



Research article

Stationary moments, distribution conjugation and phenotypic regions in stochastic gene transcription

Jiajun Zhang^{1,2} and Tianshou Zhou^{1,2,*}

¹ Guangdong Province Key Laboratory of Computational Science, Guangzhou 510275, P.R. China

² School of Mathematics, Sun Yat-Sen University, Guangzhou 510275, P.R. China

* **Correspondence:** Email: mcszhtsh@mail.sysu.edu.cn.

Abstract: Transcription is a pivotal step in gene expression yet a complex biochemical process. It occurs often in a bursty manner, leading to cell-to-cell variability important for the cells surviving in complex environments. Quantitative experiments of such stochastic transcription call for model analysis in a systematic rather than case-by-case fashion. Here we analyze two general yet biologically-reasonable classes of stochastic transcription models: the first class called promoter models considers that promoter structure is general, and the second class called queuing models considers that waiting-time distributions are general. For the former, we show that there are conjugate relationships between models with different transcription exits. This property well reveals the mechanic principle of stochastic transcription in complex cases of transcription factor regulation. For the latter, we establish an integral equation for the mRNA moment-generating function, through which stationary mRNA moments of any orders can be analytically derived. Finally, we analyze parametric regions for robust unimodality and bimodality. The overall analysis not only lays a foundation for quantitative analysis of stochastic transcription but also reveals how the upstream promoter kinetics impact the downstream expression dynamics.

Keywords: stochastic transcription; chemical master equation; unimodal distribution; bimodal distribution; cellular phenotype

1. Introduction

Transcription is not only a central process in life but also a pivotal step in gene expression since it determines when and how the genetic information stored in DNA is transcribed into messenger RNA (mRNA) molecules that are further translated into proteins that virtually participate in all cell functions. Transcription occurs often in a bursting fashion [1–5], leading to substantial variations in mRNA and further protein abundances. Single molecule measurement technologies have provided

evidence for transcriptional bursting, i.e., production of mRNAs in bursts [6–9]. This noise has been identified as an important source of cell-to-cell variability [3, 10–16]. Owing to its importance, stochastic transcription has received extensive attention in recent years. However, quantitative experiments of stochastic transcription call for a modeling effect that addresses transcriptional mechanisms in a systematic rather than case-by-case fashion.

Transcription is the most complex step in gene expression, involving ordered assembly of pre-initiation complexes, recruitment of various polymerases, chromatin remodeling, binding and unbinding of transcription factors (TFs) to DNA regulatory sites, etc. [17–19]. These factors or processes either result in non-exponential distributions of ON and OFF times (i.e., the times that the gene dwells at active and inactive states) [11, 16, 20], or create complex promoter structures [18, 21–25]. The former case occurs mainly because, e.g., some of the chromatin states are transcriptionally permissive, and together make up the ON-phase of the gene and allow for (repeated) multistep assembly of the preinitiation complex and promoter escape. This will motivate us to introduce a so-called queuing model of stochastic transcription, where the distributions of the ON and OFF times may be arbitrary. The latter case may occur in prokaryotic and eukaryotic cells. For example, the PRM promoter of phage lambda in *E. coli* is regulated by two different TFs binding to two sets of three operators that can be brought together by looping out the intervening DNA and as a result, the number of regulatory states of the PRM promoter may be up to 128 [26]. In contrast, eukaryotic promoters may be more complex since they would involve, e.g., nucleosomes competing with or being removed by TFs [27], and epigenetic regulations via histone modifications [19, 28, 29]. In a word, complex promoters with more than two activity states are not the exception but the rule as combinatorial control of gene regulation by multiple TFs is widespread [9, 25, 30, 31]. This will motivate us to introduce another model of stochastic transcription (called the promoter model), where the promoter may contain many active and inactive states and there are random transitions among them. For both generalized models, an unsolved issue is how the upstream promoter dynamics affects the downstream expression dynamics, and in particular, how mRNA distribution is derived.

On the other hand, a different mRNA distribution describes a different pattern of cell-to-cell variability at the transcription level, or represents a different cellular phenotype. As a matter of fact, bimodality of gene expression, as a mechanism driving to phenotypic diversity, can enhance the survival of cells in a fluctuating environment [32]. Unimodality also has similar biological function, e.g., the unimodal expression of Tat (a vital gene in HIV-1 viral) can drive phenotypic diversity of clonal populations through inducing phenotypic bifurcation [33, 34]. In simplified cases or under strong assumptions, previous studies showed that distributions of gene product (mRNA or protein) may follow simple distributions such as Poisson distribution [4, 35], Gamma distribution [36, 37] and Beta distribution [9, 38, 39]. These simple distributions provide explicit and intuitive relationships between cellular phenotypes and expression patterns. In general, however, the mRNA numbers follows more complex distributions expressed by confluent hypergeometric functions, as shown in previous studies [24, 39]. This raises another question: How are unimodal or bimodal mRNA expressions robustly shaped? From the viewpoint of synthetic biology, studying this question is significant since it can help biologists design robust experiments and interpret their experimental results. We point out: (1) with the emergence of single cell measurement techniques, there have been increasingly rich data on distributions of gene systems [40, 41]. These distributions, which are in general stationary, are still biologically reasonable if a set of reactions proceeds much faster than

environmental changes; (2) the stationary distribution and the corresponding statistics are important quantities characterizing the steady-state behavior of many biochemical systems. Therefore, we will focus on the steady-state distribution and moments of mRNA.

This article is organized as follows. In section 2, we analyze a promoter model with arbitrarily complex promoter structure at the stochastic transcription level. By analysis, we find that there are conjugate relationships between mRNA distributions, which well reveal the essential principle of stochastic transcription in complex cases of TF regulation. In section 3, we analyze a queuing model of stochastic transcription, where waiting-time distributions are general (e.g., non-exponential). For this model, we derive an analytical formula for calculating raw moments, and give a strategy for reconstructing the mRNA distribution from raw moments. In section 4, by mathematical analysis combined with numerical simulation, we give robust regions for unimodal and bimodal expressions in two specific models of stochastic transcription. Finally, we conclude this paper and give some discussions.

2. Analysis of a gene model with arbitrary promoter structure

2.1. Model description and stationary mRNA distribution

As pointed out in the introduction, a gene promoter (i.e., a regulatory sequence) may contain many activity states (referring to Figure 1, which shows a biologically feasible example [21]), and there are random transitions between these states. We assume that the promoter has N activity states in total. For analysis simplicity, we further assume that there is transcription only at ON state whereas there is no transcription at OFF state, and there are transitions among these states with constant transition rates. To incorporate effects of TF regulation, we assume that the TFs act directly on transition rates [21], so that these rates may change in some finite ranges. In addition, we assume that all the produced mRNAs degrade in a linear manner. Here we do not consider protein levels since mRNA levels can well explain stationary protein levels in many situations [42].

