Semi-Analytical Methods for RNA burst models

Gennady Gorin[1,2], Lior Pachter[2]

[1]Division of Chemistry and Chemical Engineering, Caltech [2]Division of Biology and Biological Engineering, Caltech

Abstract: Multimodal single-cell genomics data present an appealing substrate for improved physical modeling of biological systems. We present a model for bursty, multi-stage mRNA production, introduce and characterize a power series-based method for the computation of likelihoods, and motivate applications to statistical inference.

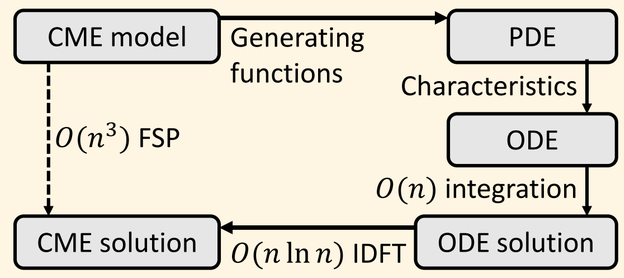
Background

Genomic data provide statistical challenges due to stochasticity of mRNA production and experimental sampling. To maximize their biological interpretability, we are developing Chemical Master Equation-based models and solvers for the Bayesian inference of underlying biophysical parameters. With an eye toward extrapolation, we explore burst models with a distinct pre-RNA population[1,2].

Mathematical background

Given the model:

where $B \sim Geom(b)$, we can compute the steady-state $P(\text{pre-mRNA, mRNA})$ through the generating function $\phi = k\int_0^\infty F(U;b)ds$.



U is in the form of $Ae^{-\beta s} + Be^{-\gamma s}$ for $\beta \neq \gamma$ and $e^{-\gamma s}(A + Bs)$ for $\beta = \gamma$.

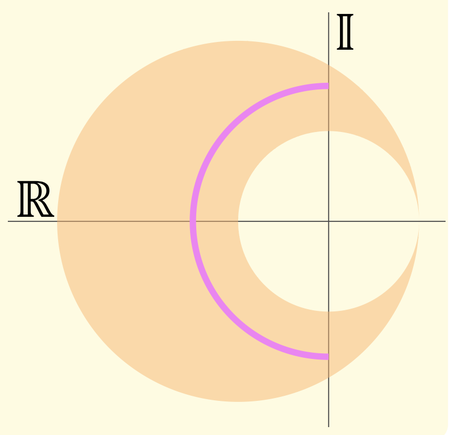
Thus, U represents downstream dynamics. For a given U, F is solely a function of the burst distribution, e.g. $\frac{bU}{1-bU}$.

By integrating $F(U;b)$, we can compute the steady-state distribution[3].

ODE decomposition

We can decompose $F(U(s;u,v;\beta,\gamma);b)$ into disjoint power series approximations $\mathrm{F}(s \in S_k) \approx \sum_i \Omega_{k,i}U^i$ integrable through the evaluation of special functions, then compute $\phi \approx \hat{\phi} = \sum_k \sum_i \Omega_{k,i}\int_{S_k} U^i ds$.

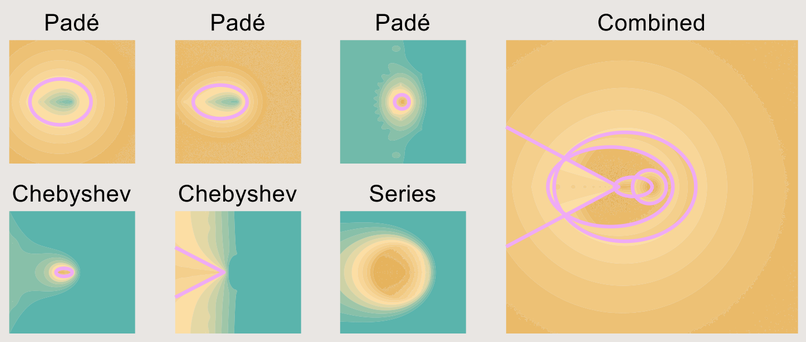
The functional form of U guarantees that $Re(U) < 0$; therefore, we decompose based on a **threshold value** of $|U|$ within the **common domain of convergence**.

