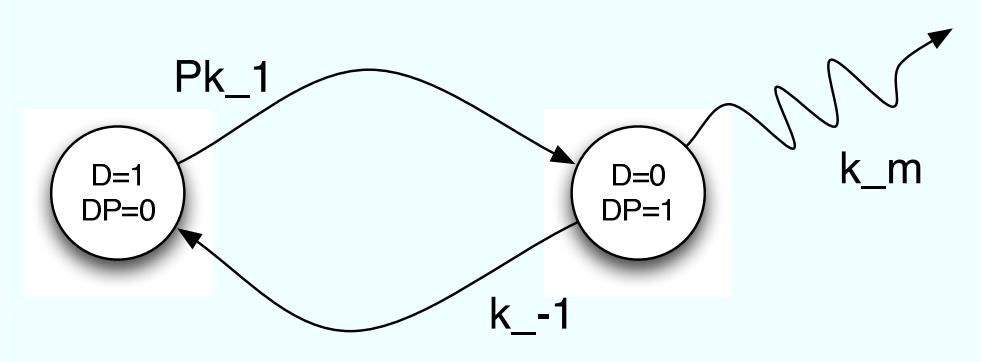
$$\dot{x}_g(t) = \alpha_g + f(\eta_1(t), \eta_2(t), \boldsymbol{\theta}_g) - \delta_g x_g(t)$$

- What is a good $f(\eta_1(t), \eta_2(t), \boldsymbol{\theta}_g)$?
- and **when** is it good?

mRNA production

- Return to SIM, one TF
- We have the following reaction set

$$D + P \xrightarrow{k_1} DP, \quad DP \xrightarrow{k_{-1}} D + P, \quad DP \xrightarrow{k_m} DP + M.$$



- Can be thought of as a two state Continuous time Markov Chain where the time between state transitions is Exp(Pk_1) and Exp(k_{-1}).

Stochastic Quasi-Steady-State assumption

- For large enough T, we can calculate quantity of mRNA produced using expectations with respect to the stationary ditribution.

$$p(DP = 1) = \frac{Pk_1}{Pk_1 + k_{-1}}$$

$$\begin{aligned} p(M_T|\dots) &= p(DP = 1) \times \text{Poiss}(M_T|Tk_m) \\ &= \text{Poiss}(M_T|p(DP = 1)Tk_m) \end{aligned}$$

- Which is equivalent to removing the protein binding-unbinding reactions and modifying the mRNA production rate.

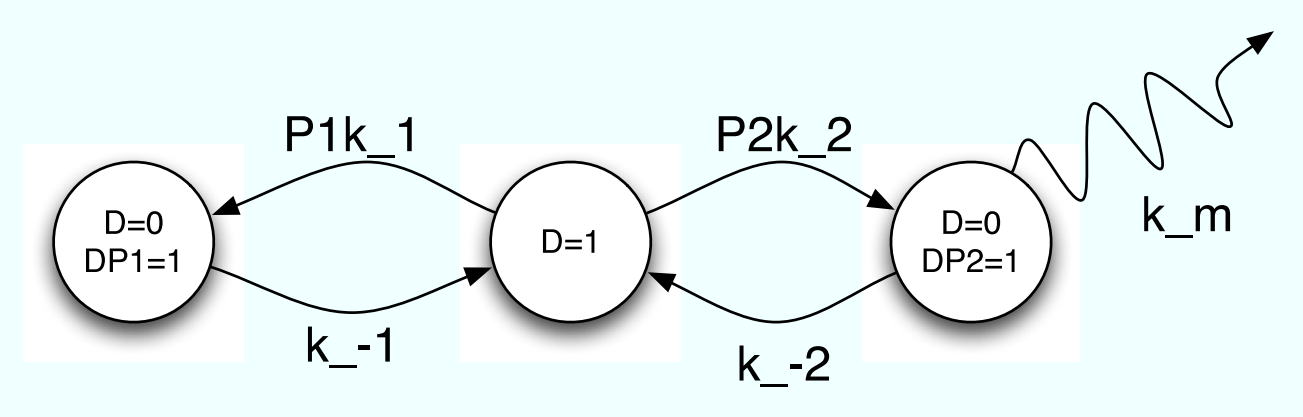
$$\hat{k_m} = k_m p(DP = 1) = k_m \frac{Pk_1}{Pk_1 + k_{-1}}$$