Denote by m the number of mRNA molecules, and let $P(m; t)$ represent the distribution that mRNA has m molecules at the k^{th} state of promoter and denote by $\mathbf{P} = (P_1, \dots, P_N)^T$ the column vector consisting of these factorial distributions. Denote by λ_{jj} the transition rate from state- i to state- j ($\lambda_{jj} = 0$ means that no transition occurs). Let the $N \times N$ matrix $\mathbf{W} = (\lambda_{jj})$ describe transitions between promoter states (called transition matrix), the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\mu_1, \dots, \mu_N)$ describe exits of transcription (called transcription matrix) with μ_i representing the transcription rate of mRNA in state- i ($\mu_i = 0$ means that no transcription takes place), and the diagonal matrix $\delta = \text{diag}(\delta_1, \dots, \delta_N)$ describe the degradation of mRNA. The chemical master equation describing mRNA dynamics takes the form

$$\frac{\partial}{\partial t} \mathbf{P}(m; t) = \mathbf{W} \mathbf{P}(m; t) + \mathbf{\Lambda} (\mathbf{E}^{-1} - \mathbf{I}) [\mathbf{P}(m; t)] + \delta (\mathbf{E} - \mathbf{I}) [m \mathbf{P}(m; t)], \quad (2.1)$$

where \mathbf{E} and \mathbf{E}^{-1} are step operators and \mathbf{I} is the identity operator. Clearly, the first term in Eq (2.1) describes promoter dynamics with transition matrix \mathbf{W} being actually an M-matrix (i.e., the sum of every column or row elements is equal to zero), the second term describes the exits of transcription with transcription matrix $\mathbf{\Lambda}$, and the third term describes the degradation dynamics of mRNA with degradation matrix δ (throughout this paper, we will assume that all the degradation rates are the same. Without loss of generality, we further assume that the common degradation rate is 1. This is equivalent

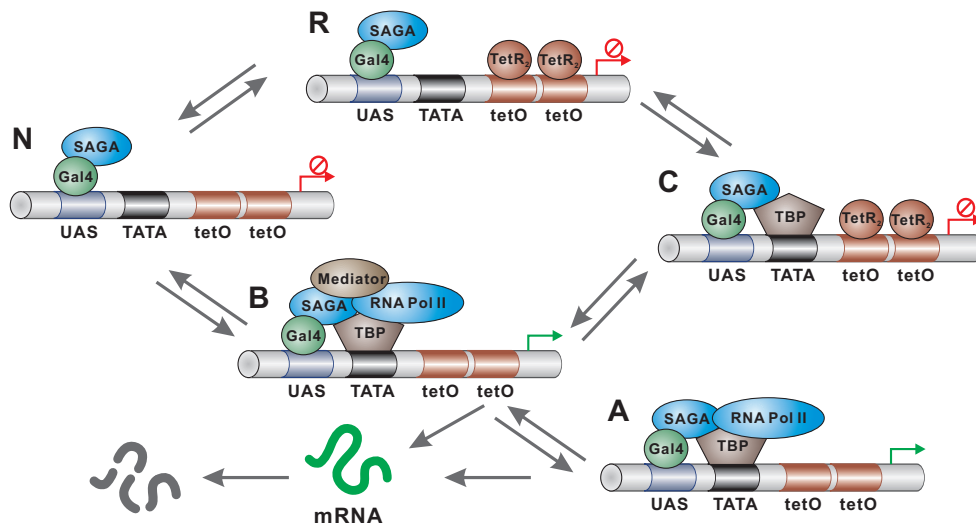


Figure 1. A biological example for gene model with multiple activity states [21], where the five promoter states indicated by C, R, N, B, and A correspond to binding of TetR, TBP, and RNA Pol II to promoter DNA in the following combinations: both TBP and TetR bound (“Combined” state C), only TetR bound (“Repressed” state R), only TBP bound (“Bound” state B), neither TBP nor TetR bound (“Neutral” state N), and TBP and RNAPol II bound (“Active” state A). States C, R, and N (TBP not bound) represent inactive promoter states, while states B and A (TBP bound) represent active promoter states. The transition rate from B to N (and C to R) depends on TBP-DNA stability, while the transition rate from state N to R depends on the inducer concentration.

to the normalization of all other parameters by this common rate). We point out that model (2.1) includes many previously-studied gene models as its particular case.

In order to solve (2.1), we simply introduce the convergent moment (CM) approach that we developed previously [43]. CMs of a distribution in the one-dimension $b_k = \sum_{n \geq k} \binom{n}{k} P(n)$, $k = 0, 1, 2, \dots$, all cases are defined as

$$b_k = \sum_{n \geq k} \binom{n}{k} P(n), k = 0, 1, 2, \dots, \quad (2.2)$$

with $b_0 = 1$ that corresponds to the conservative condition of probability. It is easy to verify that the CMs defined in such a way are actually coefficients in the Taylor series of the probability-generating function, denoted by $g(z)$, corresponding to the probability $p(n)$, that is, $b_k = (1/k!) \partial^k g(z) / \partial z^k |_{z=1}$. Importantly, if all the CMs are finite and given, we can have the following formula for reconstructing the distribution

$$P(n) = \sum_{k \geq n} (-1)^{k-n} \binom{n}{k} b_k, n = 0, 1, 2, \dots, \quad (2.3)$$

where symbol $\binom{n}{k}$ represents the common binomial coefficients.

Now, we return to (2.1). Denote by b_k the k^{th} order CM of the total distribution $P = \sum_{i=1}^N P_i$ and by $b_k^{(i)}$ the k^{th} order CM of the factorial distribution P_i . Next, we can derive the following CM equation

from (2.1)

$$\frac{d\mathbf{b}_k}{dt} = \mathbf{W}\mathbf{b}_k + \mathbf{\Lambda}\mathbf{b}_{k-1} - k\mathbf{b}_k, k = 1, 2, \dots, \quad (2.4)$$

where \mathbf{W} and $\mathbf{\Lambda}$ have been rescaled by δ , and $\mathbf{b}_k = (b_k^{(1)}, \dots, b_k^{(N)})^T$ is a column vector consisting of factorial CMs. See Appendix A for the derivation of (2.4). Finding steady-state mRNA distribution is common interest. According to (2.3), the key is to find stationary CMs b_k . For this, we have

Proposition 1 Steady-state CMs b_k are given by

$$b_k = \frac{1}{k! \prod_{i=1}^k \prod_{j=1}^{N-1} (i + \alpha_j)} \mathbf{u}_N \prod_{i=k}^1 [(\mathbf{iI} - \mathbf{W})^* \mathbf{\Lambda}] \mathbf{b}_0, k = 1, 2, \dots, \quad (2.5)$$

where $\mathbf{u}_N = (1, 1, \dots, 1)$ is an N -dimensional row vector, $0, -\alpha_1, -\alpha_2, \dots, -\alpha_{N-1}$ are nonzero characteristic values of M-matrix \mathbf{W} ; $(\mathbf{iI} - \mathbf{W})^*$ is the adjacency matrix of matrix $(\mathbf{iI} - \mathbf{W})$; and $\mathbf{b}_0 = (b_0^{(1)}, \dots, b_0^{(N)})^T$ is given by

$$b_0^{(j)} = \prod_{i=1}^{N-1} \frac{\beta_i^{(j)}}{\alpha_i}, 1 \leq j \leq N, \quad (2.6)$$

in which $-\beta_1^{(j)}, -\beta_2^{(j)}, \dots, -\beta_{N-1}^{(j)}$ are the characteristic values of \mathbf{W}_j , the minor matrix of \mathbf{W} after crossing out the j^{th} row and j^{th} column of the entry λ_{jj} of \mathbf{W} .

Proof Here, we only provide main steps of the proof. See Appendix B for more details.

First, note that at steady state, (2.4) becomes the following iterative form

$$(\mathbf{kI} - \mathbf{W})\mathbf{b}_k = \mathbf{\Lambda}\mathbf{b}_{k-1}, k = 1, 2, \dots. \quad (2.7)$$

Using the fact that \mathbf{W} is an M-matrix, i.e., the sum of every row elements is equal to zero, it is not difficult to verify (2.5).

Second, note that the Laplace's formula for M-matrix \mathbf{W} gives

$$\mathbf{W} (\det(\mathbf{W}_1), \dots, \det(\mathbf{W}_N))^T = \det(\mathbf{W})\mathbf{I} = \mathbf{0}, \quad (2.8)$$

the null space of \mathbf{W} is one-dimensional, and (2.7) holds for $k = 0$ (implying $\mathbf{W}\mathbf{b}_0 = \mathbf{0}$). Therefore, we can set the formal expression of \mathbf{b}_0 as $\mathbf{b}_0 = c(\det(\mathbf{W}_1), \dots, \det(\mathbf{W}_N))^T$, where c is a constant, given by

$$c = \left[\sum_{k=1}^N \det(\mathbf{W}_k) \right]^{-1} = (-1)^{N-1} + \left[\sum_{k=1}^N \prod_{i=1}^{N-1} \beta_i^k \right]^{-1}. \quad (2.9)$$

Third, if we denote by

$$f_{\mathbf{W}}(x) = \det(x\mathbf{I}_N - \mathbf{W}) = x(x + \alpha_1) \cdots (x + \alpha_{N-1}), \quad (2.10)$$

the characteristic polynomial of matrix \mathbf{W} , it is not difficult to show

$$\frac{df_{\mathbf{W}}(x)}{dx} \Big|_{x=0} = \text{tr}((-\mathbf{W})^*) = (-1)^{N-1} \sum_{k=1}^N \det(\mathbf{W}_k) = \sum_{k=1}^N \prod_{i=1}^{N-1} \beta_i^k = \prod_{i=1}^{N-1} \alpha_i \quad (2.11)$$

where "tr" represents the matrix trace.