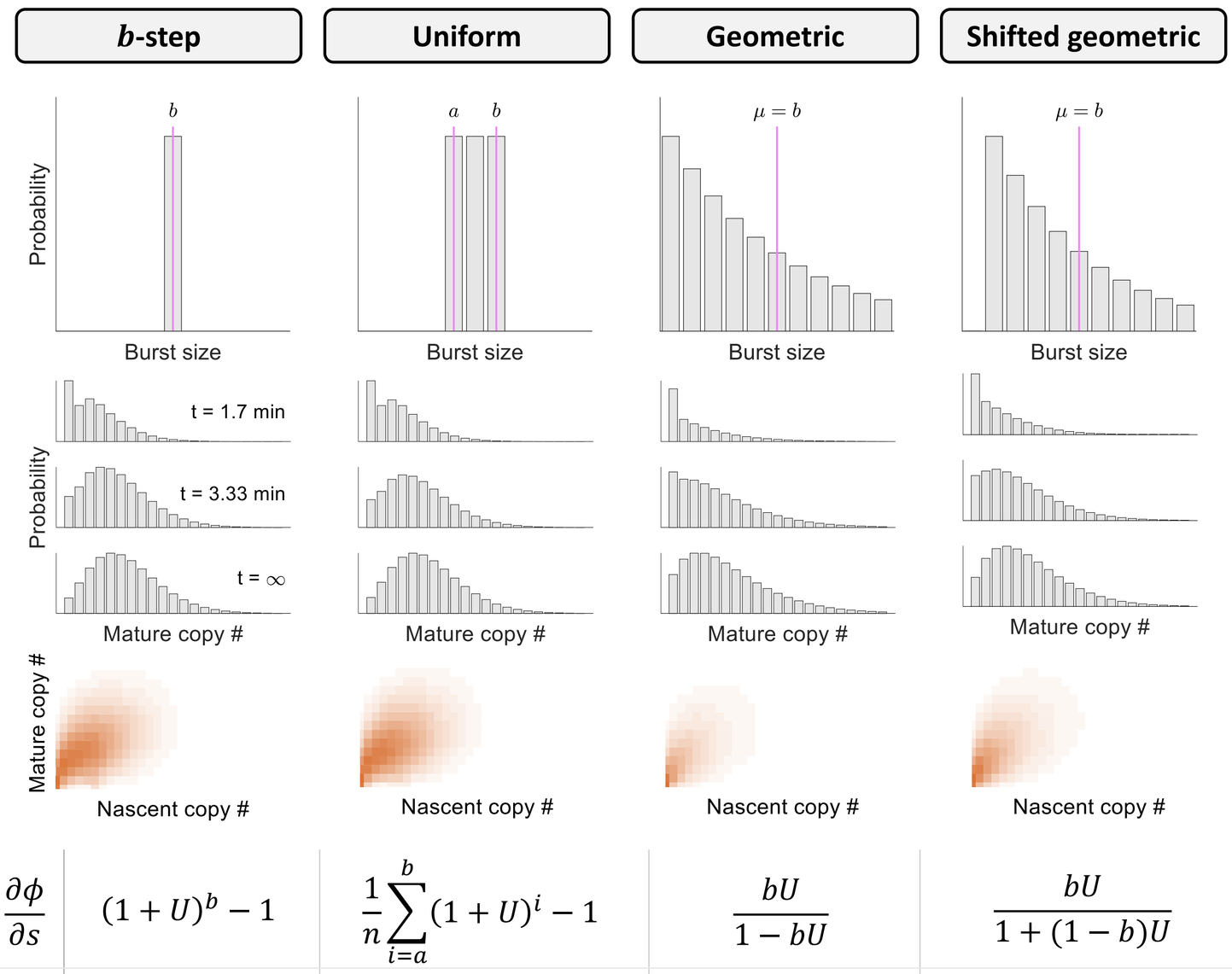


Special function solutions

$$\int U^i ds \sim \begin{cases} {}_2F_1(-i,\mu i,\mu i+1,z) & \gamma \neq \beta \\ \Gamma(1+i,\alpha i) & \gamma = \beta \end{cases}$$

Fast, general-purpose algorithms for evaluation with $\alpha, \mu, z \in \mathbb{C}$ are scarce. We are designing dedicated algorithms to extract and combine **high-performing regions** from approximations, while avoiding **divergent regions**.