- $\hat{k_m}$ gives the expected quantity of mRNA produced per unit time.
- Using this as our production term in a continuous representation gives us exactly the Michaelis-Menten expression shown previously.

$$\dot{x}_g(t) = \alpha_g + k_{mg} \frac{P}{P + k_{-1g}/k_{1g}} - \delta_g x_g(t)$$

- If we can compute the stationary distribution for a particular motif, we can derive an MM-type expression.
- Approximation is reasonable as long as the probability of being in a particular state can be considered to be stationary.
- i.e., binding and disassociation of protein is fast relative to changes in P.

Competitive transcription factors

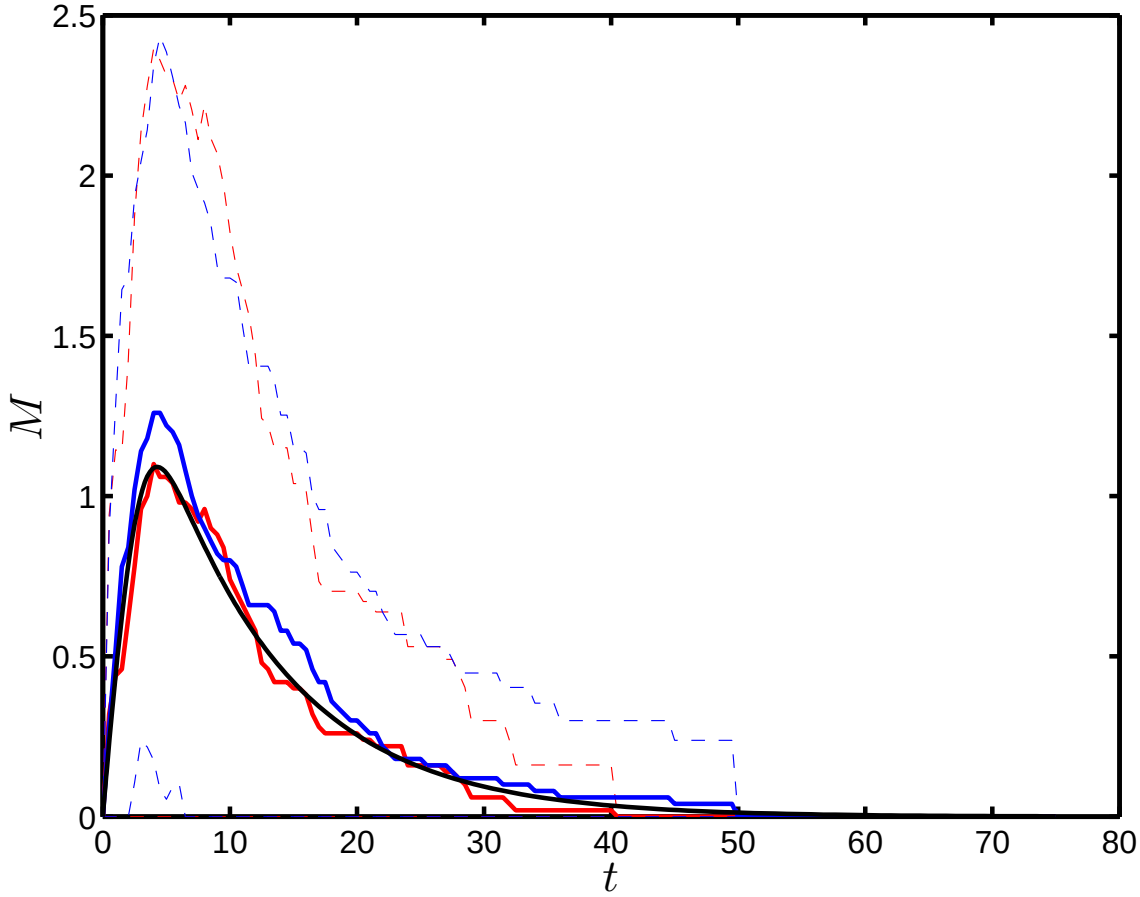


$$p(DP2 = 1) = \frac{k_2 P_2 k_{-1}}{k_2 P_2 k_{-1} + k_{-2}(P_1 k_1 + k_{-1})}$$

$$\hat{k_m} = k_m \frac{P_2}{P_2 + K P_1 + \gamma}$$

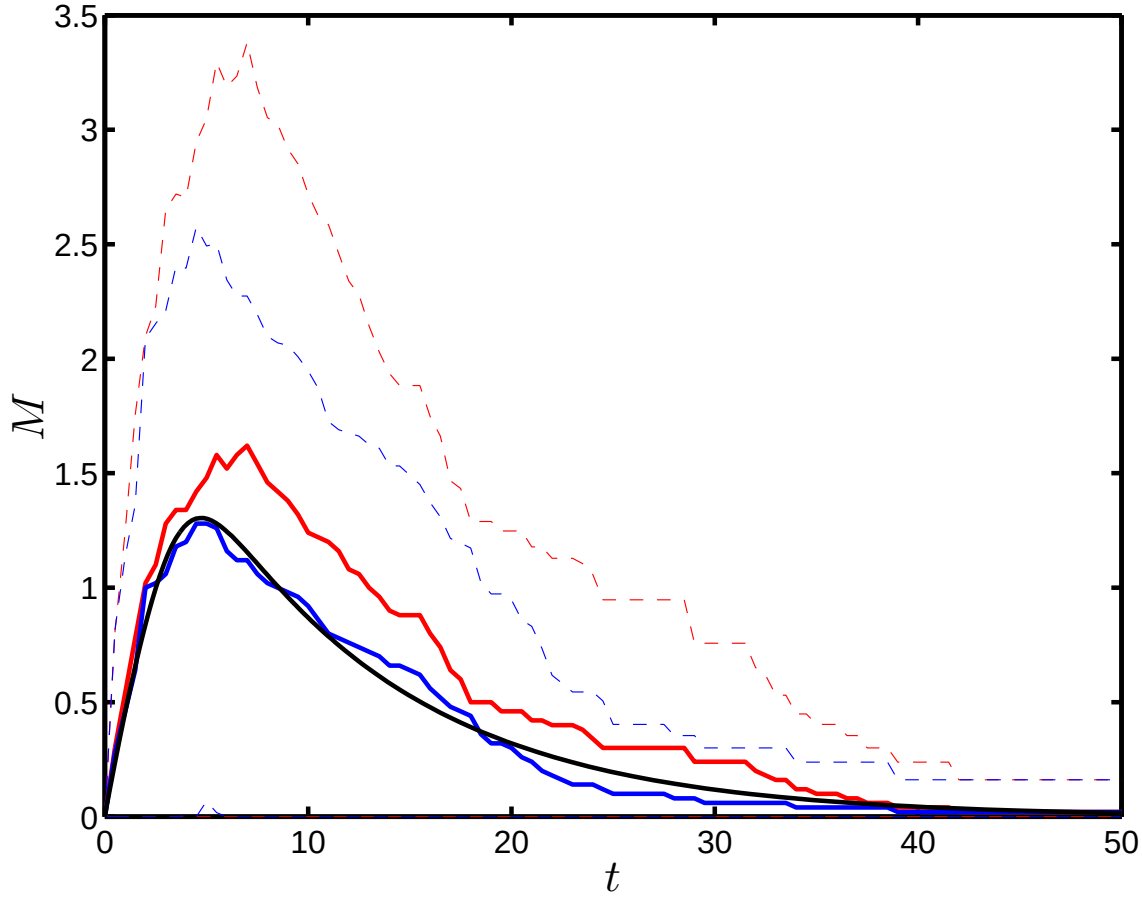
$$\dot{x_g}(t) = \alpha_g + \frac{k_{mg} P_2}{P_2 + K_g P_1 + \gamma_g} - \delta_g x_g(t)$$

