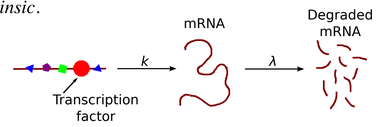
Analytical solutions of separable Master Equations and applications to gene regulatory networks

Martin Hemberg[1], Justine Dattani[2] and Mauricio Barahona[2]

[1] Department of Ophthalmology, Children's Hospital Boston, Harvard Medical School, Boston, MA; [2] Department of Mathematics, Imperial College London, UK

Motivation

- Most cellular processes are subject to significant expression noise, even between genetically identical cells in a homogeneous environment.
- The precise sources of variability, how the noise is harnessed, and the possible functional roles for noise remain poorly understood.
- Noise is fundamentally important in *gene regulation*.
- One source of stochasticity in gene transcription is the small number of molecules involved. In this sense, it is *intrinsic*.
- Additional variability is caused by the intracellular environment and the uncertainty of the biochemical processes involved, leading to cell-to-cell variability. In this sense, it is *extrinsic*.



Analytical solution of separable Master Equations

The most fundamental way to represent gene transcription is with the Master Equation (ME), but it is difficult to solve. Most researchers simulate the ME or consider only linear cases.

Here we consider systems of the form:

$$\varnothing \xrightarrow{k(t)} n(t) \xrightarrow{\lambda(t)} \varnothing$$

$$\frac{\partial P_n}{\partial t} = k(t)P_{n-1} + (n+1)\lambda(t)P_{n+1} - k(t)P_n - n\lambda(t)P_n$$

Definition 1 (Separable ME) *A ME of the form above is* ***separable*** *if the reaction rates $k(t)$ and $\lambda(t)$ do not depend on n.*

Theorem 2 (Solution of the separable ME) *The solution of a separable ME with initial conditions $P_{n_0}(t_0) = 1$ is given by*

$$P(n,t|n_0,t_0) = e^{-\kappa(t)} \sum_{p=0}^{\min\{n_0,n\}} \binom{n_0}{p} \frac{e^{-p\int_{t_0}^{t}\lambda(\tau)\,\mathrm{d}\tau}}{(n-p)!} \left(1 - e^{-\int_{t_0}^{t}\lambda(\tau)\,\mathrm{d}\tau}\right)^{n_0-p} \kappa(t)^{n-p}$$

where $\kappa(t) := \int_{t_0}^{t} k(\tau) e^{-\int_{\tau}^{t}\lambda(\tau')\,\mathrm{d}\tau'}\,\mathrm{d}\tau$

This theorem generalizes previous results for linear ME's, and for the case $n_0 = 0$.

Theorem 3 (Networks of separable reactions) *Consider a network $\mathcal{N}$ of uni-molecular reactions, as above. If $\mathcal{N}$ does not contain directed cycles, and if the reaction rates for each molecular species s_j are independent of s_j, then we can solve the ME for the network sequentially using Theorem 2.*

Acknowledgements We thank Anders Ståhlberg for helpful discussions. JD is funded through a Doctoral Grant of the Engineering and Physical Sciences Council (EPSRC) of the UK.

References

[1] O G Berg. A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of theoretical biology*, 71(4):587–603, April 1978.

[2] Julie Lyng Forman and Michael Sørensen. The Pearson Diffusions: A Class of Statistically Tractable Diffusion Processes. *Scandinavian Journal of Statistics*, 35(3):438–465, September 2008.

[3] Christian Gourieroux and Joann Jasiak. Multivariate Jacobi process with application to smooth transitions. *Journal of Econometrics*, 131(1-2):475–505, March 2006.

[4] VM Kenkre, EW Montroll, and MF Schlesinger. Generalized master equations for continuous-time random walks. *Journal of Statistical Physics*, 9(1):45–50, 1973.

[5] Anders Ståhlberg, Martin Bengtsson, Martin Hemberg, and Henrik Semb. Quantitative transcription factor analysis of undifferentiated single human embryonic stem cells. *Clinical chemistry*, 55(12):2162–70, December 2009.

Coupled stochastic dynamics and mixture models

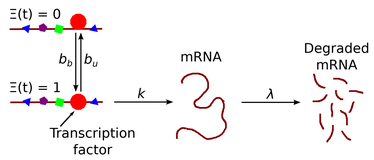
- We need to include the extra variability surrounding the process of transcription. Specifically, we model the stochastic processes at the **promoter level**.
- Treating the reaction rates as Markovian is a gross oversimplification.
- In particular, Markovian models provide poor fits to data.
- Doubly stochastic processes *mixture models* take this complexity into account.

Definition 4 (Mixture models) *A ME where the rate $k(t)$ is governed by a stochastic process with density $\pi(k,t)$ is referred to as a* ***creation mixture model*** *and if $\lambda(t)$ is governed by a stochastic process with density $q(\lambda,t)$ it is referred to as a* ***degradation mixture model****.*

The (Markovian) Berg model

The seminal gene transcription model by Berg [1] can be represented as a Markovian mixture model:

- The promoter toggles via a telegraph process between a bound (B) and an unbound (U) state with constant rates k_b and k_u.
- The mRNA creation rate is $k(t) = k\Xi(t)$, where Ξ is a (random) indicator variable that takes the value 1 when the promoter is in state B, and 0 otherwise:



One can instead describe the reaction rates as stochastic variables governed by a wide class of solvable stochastic processes known as *Pearson diffusions* [2]:

$$dK(t) = -a(K(t)-b)dt + \sqrt{2a(c_0K(t)^2 + c_1K(t) + c_2)}dW(t),$$

We take this approach for the two models below.

