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 ρ and are removed from the system (degraded) after a fxed time delay τ:

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

The training set is the distribution from 1×10^4 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. The fitting performance after training is shown in Fig. 1.
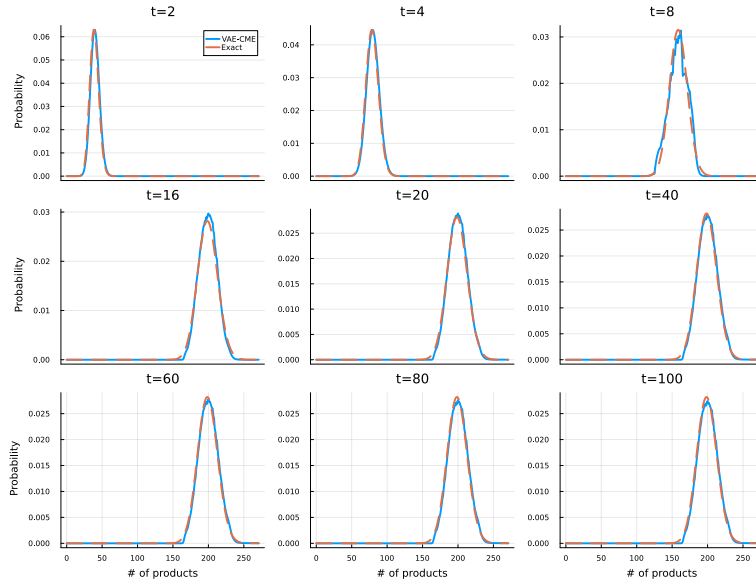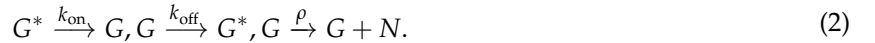

Figure 1: Birth Death Model Fitting

2 On Off Model

We also consider On Off Model wherein the promoter switches between an active and inactive state, RNAP binding occurs only in the active state, which is followed by delayed degradation modelling the RNAP movement along the gene and its detachment; this can be described by the reaction scheme:

$$G^* \xrightarrow{k_{\mathrm{on}}} G, G \xrightarrow{k_{\mathrm{off}}} G^*, G \xrightarrow{\rho} G + N. \tag{2}$$

The same as Birth Death Model, The training set is the distribution from 1×10^4 samples using the SSA.In the experiment, we assume $k_{\mathrm{on}} = 1$, $k_{\mathrm{off}} = 1$, $\rho = 20$, and truncation $N = 45$.

The same objective function, optimizer, and adjustments to weights and learning rate are used for training as in the Birth Death Model. The fitting performance after training is shown in Fig. 2.
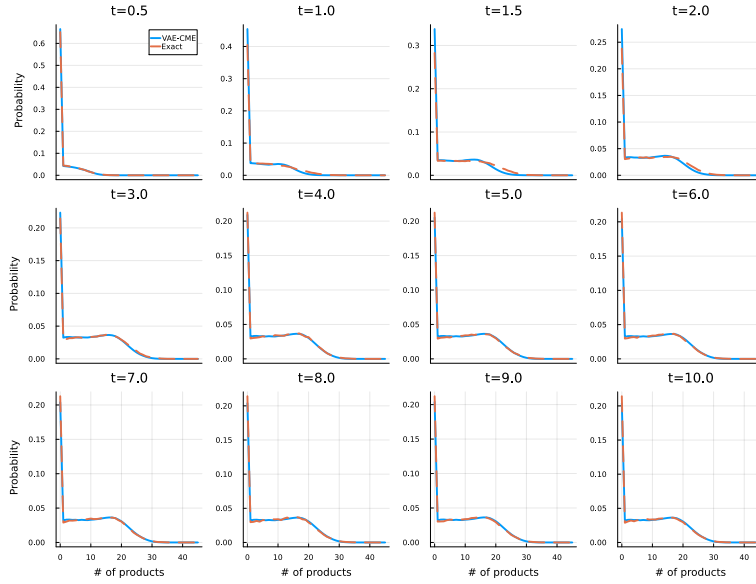
Figure 2: On Off Model fitting

