STOCHASTIC ANALYSIS AND INFERENCE FOR OSCILLATING REACTION NETWORKS USING THE LINEAR NOISE APPROXIMATION

ARAUJO, Guilherme D. **MAIA, Leonardo P.**

Abstract

In systems biology, complex biochemical systems can be modelled by a Reaction Network. It is a mathematical framework that aims at mechanistically represent essential reaction dynamics of a system. Reaction networks can be translated to dynamical systems for analysis. We focus on systems of genetic regulation, involving low molecular counts in an intracellular environment. These systems demand a stochastic dynamical treatment that is made as a markovian jump process. We approximate the nonlinear resulting equations by a systematic expansion called linear noise approximation (LNA). Finally, we use bayesian inference methods to uncover unknown parameters of models using measurements data. We study this analysis and inference methodology in the context of oscillatory systems using a stochastic representation of the Goodwin oscillator model. Circadian clocks are an example of oscillating genetic systems.

1. Introduction

The goal of this presentation is to clearly expose our methodology and connect it with our research aims. We then consider as example a simple, non-oscillating, genetic negative nonlinear feedback model, with a protein repressing its own mRNA production by a Hill activation function. The Goodwin oscillator can be obtained by replicating the second set of reactions, adding a third species and making the regulation indirect. This introduced delay is what makes the system oscillate.

2. Reaction Networks

A Reaction Network is defined by a set of species, a set of reaction complexes, and a set of reactions [1]; it is represented as:

$$\sum_{i=1}^{N} s_{ir} Y_i \xrightarrow{k_r} \sum_{i=1}^{N} s'_{ir} Y_i, \quad r = 1, 2...R$$

Below is the reaction network for the negative autoregulation model with a Hill function transcription rate:

$$\emptyset \xrightarrow{F(P)} mRNA, \quad mRNA \xrightarrow{g_1} \emptyset$$
$$mRNA \xrightarrow{k_2} mRNA + P, \quad P \xrightarrow{g_2} \emptyset \qquad F(P) = \frac{k_1}{1 + KP^n}$$

We can translate Reaction Networks to a deterministic kinetic framework:

$$\frac{dX_i}{dt} = \sum_{r=1}^{R} S_{ir} f_r^{(0)}(X_1, ..., X_N) \Rightarrow \begin{aligned} \frac{dX_1}{dt} &= F(X_2) - g_1 X_1 \\ \frac{dX_2}{dt} &= k_2 X_1 - g_2 X_2 \end{aligned}$$

Or to a stochastic framework. It considers a markov process modelled by a master equation (Kolmogorov forward equation). This equation usually is unsolvable for nonlinear propensities. It can be exactly simulated (Fig. 1(a)) or approximated by analytical expansions (LNA).

$$\frac{\partial P(\mathbf{n}, t)}{\partial t} = \sum_{r=1}^{R} \Big(f_r(\mathbf{n} - \mathbf{S}_r) P(\mathbf{n} - \mathbf{S}_r, t) - f_r(\mathbf{n}) P(\mathbf{n}, t) \Big)$$

3. Linear Noise Approximation

LNA Is the first order approximation of a systematic expansion of the master equation in terms of the system's size, based on van Kampen's ansatz [2]. It connects the stochastic and deterministic frameworks and results in a linear Fokker-Planck form with dynamical gaussian solutions. Intrinsic noise analysis can be performed (Fig. 1(b)).

$$n_i = \Omega X_i(t) + \sqrt{\Omega}\eta_i$$

$$\frac{\partial \Pi(\eta, t)}{\partial t} = \sum_{r=1}^{R} \Big(-\sum_{i=1}^{N} S_{ir} \frac{\partial}{\partial \eta_i} \sum_{j=1}^{N} \eta_j \frac{\partial f_r^{(0)}(X)}{\partial X_j} + \frac{1}{2} \sum_{i,j=1}^{N} S_{ir} S_{jr} f_r^{(0)}(X) \frac{\partial}{\partial \eta_i} \frac{\partial}{\partial \eta_j} \Big) \Pi(\eta, t)$$