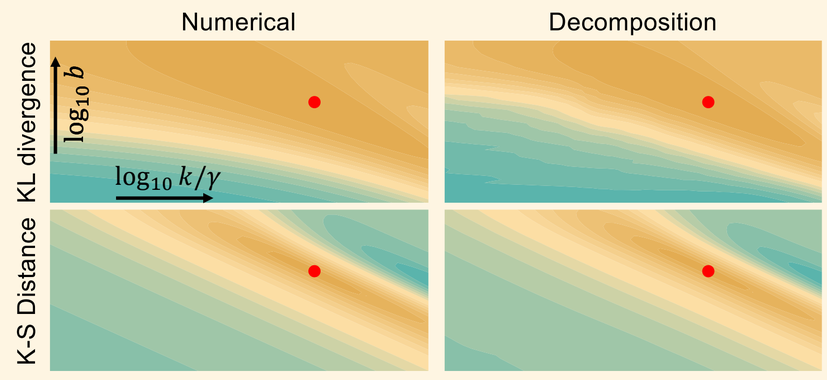




	***b*-step**	**Uniform**	**Geometric**	**Shifted geometric**
$\frac{\partial\phi}{\partial s}$	$(1+U)^b - 1$	$\frac{1}{n}\sum_{i=a}^{b}(1+U)^i - 1$	$\frac{bU}{1-bU}$	$\frac{bU}{1+(1-b)U}$
Ω_j	$\binom{b}{j}$	$\frac{1}{n}\left[\binom{b+1}{j+1} - \binom{a}{j+1}\right]$	$b^j\sum_{i=j}^{N}\frac{1}{2^{i+1}}\binom{i}{j} - \frac{1}{b^j}$	$\beta(b-1)^j\sum_{i=j}^{N}\frac{1}{2^{i+1}}\binom{i}{j} - \frac{1}{\beta(b-1)^j}$
j	$1,\ldots,b$	$1,\ldots,b$	$1,\ldots,N$ $0,-1,\ldots,-N$	$1,\ldots,N$ $0,-1,\ldots,-N$
U	$\mathbb{C}$	$\mathbb{C}$	$\lvert U\rvert < \frac{1+\sqrt{3}}{2b}$ $\lvert U\rvert > \frac{1+\sqrt{3}}{2b}$	$\lvert U\rvert < \frac{\sqrt{3}}{b-1}$ $\lvert U\rvert > \frac{\sqrt{3}}{b-1}$

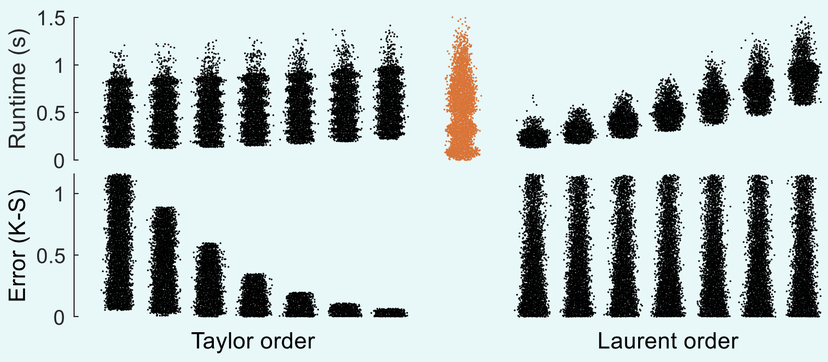
Parameter estimation

Decomposition reproduces the likelihood and Kolmogorov-Smirnov landscapes produced *via* numerical integration, motivating **parameter** estimation by searching for **optimal basins**[4].



Method performance

The **decomposition method** provides excellent control of runtime compared to **numerical integration**. Taylor approximations improve precision over a broad parameter domain.



Next steps

This approach suggests routes to *non-perturbative* bursty joint RNA-protein distribution models, which have been challenging to model[5]. The availability of entire distributions enables explicit modeling and identification of sampling processes, improving the physical interpretability of single-cell genomics data[6].

References

1. Xu et al. *Phys Rev Lett* **2016**, *117*, 128101.
2. La Manno et al. *Nature* **2018**, *560*, 494-498.
3. Singh and Bokes. *Biophys J* **2012**, *103*, 1087-1096.
4. Munsky et al. *PNAS* **2018**, *115*, 7533-7538.
5. Veerman et al. *J Math Biol* **2018**, *77*, 261-312.
6. Grün et al. *Nat Methods* **2014**, *11*, 637-640.