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:

$$\varnothing \xrightarrow{\frac{\alpha\beta^i}{(1+\beta)^{i+1}}} iN, i = 1, 2, 3, ... \tag{3}$$
$$N \xRightarrow{\tau} \varnothing$$

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. The weight of mean-squared-error needs to be increased, and the learning rate needs to be decreased. The fitting performance after training is shown in Fig. 3
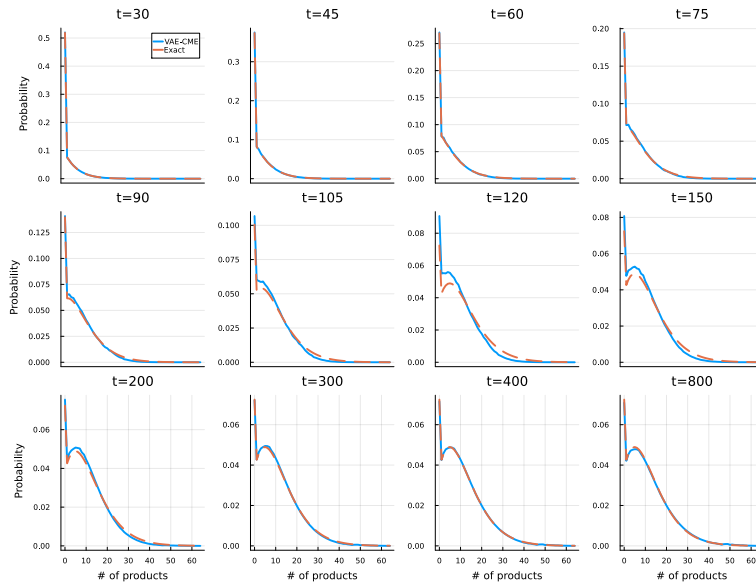

Figure 3: Bursty Model Fitting

3.1 Control Mean of time delay τ

3.2 Control Variance of time delay τ

4 Oscillation Model

Here we consider Oscillation Model, which is a simple genetic negative feedback loop whereby (i) a protein X is transcribed by a promoter, (ii) subsequently after a fixed time delay τ, X turns (via some set of unspeci"ed biochemical processes) into a protein Y and (iii) finally Y binds the promoter and reduces the rate of transcription of X. This can be described by the reaction scheme:

$$\varnothing \xrightarrow{J_1(Y)} X, X \overset{\tau}{\Rightarrow} Y, Y \xrightarrow{J_2(Y)} \varnothing. \tag{4}$$

The function $J_1(Y)$ and $J_2(Y)$ is defined as follows:

$$
\begin{aligned}
J_1(Y) &= k_1 S \frac{K_d^p}{K_d^p + Y^p}, \\
J_2(Y) &= k_2 E_T \frac{Y}{K_m + Y}.
\end{aligned}
\tag{5}
$$

In this example, we assume $k_1 = k_2 = S = E_T = K_d = K_m = 1, p = 2$ for simplicity. Unlike Models considered earlier, the delay master equation corresponding to this model has no known analytical solution. The training set is the distribution from 1×10^4 samples using the SSA. Moreover, we assume $\tau = 10$ and truncation $N = 26$.