Examples - how good is the approximation?



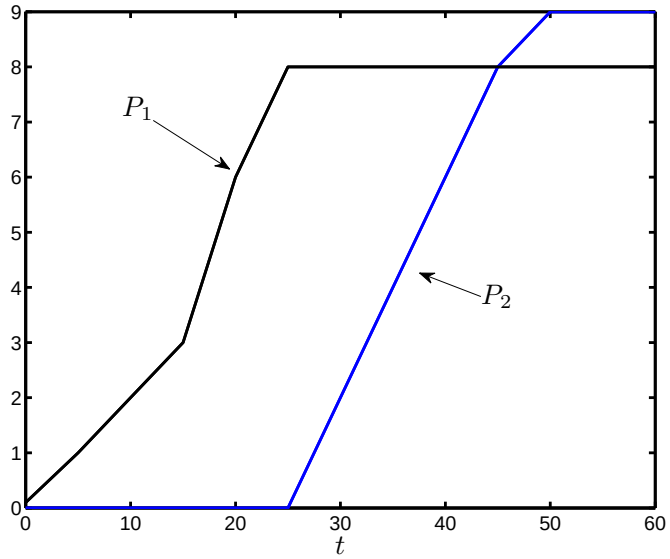
Protein degradation slow with respect to state transition.

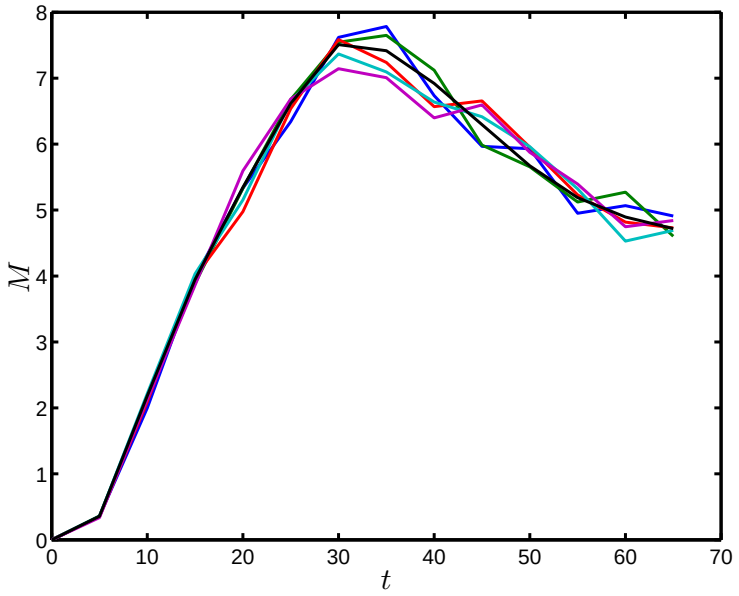
Examples - how good is the approximation?