The Poisson-Jacobi (PJ) model

We model the binding and unbinding rates as CIR processes (a particular case of Pearson diffusion):

$$dK_b(t) = -a(K_b(t)-b_b)dt + c\sqrt{K_b(t)}dW_b(t)$$
$$dK_u(t) = -a(K_u(t)-b_u)dt + c\sqrt{K_u(t)}dW_u(t).$$

The stochastic process governing the promoter occupancy, $\Xi(t) = K_b(t)/(K_b(t)+K_u(t))$, is a Jacobi process [3] and the creation rate in the ME is $k(t) = k\Xi(t)$.

The stationary distribution of the mRNAs is a **Poisson-Beta (PB) mixture distribution**:

$$PB(n;k,\lambda,\alpha,\beta) = \frac{(\alpha)_n k^n}{(\alpha+\beta)_n n!\lambda^n}\Phi(\alpha,\alpha+\beta;-k/\lambda),$$

where $(x)_n$ is the Pochammer symbol and $\Phi(a,c;z)$ is the confluent hypergeometric function.

The Poisson-Inverse-Gamma (PIG) model

- Model the degradation rate $\Lambda(t)$ as governed by a CIR process.
- The stationary distribution of the mRNAs is a **Poisson-Inverse-Gamma (PIG) mixture distribution**:

$$PIG(n;\alpha_d,\beta_d) = \frac{2\alpha_d^{\frac{\beta_d+n}{2}}}{\Gamma(n+1)\Gamma(\alpha_d)}K_{\beta_d-n}(2\sqrt{\alpha_d}),$$

where K_a is the Bessel function of the third kind.

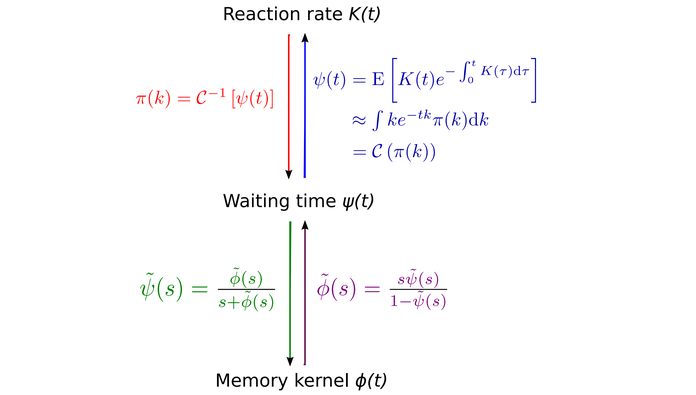
Stochastic Reaction Rates, Non-Exponential Waiting Times and Memory Kernels

Stochastic reaction rates can be interpreted as other types of non-linearities: non-exponential waiting times and memory-kernels. Kenkre *et al* [4] showed that the description of the ME using non-exponential waiting times or memory-kernels are equivalent. Using a memory kernel, the ME can be written as:

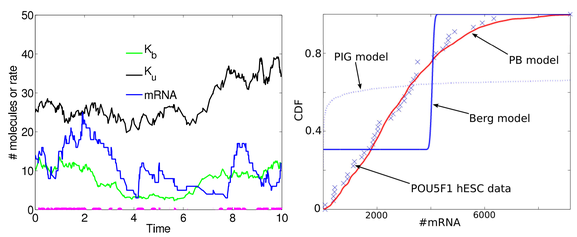
$$\dot{P}_n(t) = \int_0^t \left[\phi_{n-1}(t-\tau)P_{n-1} + (n+1)\lambda P_{n+1} - \phi_n(t-\tau)P_n - n\lambda P_n\right]\,\mathrm{d}\tau.$$

Assuming stationarity of the mixing process, $\pi(k,t)$, we can extend the equivalence.

Below, $\mathcal{C}$ is the Carson-Laplace transform.



Data fitting and temporal characteristics



(*Left*) Simulation of the PJ model with $a = 0.02$, $b_b = 10$, $b_u = 25$, $c = 1$, $k = 50$, $\lambda = 1$. The magenta marks at the bottom show when the promoter was active. (*Right*) Fits of the Berg, PJ and PIG data from for the POU5F1 (OCT4) gene from single-cell human embryonic stem cells [5]. The PJ model is clearly superior in fitting the data, as shown by the Akaike Information Criterion.

Summary

- We obtain full analytical solutions for a large class of MEs, including networks of arbitrary size that fulfill certain broad conditions.
- Applying the method to *mixture models* provides a general class of solutions that:
  - fit experimental data exceptionally well
  - are biologically interpretable
  - are analytically tractable.
- We demonstrate: stochastic reaction rates ↔ non-exponential waiting times ↔ memory kernels.