We only use the simulated trajectories of mature protein Y to train the VAE. In other words, the data of protein X is not required during training. However, an additional regularization term which is the first derivative of the probability distribution of protein X obtained from the chemical master equation (CME), needs to be added to the objective function. Therefore, the final objective function of this experiment is the sum of mean-squared-error of protein Y, KL divergence and regularization term mentioned before. The fitting performance of protein Y after training is shown in Fig. 4 and the predicting performance of protein X is shown in Fig. 5.
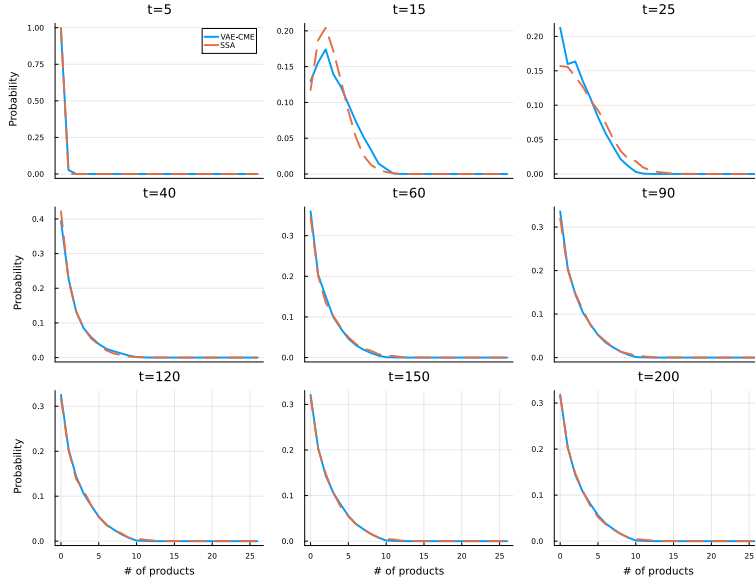

Figure 4: Oscillation Y fitting

4.1 Reducing sample size

Due to the poor predicting performance of MLP [1] on protein X in the Oscillation model, we studied the predicting performance of VAE under small sample conditions. We trained MLP and VAE using probability distributions obtained from simulating SSA for 100, 300, 1000, 3000, and 10000 iterations as training sets, respectively. Since the distribution from 1×10^4 samples using the SSA closely approximate the true probability
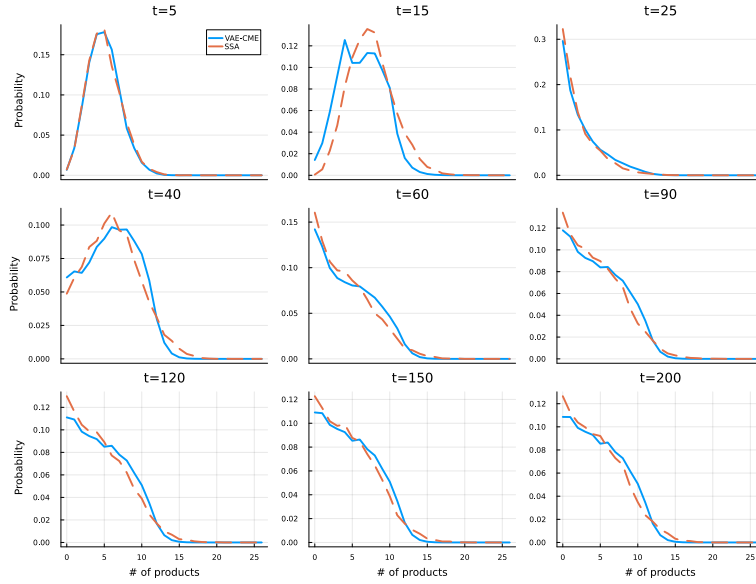
Figure 5: Oscillation X predicting

distribution, we consider it as exact distributions. The measure of the accuracy is the mean squared error between the network(MLP or VAE) and exact distributions.

We will trained MLP and VAE using three sets of probability distributions obtained from SSA for 100, 300, 1000, 3000 and 10000 iterations respectively. The average mean squared error is calculated and compared(See 6).

Note that the VAE obtained from training with 1×10^2 samples produces a distribution that is as precise as that from 1.6×10^3 samples trained by MLP. In this case the training time of the VAE is also just about 1/3 of the MLP for every epoch.
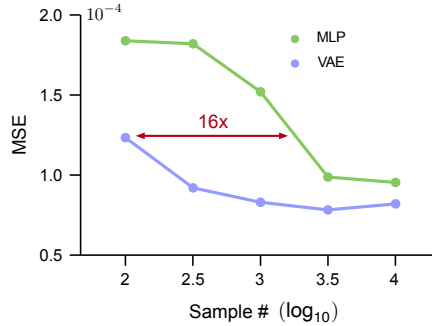

Figure 6: MSE VAE and MLP

5   Exact solution for variable time delay τ

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.