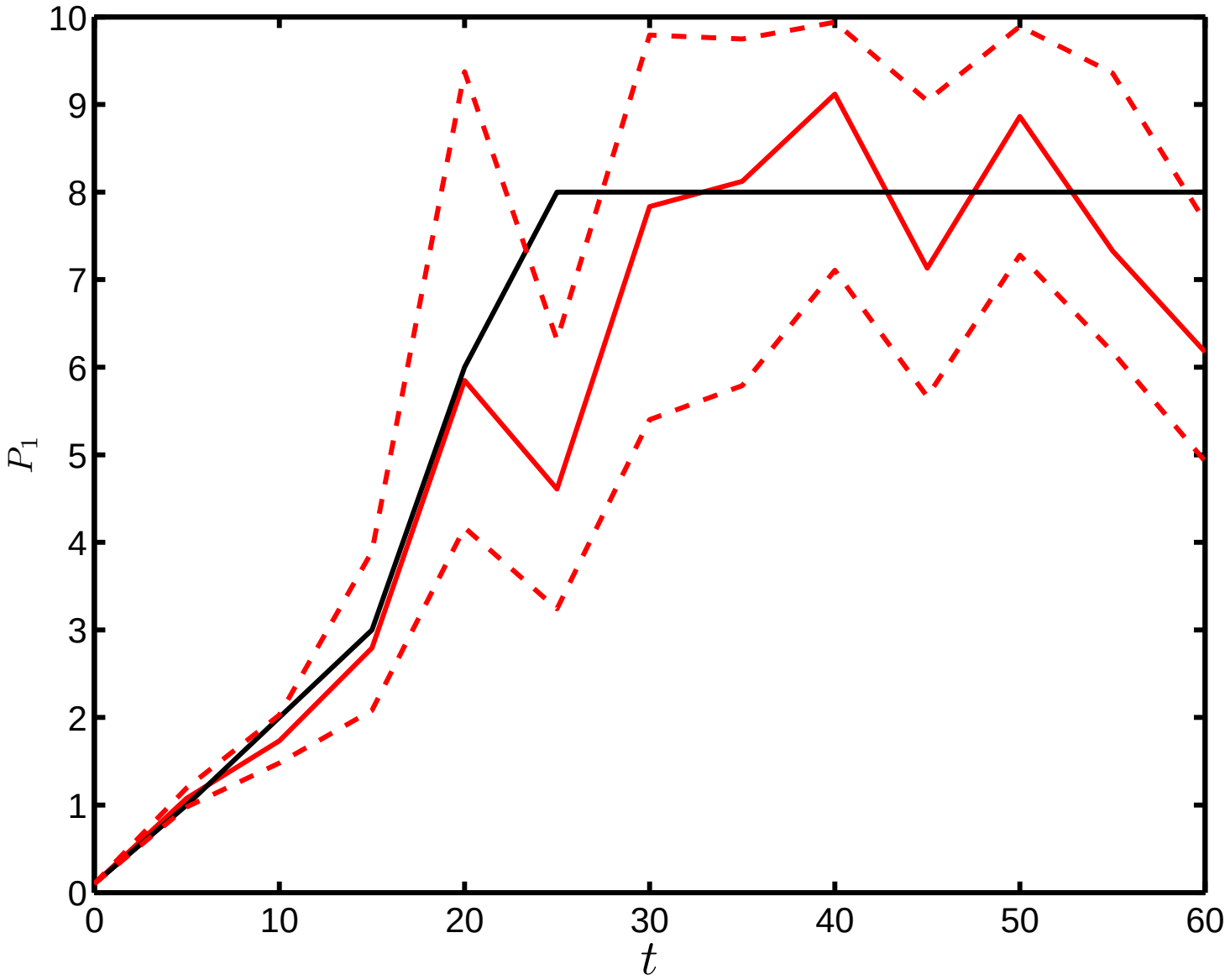
Protein degradation and state transition rates of the same order.

Example - inference





- Fixed P_2, can we infer P_1?
- Metropolis Hastings algorithm, 20000 Burn-in samples
- Log-normal likelihood


10
9
8
7
6
5
4
3
2
1
0
P_1
0
10
20
30
40
50
60
t

Conclusions

- Derived a deterministic MM-type function for the specific Biological model
- Verified the inference of activator protein in simple example

Future work

- More realistic examples
- More sophisticated sampling schemes - multi-modal posterior
- Incorporate ChIP binding data