Finally, using (2.10) and considering the components of vector \mathbf{b} , we have

$$b_0^{(k)} = c \cdot \det(\mathbf{W}_k) = (-1)^{N-1} \left(\prod_{i=1}^{N-1} \alpha_i \right)^{-1} (-1)^{N-1} \prod_{i=1}^{N-1} \beta_i^k = \prod_{i=1}^{N-1} \frac{\beta_i^k}{\alpha_i}, 1 \leq k \leq N, \quad (2.12)$$

thus proving Proposition 1.

After having obtained all the stationary CMs, we can reconstruct the stationary mRNA distribution using (2.3). From (2.7) and using the fact that \mathbf{W} is an M-matrix, we can show

$$b_k = \mathbf{u}_N \mathbf{b}_k = \frac{1}{k} \sum_{i=1}^N \mu_i b_{k-1}^{(i)} = \frac{1}{k} \mathbf{u}_N \mathbf{\Lambda} \mathbf{b}_{k-1}, k = 0, 1, 2, \dots \quad (2.13)$$

Thus, it is particularly interesting that if all the states of the promoter are ON with the same transcription rate μ , i.e., if $\mathbf{\Lambda} = \mu \mathbf{I}_N$, the mRNA number obeys to the following Poisson distribution

$$P(m) = e^{-\mu} \frac{\mu^m}{m!}, m = 0, 1, 2, \dots \quad (2.14)$$

This indicates that the mRNA distribution is completely determined by a single parameter (transcription rate), independent of promoter structure. This property has the following implication: if transcription rates are not exactly the same, the corresponding mRNA distribution is definitely not a simple Poissonian one. In this case, deriving its analytical expression is challenging.

2.2. Conjugation

In order to address the above challenge, here we introduce the conception of conjugation. By conjugation we mean that for two gene models with the same transition pattern between promoter states (i.e., both have the same transition rate matrix \mathbf{W}), denoted by Model A and Model B, if their transcription matrices take the specific forms: $\mathbf{\Lambda}_A = \text{diag}((\mu + \varepsilon) \mathbf{I}_k, \mu \mathbf{I}_{N-K})$ for Model A and $\mathbf{\Lambda}_B = \text{diag}(\mu \mathbf{I}_K(\mu + \varepsilon) \mathbf{I}_{N-K})$ for Model B, where ε is a real number and may be equal to $-\mu$, and K is a positive integer of less than N , next we call Model A conjugate to Model B or vice versa. Such conjugation is existent in natural or synthetic biological systems, e.g., for the example illustrated in Figure 1(A) (denoted by Model A), if states C, R, and N are active while states B and A are inactive, the resulting model becomes a conjugate model of Model A. The conjugation is useful in revealing relationships between complex mRNA distributions. In the following, for a system of gene expression with promoter states, let $\mathbf{P}(m; t) = (P_1(m; t), \dots, P_N(m; t))^T$ and $\mathbf{G}(z; t) = (G_1(z; t), \dots, G_N(z; t))^T$ be the column vectors consisting of factorial probabilities $P_i(m; t)$ with $1 \leq i \leq N$ and the corresponding factorial probability-generating functions $G_i(z; t)$ with $G_i(z; t) = \sum_{m \geq 0} P_i(m; t) z^m$ and $1 \leq i \leq N$, respectively; let $P(m; t) = \sum_{i=1}^N P_i(m; t)$ and $G(z; t) = \sum_{i=1}^N G_i(z; t)$ represent the total probability and the corresponding total probability-generating function, respectively.

2.3. Distribution properties

The above CM approach can in principle give the exact mRNA distributions in gene models with arbitrary promoter structure, which however may be complex in form [25]. For example, for the

common ON-OFF model of stochastic transcription, we know that the mRNA distribution is expressed by a complex confluent hypergeometric function [24]. To see that these reconstructed distributions belong to which types, we check the corresponding Fano factors defined as the ratio of variance over mean. This is because if the Fano factor is less than 1, the distribution is sub-Poissonian, if it is equal to 1, the distribution is Poissonian, and if it is greater than 1, the distribution is sup-Poissonian [45]. For clarity, we consider a specific case, that is, the promoter has one ON state and $N - 1$ OFF states, which altogether form a loop (i.e., the so-called DNA loop [46]). In this case, we can show

$$Fano - Factor = 1 + \langle m \rangle \left[\left(\prod_{i=1}^{N-1} \frac{\alpha_i}{\beta_i^{(N)}} \right) \left(\prod_{i=1}^{N-1} \frac{\beta_i^{(N)} + 1}{\alpha_i + 1} \right) - 1 \right], \quad (2.15)$$

where $\langle m \rangle$ represents the mean number of mRNA molecules and can be expressed as $\langle m \rangle = \mu \langle \tau_{on} \rangle / (\langle \tau_{on} \rangle + \langle \tau_{off} \rangle)$ with μ being the transcription rate, and $\langle \tau_{on} \rangle$ and $\langle \tau_{off} \rangle$ representing the mean times dwelling at ON and OFF states, respectively. We can prove that $\beta_i^{(N)} \leq \alpha_i$ for any $i = 1, 2, \dots, N - 1$. Therefore, the term in the big bracket of (2.15) is larger than or equal to zero, implying that for stochastic gene transcription, the mRNA distribution is either sup-Poissonian or Poissonian, independent of promoter topology. Regarding conjugation between two systems of gene expression with the same number of promoter states, we have the following theorem.

Theorem 1 For two conjugate stochastic transcription Models A and B, if we let two pairs $P_A(m; \mu)$, $G_A(z)$, and $P_B(m; \mu)$, $G_B(z)$ represent the distributions and probability-generating functions

$$G_A(s) = e^{\varepsilon s} G_B(-s), P_A(m; \mu) = e^{\varepsilon s} \sum_{k=0}^m \frac{e^{m-k}}{(m-k)!} P_B(m; -\mu), \quad (2.16)$$

where $s = z - 1$.

Proof Consider the following two models of stochastic transcription

Model A:

$$\frac{\partial}{\partial t} \mathbf{P}(m; t) = \mathbf{W}\mathbf{P}(m; t) + \mathbf{\Lambda}_A(\mathbf{E}^{-1} - \mathbf{I})[\mathbf{P}(m; t)] + \delta(\mathbf{E} - \mathbf{I})[m\mathbf{P}(m; t)].$$

Model B:

$$\frac{\partial}{\partial t} \mathbf{Q}(m; t) = \mathbf{W}\mathbf{Q}(m; t) + \mathbf{\Lambda}_B(\mathbf{E}^{-1} - \mathbf{I})[\mathbf{Q}(m; t)] + \delta(\mathbf{E} - \mathbf{I})[m\mathbf{Q}(m; t)].$$

Introduce probability-generating functions: $\mathbf{G}_A(z; t) = \sum_{m \geq 0} \mathbf{P}(m; t) z^m$, $\mathbf{G}_B(z; t) = \sum_{m \geq 0} \mathbf{Q}(m; t) z^m$.

