# Contents

# VAE-CME

Xinyi Zhou

2023.6.23

## 1 Birth Death Model

Consider a simple non-Markovian system where molecules are produced at a rate $\rho$ and are removed from the system (degraded) after a fxed time delay $\tau$:

$$\varnothing \xrightarrow{\rho} N, N \xRightarrow{\tau} \varnothing \tag{1}$$

The training set is the distribution from $1 \times 10^4$ samples using the SSA.In the experiment, we assume $\rho = 20$, $\tau = 10$ and truncation $N = 271$.

Both encoder and decoder are multilayer perceptron with one hidden layer. The objective function is chosen as the sum of mean-squared-error and KL divergence. For the training we used the standard adaptive moment estimation algorithm (ADAM). The weight of mean-squared-error needs to be increased, and the learning rate needs to be decreased from approximately 0.25 to 0.01 during the training process.

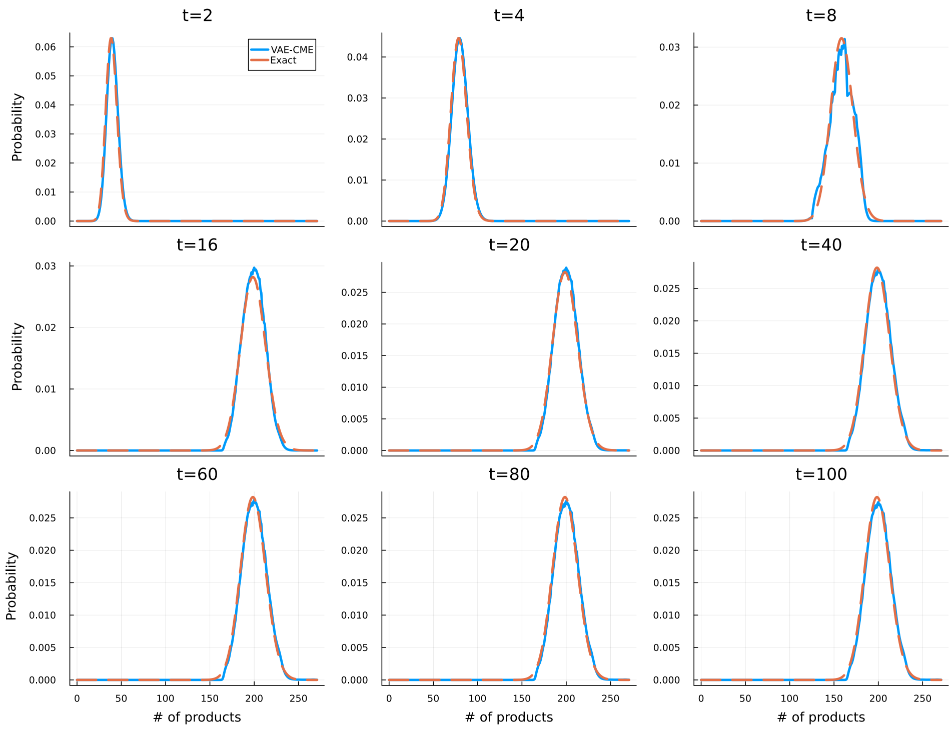


Figure 1: Birth Death Model Fitting

## 2 On Off Model

## 3 Bursty Model

We consider Bursty Model, which is the same as Birth Death Model, except that the binding of RNAPs to the promoter occurs in bursts whose size $i$ is distributed according to the geometric distribution $b^i/(1+b)^{i+1}$; this can be described by the reaction scheme:

$$\begin{aligned} &\varnothing \xrightarrow{\frac{\alpha\beta^i}{(1+\beta)^{i+1}}} iN, i = 1, 2, 3, \ldots \\ &N \overset{\tau}{\Rightarrow} \varnothing \end{aligned} \tag{2}$$

To achieve the best fitting performance, the analytical solution of the Bursty model's probability distribution (See SI in [1]) is used as the training set. In the experiment, we assume $\alpha = 0.0282$, $\beta = 3.46$, $\tau = 120$ and truncation $N = 64$.

The same as before, we choose the sum of mean-squared-error and KL divergenc as the objective function, ADAM as the optimizer. And the weight of mean-squared-error needs to be increased, the learning rate needs to be decreased

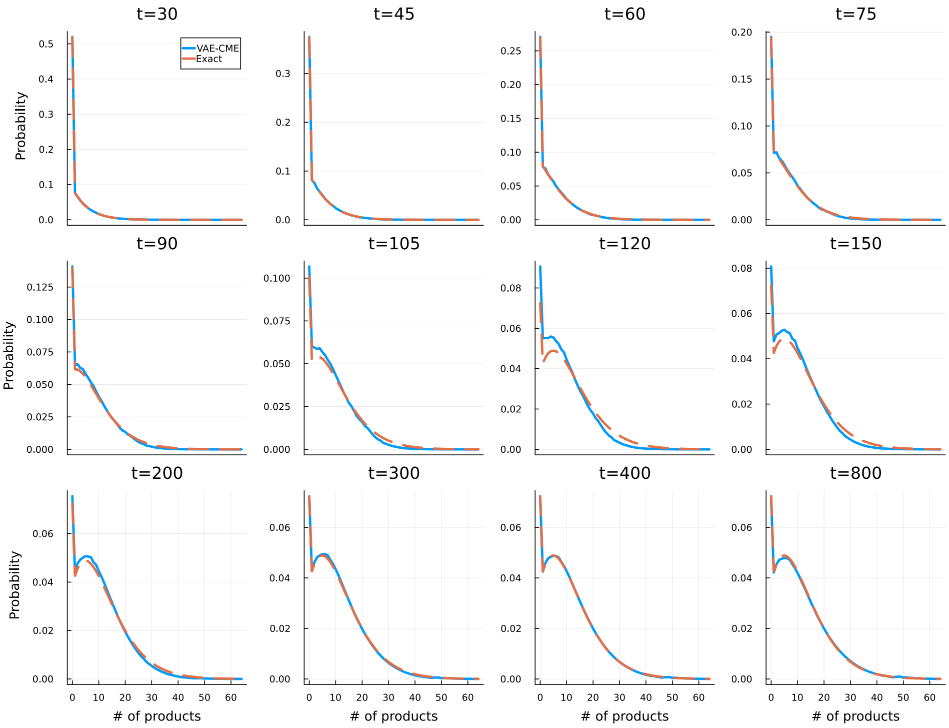


Figure 2: Bursty Model Fitting

## 3.1 Control Mean of time delay $\tau$

## 3.2 Control Variance of time delay $\tau$

# 4 Oscillation Model

## 4.1 Reducing sample size

# 5 Exact solution for variable time delay $\tau$

# References

[1] Qingchao Jiang, Xiaoming Fu, Shifu Yan, Runlai Li, Wenli Du, Zhixing Cao, Feng Qian, and Ramon Grima. Neural network aided approximation and parameter inference of non-markovian models of gene expression. Nature communications, 12(1):2618, 2021.