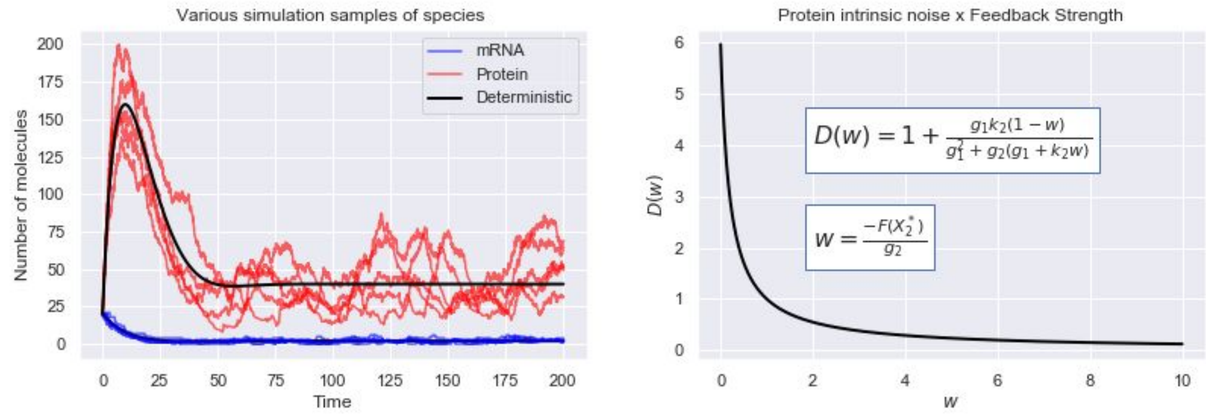

Fig. 1 - (a) Deterministic solution and some stochastic simulation samples. (b) Analysis of noise profile in terms of feedback strength using LNA, at equilibrium.

4. Inference Model

Parameter inference puts a measurement model over the reaction network and uses Bayesian inference to calculate a posterior distribution for unknown parameters [3]. In Fig. 2 we show the estimation of the Hill exponent using the LNA as data likelihood. We use simulated independent measurements and a Metropolis MCMC algorithm for sampling the posterior.

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta} \Rightarrow P(\theta|n_2) \propto \prod_i LNA(n_2^{(i)}|\theta) Prior(\theta)$$

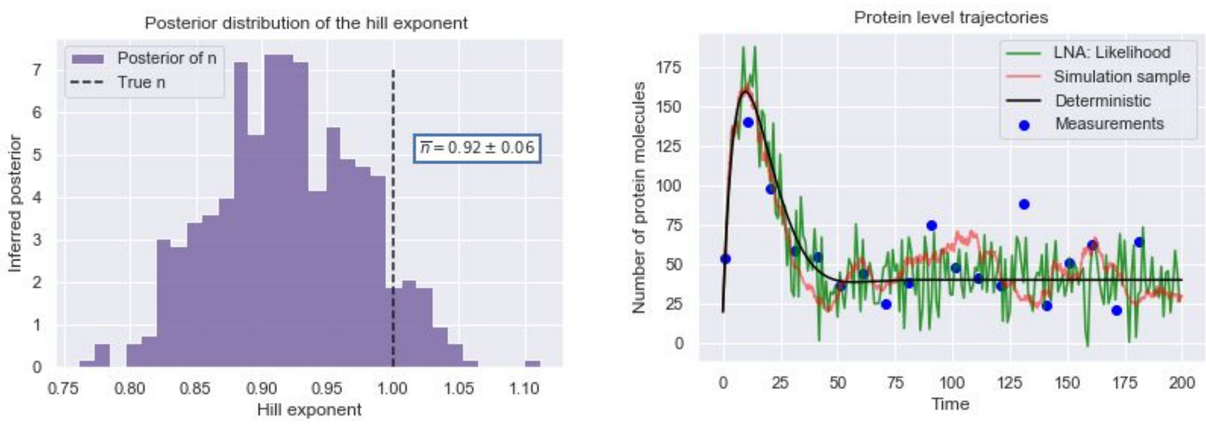

Fig. 2 - (a) Bayesian estimation of the Hill exponent (binding cooperativity). (b) Protein trajectories showing simulated measurements.

5. Conclusion

Both the LNA and the inference model change behavior near bifurcations and in oscillations. Accuracy and computational cost are quality metrics for modelling frameworks. Examples of variations we study include Resetting LNA methods and extended Kalman filters.

6. References

[1] SCHNOERR, D.; SANGUINETTI, G.; GRIMA, R. Approximation and inference methods for stochastic biochemical kinetics: a tutorial review. **Journal of Physics A**, v. 50, n. 9, p. 093001-1-093001-60, 2017.

[2] van KAMPEN, N. G. **Stochastic processes in physics and chemistry**. Amsterdam: Elsevier. 1992. v. 1.

[3] WILKINSON, D. J. Stochastic modelling for systems biology. CRC press, 2011.

CNPq

UNIVERSIDADE DE SÃO PAULO
Instituto de Física de São Carlos