Then

$$\begin{aligned} \frac{\partial}{\partial t} \mathbf{G}_A(z; t) &= \sum_{m \geq 0} z^m \frac{\partial}{\partial t} \mathbf{P}(m; t) \\ &= \sum_{m \geq 0} z^m (\mathbf{W}\mathbf{P}(m; t) + \mathbf{\Lambda}_A(\mathbf{E}^{-1} - \mathbf{I})\mathbf{P}(m; t) + \delta(\mathbf{E} - \mathbf{I})(m\mathbf{P}(m; t))) \\ &= \mathbf{W}\mathbf{G}_A(z; t) + \mathbf{\Lambda}_A \left(\sum_{m \geq 0} z^m \mathbf{P}(m-1; t) - \sum_{m \geq 0} z^m \mathbf{P}(m; t) \right) + \delta \left(\sum_{m \geq 0} (m+1) z^m \mathbf{P}(m+1; t) - \sum_{m \geq 0} m z^m \mathbf{P}(m; t) \right) \\ &= \mathbf{W}\mathbf{G}_A(z; t) + (z-1)\mathbf{\Lambda}_A \mathbf{G}_A(z; t) - (z-1)\delta \frac{\partial}{\partial z} \mathbf{G}_A(z; t). \end{aligned}$$

Assume $\delta = \delta \mathbf{I}$ with δ being a positive constant, and set $s = z - 1$. Next, we have

$$\frac{\partial}{\partial \tau} \mathbf{G}_A = \tilde{\mathbf{W}} \mathbf{G}_A + s \tilde{\Lambda}_A \mathbf{G}_A - s \frac{\partial}{\partial s} \mathbf{G}_A, \quad (2.17)$$

where $\tau = \delta t$, $\tilde{\mathbf{W}}$ and $\tilde{\Lambda}_A$ have been scaled by δ . Similarly, we can obtain

$$\frac{\partial}{\partial \tau} \mathbf{G}_B = \tilde{\mathbf{W}} \mathbf{G}_B + s \tilde{\Lambda}_B \mathbf{G}_B - s \frac{\partial}{\partial s} \mathbf{G}_B, \quad (2.18)$$

where $\tilde{\Lambda}_B$ has been scaled by δ . Multiplying $e^{-\varepsilon s}$ on both sides of (2.17) and noting that s is unrelated to τ yield

$$\begin{aligned} \frac{\partial}{\partial \tau} (e^{-\varepsilon s} \mathbf{G}_A) &= \tilde{\mathbf{W}} (e^{-\varepsilon s} \mathbf{G}_A) + s \tilde{\Lambda}_A (e^{-\varepsilon s} \mathbf{G}_A) - s e^{-\varepsilon s} \frac{\partial}{\partial s} \mathbf{G}_A \\ &= \tilde{\mathbf{W}} (e^{-\varepsilon s} \mathbf{G}_A) + s \tilde{\Lambda}_A (e^{-\varepsilon s} \mathbf{G}_A) - s \frac{\partial}{\partial s} (e^{-\varepsilon s} \mathbf{G}_A) - s \varepsilon (e^{-\varepsilon s} \mathbf{G}_A). \end{aligned}$$

In other words, under the transformation $\mathbf{G}_A = e^{\varepsilon s} \mathbf{F}_A$, (2.17) is transformed into

$$\frac{\partial}{\partial \tau} \mathbf{F}_A = \tilde{\mathbf{W}} \mathbf{F}_A + s (\tilde{\Lambda}_A - \varepsilon \mathbf{I}) \mathbf{F}_A - s \frac{\partial}{\partial s} \mathbf{F}_A, \quad (2.19)$$

which is similar to (2.18) in form, where $\tilde{\Lambda}_A - \varepsilon \mathbf{I} = -\tilde{\Lambda}_B$ is a constant matrix. Thus, according to the transformation $\mathbf{G}_A = e^{\mu s} \mathbf{F}_A$ and noting that $-s \partial / \partial s = -(-s) \partial / \partial (-s)$, we know that the first equality in (2.16) holds. Furthermore, through the relationship $P(m; t) = (1/m!) (\partial^m G / \partial z^m)|_{z=0}$, where $P = \sum_{i=1}^N P_i$ and $G = \sum_{i=1}^N G_i$ are respectively the total probability distribution and the total probability-generating function, it is not difficult to verify that the second equality in (2.16) also holds. To that end, the proof of Theorem 1 is completed.

From (2.16), we see that $P_A(m; \mu)$ is actually a binomial convolution of $P_B(m; \mu)$ and a Poisson distribution. We point out that by making use of the conjugation idea, one can find analytical distributions in complex gene models, e.g., the gene model with one OFF state and two ON states is difficult to solve but the gene model with two OFF state and one ON states is easy to solve. Using the conjugation idea, we can transform the question of solving the former into that of solving the later. To show this point clearly, here we consider a more general case. Assume that for some multi-state model of stochastic transcription, the steady-state equations for probability-generating functions are

$$\begin{aligned} \sum_{k=1, \neq i}^N \lambda_{ki} G_k - \sum_{k=1, \neq i}^N \lambda_{ik} G_i - s \frac{\partial G_i}{\partial s} + \mu s G_i &= 0, i = 1, 2, \dots, N-1, \\ \sum_{k=1}^{N-1} \lambda_{kN} G_k - \sum_{k=1}^{N-1} \lambda_{Nk} G_N - s \frac{\partial G_N}{\partial s} + (\mu + \varepsilon) s G_N &= 0, \end{aligned} \quad (2.20)$$

where $s = z - 1$, and parameter ε may take any value in the interval $[-\mu, \mu]$. This equation group seems difficult to solve since the corresponding gene model contains multiple nontrivial transcription exits. However, the following equation group

$$\begin{aligned} \sum_{k=1, \neq i}^N \lambda_{ki} F_k - \sum_{k=1, \neq i}^N \lambda_{ik} F_i - s \frac{\partial F_i}{\partial s} &= 0, i = 1, 2, \dots, N-1, \\ \sum_{k=1}^{N-1} \lambda_{kN} F_k - \sum_{k=1}^{N-1} \lambda_{Nk} F_N - s \frac{\partial F_N}{\partial s} + \varepsilon s F_N &= 0 \end{aligned} \quad (2.21)$$

which corresponds to the gene model with one ON state and $N - 1$ OFF states and hence is easy to solve [24]. Moreover, using the above CM approach, we can obtain the solution for the total generating function $F_s = \sum_{i=1}^N F_i(s)$

$$F(s) = {}_{N-1}F_{N-1} \left(\beta_1^{(N)}, \dots, \beta_{N-1}^{(N)} \middle| \alpha_1, \dots, \alpha_{N-1}; \varepsilon s \right), \quad (2.22)$$

where ${}_nF_n \left(\begin{matrix} a_1, \dots, a_n \\ b_1, \dots, b_n \end{matrix} \middle| \sigma \right)$ is a confluent hypergeometric function [44]. Note that (2.20) and (2.21) are conjugate under the transformation $G_i = e^{\mu s} F_i$. Thus, we obtain the analytical expression for the total generating function $G_s = \sum_{i=1}^N G_i(s)$

$$G(s) = e^{\mu s} F(s) {}_{N-1}F_{N-1} \left(\beta_1^{(N)}, \dots, \beta_{N-1}^{(N)} \middle| \alpha_1, \dots, \alpha_{N-1}; -\varepsilon s \right). \quad (2.23)$$

Furthermore, using the relationship $P(m) = G^m(0)/m!$, we obtain analytical steady-state mRNA distribution in the underlying model of stochastic transcription, which is given by

$$P(m) = \frac{e^{-\mu}}{m!} \sum_{k=0}^m \binom{m}{k} \varepsilon^{m-k} \mu^k \prod_{i=1}^{N-1} \frac{(\beta_i^{(N)})_k}{(\alpha_i)_k} {}_{N-1}F_{N-1} \left(\begin{matrix} k + \beta_1^{(N)}, \dots, k + \beta_{N-1}^{(N)} \\ k + \alpha_1, \dots, k + \alpha_{N-1} \end{matrix} \middle| -\varepsilon \right). \quad (2.24)$$

Before concluding this section, we discuss the relationship between this hypergeometric type distribution and Poissonian distribution. In fact, if there is only one ON state without OFF state, the distribution is Poissonian. Also, if there is only one OFF state without ON state, the distribution is degenerative. Consider a promoter structure with K ON states and L OFF states. For slow promoter switch rates, one observes the characteristic distribution that results from the mixture of K Poissonian distributions and L degenerate distributions approximately. For fast promoter switch rates, the characteristic distribution is a Poissonian distributions with mean synthetic rate approximately. For moderate promoter switch rates, the characteristic distribution is very complicated hypergeometric type.

3. Analysis of a gene model with any waiting-time distributions

3.1. Calculation of raw moments

The extensively used two-state model of gene expression [37, 47–50] assumes that gene promoter has two states: one active (ON) state and one inactive (OFF) state and two transition rates between these states are constants (implying that times that the gene dwells at ON and OFF states, i.e., ON and OFF times, follow exponential distributions). However, transcription depends on the chromatin template that accumulates over time until the promoter becomes active [24]. This implies that OFF times do not follow a simple exponential distribution but may be general [51]. Similarly, ON times may also follow a non-exponential distribution [20]. Some recent experimental papers [52, 53] also showed non-exponential promoter switching times. In addition, the mRNA degradation may undergo a multistep process, implying that the distribution of mRNA's half lifetimes may be non-exponential. For such a gene model that one can find its prototypes in the real biological systems [16], there are few studies. Here, we apply standard results from the queuing theory [50, 51, 54, 55] to analyze a state-waiting model of stochastic transcription with general waiting-time distributions $f_{on}(t)$, $f_{off}(t)$, and $h(t)$,

referring to Figure 2. A similar model was previously studied in [56]. Importantly, we derive explicit formulae for calculating the moment-generating function corresponding to the mRNA distribution, which can in turn be used for quantitative analysis of mRNA noise and even for reconstruction of the mRNA distribution. We point out that the gene model considered here can be viewed as an extension of the gene model considered above. In fact, for the later, we can show [25]

$$\begin{aligned} f_{on}(t) &= \mathbf{u}_K \mathbf{W}_{10} \exp(\mathbf{W}_{11}t) \mathbf{W}_{01} \mathbf{u}_K^T, \\ f_{off}(t) &= \mathbf{u}_L \mathbf{W}_{01} \exp(\mathbf{W}_{00}t) \mathbf{W}_{10} \mathbf{u}_L^T, \end{aligned} \quad (3.1)$$

where $\mathbf{W} = \begin{pmatrix} \mathbf{W}_{00} & \mathbf{W}_{10} \\ \mathbf{W}_{01} & \mathbf{W}_{11} \end{pmatrix}$ is a block matrix, $\mathbf{u}_K = (1, 1, \dots, 1)$ and $\mathbf{u}_L = (1, 1, \dots, 1)$ are two row vectors. Here, \mathbf{W}_{11} and \mathbf{W}_{00} describe transitions among the ON states and among the OFF states respectively, whereas \mathbf{W}_{10} and \mathbf{W}_{01} describes how the ON states transition to the OFF states and vice versa respectively. Note that two functions $f_{on}(t)$ and $f_{off}(t)$ are in general the sums of exponential functions of t , implying that ON- or OFF-times follows a non-exponential distribution.

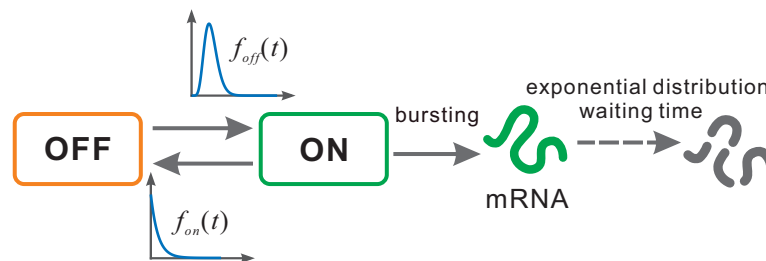


Figure 2. The schematic for a queuing model of stochastic transcription, where the times that the gene dwells at ON and OFF states are assumed to follow general distributions, and mRNA is synthesized in a bursty manner and degrades in a multistep manner.

Assume that the synthesis rate of mRNA, μ , is a constant. Denote by $P(m; t)$ the mRNA distribution at time t and by $W(z; t)$ the corresponding moment-generating function with the initial condition denoted by $W_{init}(z) = W(z; 0)$. Unlike the case of common gene models where waiting-time distributions are exponential, we in general cannot establish the master equation for $P(m; t)$ in the case of general waiting-time distributions $f_{on}(t)$, $f_{off}(t)$, and $h(t)$, since the corresponding processes are non-Markovian. Therefore, we will turn to consider moments of $P(m; t)$, and apply queuing theory to establish integral equations for moment-generating functions.

First, we establish the following proposition.

Proposition 2 Let $h(t)$ be the survival probability density function and $H(t) = \int_t^\infty h(\tau) d\tau$ be the survival function of mRNA. We have

$$W(z; t) = W_{init}(\log(1 + H(t)(e^z - 1))). \quad (3.2)$$

Proof In order to prove (3.2), we need to make assumptions: (I) there are N mRNAs produced at time $t = 0$. In general, N itself is a random variable, and every mRNA X_i ($1 \leq i \leq N$) at time t is also a stochastic variable, taking the value of either 1 (representing survival) or 0 (representing death); (II)

every X_i is independent of random variable N (this assumption seems reasonable from the viewpoint of biology); (III) all the stochastic variables X_i ($1 \leq i \leq N$) are independent of one another, each following a Bernoulli distribution with the moment-generating function given by $M(z) = 1 + H(t)(e^z - 1)$ [20, 50, 54].

Note that the total molecule number of mRNA at time t , given by $S = X_1 + \dots + X_N$, is a random variable. By the definition of moment-generating function, we know $W(z; t) = \langle e^{zS} \rangle_S$. Also note that the mean $\langle e^{zS} \rangle_S$ can be expressed as $\langle e^{zS} \rangle_S = \langle \langle e^{z(X_1 + \dots + X_N)} | N \rangle_X \rangle_N$. The above assumption (III) implies $\langle \langle e^{z(X_1 + \dots + X_N)} | N \rangle_X \rangle_N = \langle \langle e^{zX_1} | N \rangle_X \dots \langle e^{zX_N} | N \rangle_X \rangle_N$. Again by the definition of moment-generating function, we have $\langle e^{zX_i} | N \rangle_{X_i} = M(z)$ with $1 \leq i \leq N$ and the above assumption (II) yields $\langle e^{zX_i} | N \rangle_{X_i} = \langle (M(z))^N \rangle_N = \langle e^{N \log(M(z))} \rangle_N$. According to the meaning of $W_{init}(z)$, we know $W_{init}(\log(M(z))) = \langle e^{N \log(M(z))} \rangle_N$. Substituting the known expression $M(z) = 1 + S(t)(e^z - 1)$ into $W_{init}(\log(M(z)))$, we immediately obtain (3.2) and Proposition 2 is thus proved.

Next, we follow the standard derivation of the moment-generating function in queuing theory to derive an integral equation for $W_{init}(z)$ [20, 54], which is a key for calculating raw moments. For this, we give the following proposition.

Proposition 3 The initial moment function $W_{init}(z)$ satisfies the following integral equation

$$W_{init}(z) = \int_{s=0}^{\infty} \int_{t=0}^{\infty} W_{init}(\log(1 + H(s+t)(e^z - 1))) e^{\rho_t(e^z - 1)} f_{on}(t) f_{off}(s) dt ds, \quad (3.3)$$

where $\rho_t = \mu \int_0^t H(\tau) d\tau$.

Proof Let time $t = 0$ be the beginning of an OFF-state and $W(z; t)$ be the moment-generating function of mRNA molecules at time t ($0 \leq t \leq t_{on} + t_{off}$) as motioned above. The completion of one cycle (OFF- and ON state) define the boundary condition:

$$W(z; 0) = M(z; t_{on} + t_{off}) = \int_{s=0}^{\infty} \int_{t=s}^{\infty} W(z; t) f_{on}(t-s) f_{off}(s) dt ds. \quad (3.4)$$

This equation in principle determines $W_{init}(z)$ due to $W_{init}(z) = W(z; 0)$ if $W(z; t)$ is known. However, the expression of $W(z; t)$ seems difficult to derive. Here, we introduce another strategy to calculate $W_{init}(z)$, which is based on specification of moment-generating functions during an OFF duration and an ON duration.

Given the distribution of mRNA life times, $h(t)$, the probability that one mRNA molecule is still present after time Δt is a Bernoulli random variable with the moment-generating function given by $1 + H(\Delta t)(e^z - 1)$ [50, 51, 54]. Starting with the moment-generating function $W(z; 0) = W_{init}(z)$ and considering only degradation of mRNAs, the compound moment-generating function at time Δt is given by

$$W(z; \Delta t) = W_{init}(\log(1 + H(\Delta t)(e^z - 1))). \quad (3.5)$$

Therefore, at the end of an OFF-state (denoted by t_{off} the corresponding time), the moment-generating function $W(z; t_{off})$ should be equal to the integral of the initial moment-generating function $W_{init}(\log(1 + H(s+t)(e^z - 1)))$ multiplied by the OFF-time distribution $f_{off}(t)$ over the infinite interval $(0, \infty)$, that is,

$$W(z; t_{off}) = \int_{t=0}^{\infty} W_{init}(\log(1 + H(t)(e^z - 1))) f_{off}(t) dt. \quad (3.6)$$

To that end, we have given the expression for the moment-generating function during an OFF duration.

Next, we derive the moment-generating function during an ON duration. During this period, degradation of the mRNA molecules that are present at the beginning of bursts continues as described through the function $1 + H(t)(e^z - 1)$ with the moment-generating function given by, $W(z; t_{off} + \mu) = W_{init}(\log(1 + H(\mu + t_{off})(e^z - 1)))$, where μ represents a moment of the ON duration. Besides, mRNAs are also created and degraded according to a birth-death process with an exponential waiting time. This process can be described by a Poissonian distribution with the average $\rho_\mu = \mu \int_0^\mu H(\tau) d\tau$ (remark: if the mRNA synthesis rate is a constant μ and the survival probability density function of mRNA is $h(t)$, this mRNA distribution is Poissonian [67]). Correspondingly, according to the definition, we know that the moment-generating function is given by $W_{effect}(z; t_{off} + \mu) = e^{\rho_\mu(e^z - 1)}$. The combination of the above two aspects produces an effective burst size. Note that during an ON duration, the probability distribution of the mRNA number should be the convolution of the distribution of the number of those molecules that are still present from previous bursts and the effective burst-size distribution, that is,

$$W(z; t_{off} + \mu) = \int_{s=0}^{\infty} W_{init}(\log(1 + H(s + \mu)(e^z - 1))) e^{\rho_\mu(e^z - 1)} f_{off}(s) ds \quad (3.7)$$

Then, $W(z; t_{off} + t_{on})$ is given by integrating $W(z; t_{off} + \mu)$ over the infinite interval of $\mu \in (0, \infty)$, i.e.,

$$\begin{aligned} W(z; t_{off} + t_{on}) &= W(z; 0) = W_{init}(z) \\ &= \int_{\mu=0}^{\infty} \left(\int_{s=0}^{\infty} W_{init}(\log(1 + H(s + \mu)(e^z - 1))) e^{\rho_\mu(e^z - 1)} f_{off}(s) ds \right) f_{on}(\mu) d\mu \end{aligned} \quad (3.8)$$

which is just (3.3). Note that (3.8) can be rewritten as

$$W(z; 0) = \int_{s=0}^{\infty} \int_{t=s}^{\infty} W_{init}(\log(1 + H(t)(e^z - 1))) e^{\rho_\mu(e^z - 1)} f_{on}(t - s) f_{off}(s) dt ds \quad (3.9)$$

which is actually (3.4). To that end, we have actually proven Proposition 3.

In general, solving the integral equation (3.3) is difficult since two functions $f_{on}(t)$ and $f_{off}(t)$ are arbitrary. However, we can use (3.3) to calculate moments of the mRNA distribution at time $t = 0$. For example, by differentiating both sides of (3.3) at $z = 0$, it is not difficult for us to derive the expression of the first-order initial raw moment

$$\langle m \rangle_{init} = W'_{init}(0) = \frac{\int_{t=0}^{\infty} \rho_t f_{on}(t) dt}{1 - \int_{t=0}^{\infty} \int_{s=0}^{\infty} H(t + s) f_{on}(t) f_{off}(s) dt ds}. \quad (3.10)$$

According to the definition of moment-generating function, we know that $W_{init}(0) = 1$. Similarly, we can derive the expression of the second-order initial raw moment

$$\begin{aligned} \langle m^2 \rangle_{init} &= M''_{init}(0) \\ &= \frac{\int_{t=0}^{\infty} \int_{s=0}^{\infty} [\rho_t(1 + \rho_t) + \langle m \rangle_{init} H(s + t)(1 - H(s + t) + 2\rho_t)] f_{on}(t) f_{off}(s) dt ds}{1 - \int_{t=0}^{\infty} \int_{s=0}^{\infty} H^2(s + t) f_{on}(t) f_{off}(s) dt ds}. \end{aligned} \quad (3.11)$$

Other higher-order initial raw moments can also be derived similarly. Detailed expressions are omitted due to complexity.

Now, we give explicit formulae for calculating raw moments of any order. Let $F(s)$ and $G(s)$ be the cumulative functions of the duration distributions at OFF and ON states, respectively. Next, we have $F_{off}(t) = \int_0^t f_{off}(s)ds$ and $F_{on}(t) = \int_0^t f_{on}(s)ds$. According to definitions, the mean OFF and ON times are calculated according to

$$\langle \tau_{off} \rangle = \int_{s=0}^{\infty} s f_{off}(s) ds, \langle \tau_{on} \rangle = \int_{s=0}^{\infty} s f_{on}(s) ds, \quad (3.12)$$

respectively. Note that the derivatives of $W(z; t_{off} + \mu)$ at $z = 0$ give the raw moments of the mRNA distribution at time $t = t_{off} + \mu$ during an ON duration, denoted by $\langle m^k \rangle_{\mu}$, whereas the derivatives of $W(z; t_{off})$ at $z = 0$ give the raw moments of the mRNA distribution at the end time $t = t_{off}$ of an OFF duration, denoted by $\langle m^k \rangle_s$. Then, the total moments of the mRNA copy number distribution should be the sum of the two parts obtained by averaging over all times according to their probabilities, that is,

$$\langle m^k \rangle = \frac{1}{\langle \tau_{on} \rangle + \langle \tau_{off} \rangle} \left(\int_{t=0}^{\infty} \langle m^k \rangle_s (1 - F_{off}(s)) ds + \int_{t=0}^{\infty} \langle m^k \rangle_{\mu} (1 - F_{on}(\mu)) d\mu \right), \quad (3.13)$$

where $k = 1, 2, \dots$. In (3.13), we emphasize that $\langle m^k \rangle_s$ is obtained by calculating the k -order derivative of function $W_{init}(\log(1 + H(s)(e^z - 1)))$ at $z = 0$ whereas $\langle m^k \rangle_{\mu}$ by calculating the k -order derivative of function $\int_{s=0}^{\infty} W_{init}(\log(1 + H(s + \mu)(e^z - 1))) e^{\rho\mu(e^z - 1)} f_{off}(s) ds$ at $z = 0$. We point out that although we have derived general formula for calculating raw moments, e.g., (3.13), the actual calculation would be complex even if waiting times follow exponential distributions, referring to Appendix C.

3.2. Reconstructing the mRNA distribution from its raw moments

As shown above, we can calculate raw moments of any orders. In spite of this, the common interest is in finding the corresponding distribution. Here, we consider the question of how the mRNA distribution is reconstructed from the raw moments obtained above. For the above model with general waiting-time distributions, it is easy to show that CMs of the steady-state mRNA distribution b_k defined by (2.2) are linear combinations of raw moments that however have been given using the above method, e.g.,

$$b_1 = \langle m \rangle, b_2 = \frac{1}{2}(\langle m^2 \rangle - \langle m \rangle^2), b_3 = \frac{1}{6}(\langle m^3 \rangle - 3\langle m^2 \rangle \langle m \rangle + 2\langle m \rangle^3). \quad (3.14)$$

More generally, we can give explicit relationships between convergent moments and common moments. For example, central moments (denoted by μ_k) can be expressed as

$$\mu_k = (-b_1)^k + \sum_{i=0}^{k-1} \sum_{j=1}^{k-i} R(k, i, j) (j!) (b_1)^i b_j, \quad (3.15)$$

where $R(k, i, j) = (-1)^i \binom{k}{i} S(k - i, j)$ and $S(n, k)$ is the Stirling number of the second kind [65]. According to (3.15), we can in turn use central moments to express CMs (details are omitted here). In a word, all the CMs can be given if all the raw moments are given using the above method (but the related calculations would be complex. See the following content). Thus, according to (2.3), we can in principle reconstruct the mRNA distribution although waiting-time distributions may be general.

3.3. Compared with some known results: effect of non-Markovianity

As applications of the above general expressions, here we first give explicit formulae for the mRNA mean and the mRNA noise, which actually establish a relationship between the above two models through initial two moments. Here, we assume $h(t) = \delta e^{-\delta t}$ and δ represents the mean degradation rate of mRNA. Firstly, consider the mRNA mean. Calculations (see Appendix C) yield

$$\langle m \rangle = \frac{\mu}{\delta} \frac{\langle \tau_{on} \rangle}{\langle \tau_{on} \rangle + \langle \tau_{off} \rangle}. \quad (3.16)$$

This formula is well known for the common ON-OFF model [24]; [57] but here derived in a more general case (i.e., ON and OFF waiting times follow general distributions. In other words, the formula for the mRNA mean in the case of non-exponential waiting-time distributions is the same as in the case of exponential waiting-time distributions). Secondly, consider the mRNA noise. If the mRNA noise is defined as the ratio of the variance over the square of the mean, by complex calculations (see an example in Appendix C), we can derive

$$\eta_m^2 = \underbrace{\frac{1}{\langle m \rangle}}_{\text{internal noise}} + \underbrace{\frac{\langle \tau_{off} \rangle}{\langle \tau_{on} \rangle} - \frac{\langle \tau_{on} \rangle + \langle \tau_{off} \rangle}{\delta \langle \tau_{on} \rangle^2} \frac{(1 - L_{on}(\delta))(1 - L_{off}(\delta))}{1 - L_{on}(\delta)L_{off}(\delta)}}_{\text{promoter noise}}, \quad (3.17)$$

where $L_{off}(s)$ and $L_{on}(s)$ are the Laplace transforms of functions $f_{off}(t)$ and $f_{on}(t)$, respectively. In (3.18), the first term represents the factorial noise due to the birth and death of mRNA whereas the second term represents the factorial noise due to switching between two promoter states. In particular, if two transition rates between ON and OFF states are constants, implying that ON and OFF waiting-time distributions are exponential, (3.17) is reduced to the following known form [24, 58]

$$\eta_m^2 = \underbrace{\frac{1}{\langle m \rangle}}_{\text{internal noise}} + \underbrace{\frac{\langle \tau_{off} \rangle^2}{\langle \tau_{on} \rangle + \langle \tau_{off} \rangle + \langle \tau_{on} \rangle \langle \tau_{off} \rangle}}_{\text{promoter noise}}. \quad (3.18)$$

Therefore, (3.16) and (3.17) are the extensions of the mRNA mean and the mRNA noise in the common two-state model of stochastic transcription. In addition, if we consider a model of stochastic transcription where the promoter is assumed to have one ON state and multiple OFF states, which altogether form a loop, (3.16) and (3.18) still hold although the OFF times follow a non-exponential distribution (implying non-Markovianity) [24].

Next, we show that in the limit of fast switching, the mRNA distribution can become Poissonian irrespective of the distribution of mRNA life times. In fact, in the limit of fast switching, the second term on the right-hand-side of (3.17), which describes the promoter noise, approximates zero. Thus, the resulting mRNA noise can be approximated as $\eta_m^2 \approx \frac{1}{\langle m \rangle}$, implying that the mRNA number approximately follows a Poissonian distribution independent of the distribution of mRNA life times. Interestingly, the author [55] recently gave intuitive interpretations for Poisson steady-state distribution of the mRNA number.

In order to show effects of non-Markovianity on the mRNA mean and the mRNA noise, we consider that ON- and OFF-times follow Erland distributions:

$$f_{on}(t; k_{on}, \theta_{on}) = \frac{t^{k_{on}-1} e^{-t/\theta_{on}}}{\theta_{on}^{k_{on}} (k_{on} - 1)!}, f_{off}(t; k_{off}, \theta_{off}) = \frac{t^{k_{off}-1} e^{-t/\theta_{off}}}{\theta_{off}^{k_{off}} (k_{off} - 1)!}, \delta = 1, \quad (3.19)$$

where k_{on} and k_{off} describe the shapes of the distributions and take integers; θ_{on} and θ_{off} are two parameters and take positive real numbers. Note that if $k_{on} = 1$ (or $k_{off} = 1$), the only one ON (or OFF) state is considered, corresponding to the case of exponential distribution (Markovian). By simple calculations, we can show

$$L_{on}(s) = (1 + s\theta_{on})^{-k_{on}}, \quad L_{off}(s) = (1 + s\theta_{off})^{-k_{off}}, \quad \langle \tau_{on} \rangle = k_{on}\theta_{on}, \quad \langle \tau_{off} \rangle = k_{off}\theta_{off},$$

$$\langle m \rangle = \mu \frac{\langle \tau_{on} \rangle}{\langle \tau_{on} \rangle + \langle \tau_{off} \rangle} = \mu \frac{k_{on}\theta_{on}}{k_{on}\theta_{on} + k_{off}\theta_{off}}, \quad (3.20)$$

$$\eta_m^2 = \frac{k_{on}\theta_{on} + k_{off}\theta_{off}}{\mu k_{on}\theta_{on}} + \frac{k_{off}\theta_{off}}{k_{on}\theta_{on}} - \frac{k_{on}\theta_{on} + k_{off}\theta_{off}}{(k_{on}\theta_{on})^2} \frac{(1 - (1 + \theta_{on})^{-k_{on}})(1 - (1 + \theta_{off})^{-k_{off}})}{1 - (1 + \theta_{on})^{-k_{on}}(1 + \theta_{off})^{-k_{off}}}.$$

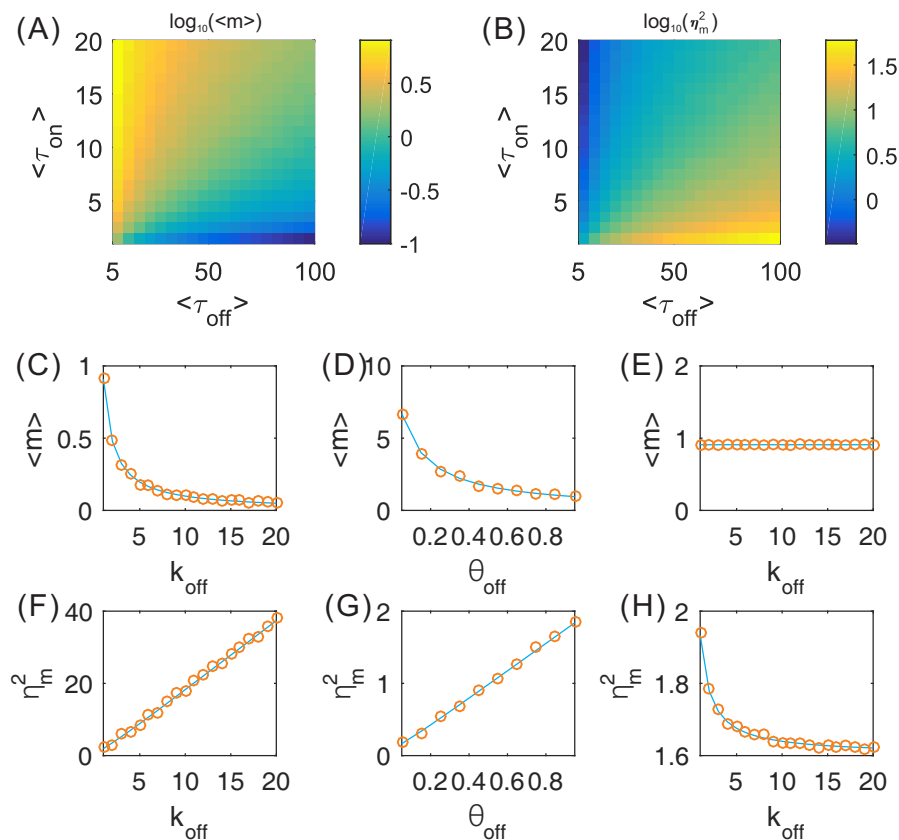


Figure 3. Effect of non-Markovianity on the mRNA noise. (A, B) phase diagrams for logarithms of means and noise intensities in the phase plane of mean OFF and ON times, $\langle \tau_{off} \rangle$ and $\langle \tau_{on} \rangle$; (C, D, E) represents the mRNA mean whereas (F, G, H) represents the corresponding mRNA noise. Parameter values: (A, B) $\mu = 10, k_{off} \in \{1, 2, \dots, 20\}$, $\theta_{off} = 5, \theta_{on} = 1, k_{on} \in \{1, 2, \dots, 20\}$; (C, F) $\mu = 10, k_{off} \in \{1, 2, \dots, 20\}, \theta_{off} = 1, k_{on} \in \{1, 2, \dots, 20\}, \theta_{on} = 0.1$; (D, G) $\mu = 10, k_{off} = 1, \theta_{off} = [0.05, 1], k_{on} = 1, \theta_{on} = 0.1$; (E, H) $\mu = 10, k_{off} \in \{1, 2, \dots, 20\}, \theta_{off} = 1/k_{off}, k_{on} = 1, \theta_{on} = 0.1$. In (C-H), empty circles are obtained by numerical realizations (Appendix D), whereas the solid lines are given by (3.16) and (3.17).

The numerical results are shown in Figure 3, which shows effects of non-Markovianity on the mRNA mean and the mRNA noise. Specifically, for the same mean ON and OFF times, a larger mRNA mean corresponds to a smaller mRNA noise (Figure 3(A, B)); the mRNA mean is a monotonically decreasing function of $k_{off} = 1$ (in a nonlinear manner. See Figure 3(C, D)) whereas the mRNA noise intensity is a monotonically increasing function of $k_{off} = 1$ (in an approximately linear manner. See Figure 3(F, G)). Figure 3(E, H) shows that if the mean mRNA is fixed, the mRNA noise is a monotonically decreasing function of $k_{off} = 1$, implying that non-Markovianity can reduce the mRNA noise.

Finally, we point out that at first glance, there is a big difference between the two models, but there is a close relationship between them. First, if the only first two moments are considered, the model analyzed in section 3 above (i.e., the queuing model) can be viewed as an extension of the model analyzed in section 2 above (i.e., the promoter model). In fact, (3.1) implies that the distributions of ON- and OFF-times in the promoter model are the algebraic sum of exponential functions but those of ON- and OFF-times in the queuing model may be more general distributions. Second, the mRNA mean is given by the same formula (3.16) for both models, and the formula for the mRNA noise in the queuing model, given by (3.17), is an extension of that for the mRNA noise in the promoter model, given by (3.18). Third, for both models, we can reconstruct the mRNA distribution from the respective CMs that are given by (2.5) for the promoter model or by the method given in the above subsection 3.2 for the queuing model. In addition, we can show that in the limit of fast switching, the mRNA distribution becomes Poissonian for both models. In fact, for the promoter model, this point is easily verified as for the common ON-OFF model. For the queuing model, the second term on the right-hand-side of (3.14) will disappear in the limit of fast switching. Some related discussions were done in [55].

4. Parametric regions for robust unimodality and bimodality

As a mechanism contributing to phenotypic diversity, bimodality of gene expression can enhance the survival of cells in a fluctuating environment. However, an unsolved question is how unimodality or bimodality appears or disappears. Here we address this question.

Although we have given analytical expressions of mRNA distributions in the previous sections, they are in general complex in form. On the other hand, a proper cellular phenotype characterized by the shape of some distribution is important for the cell to enhance the survival of cells in a fluctuating environment [33]. Thus, we naturally consider a question of how unimodal and bimodal expressions are robustly formed or how reaction rate parameters impact unimodal and bimodal regions. This nontrivial issue seems to have not been studied previously. Here we will address the issue by theoretical analysis combined with numerical simulation. The aim of this section is twofold: the one is to better help the reader understand the given-above complex distributions, and the other is to lay a theoretical foundation for finding parametric regions for unimodality and bimodality. We point out that biologically, if a unimodal distribution is wide enough, it can work as well as bimodal distribution in driving phenotypic heterogeneity [59].

For clarity, we will consider two cases separately. The following analysis is tedious, and the theoretical results are established based on properties of complex confluent hypergeometric functions.

4.1. Case of two-state gene model

Here, we consider and analyze a representative model of stochastic transcription. Assume that the gene promoter has one ON state and one OFF state (implying $N = 2$ in the above general model). For convenience, denote by λ the transition rate from ON to OFF state whereas by γ the transition rate from OFF to ON state and by μ the transcription rate of mRNA, and set the mRNA degradation rate as $\delta = 1$ (this is equivalent to the normalization of three parameters μ , λ and γ by δ). According to the above analysis, the corresponding mRNA distribution can be expressed as

$$P(m) = \frac{\Gamma(\lambda + m)\mu^m}{m!\Gamma(\lambda + \gamma + m)} \frac{\Gamma(\lambda + \gamma)}{\Gamma(\lambda)} {}_1F_1(\lambda + m, \lambda + \gamma + m; -\mu). \quad (4.1)$$

Note that $\mu > 1$ always holds since μ represents the transcription rate normalized by the degradation rate. We distinguish two cases to give theoretical results on unimodal and bimodal mRNA distributions: (A) $P(1) > P(0)$; (B) $P(1) < P(0)$.

Theorem 2 $P(1) > P(0)$. For any $\mu > 1, \lambda > 0$ and $\gamma > 0$, there is an integer $m_0 > 0$ such that $P(m + 1) - P(m) > 0$ for $0 \leq m \leq m_0$ whereas $P(m + 1) - P(m) \leq 0$ for $m > m_0$. In this case, $P(m)$ is unimodal.

Proof Using the integral expression of hypergeometric function

$${}_1F_1(\alpha, \beta; z) = \frac{\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta - \alpha)} \int_0^1 e^{zt} t^{\alpha-1} (1-t)^{\beta-\alpha-1} dt, \quad (4.2)$$

we can rewrite Eq (4.1) as

$$P(m) = \frac{\mu^m \Gamma(\lambda + \gamma)}{m! \Gamma(\lambda)\Gamma(\gamma)} \int_0^1 e^{-\mu t} t^{\lambda+m-1} (1-t)^{\gamma-1} dt. \quad (4.3)$$

Thus,

$$P(m + 1) - P(m) = \frac{\mu^m \Gamma(\lambda + \gamma)}{m! \Gamma(\lambda)\Gamma(\gamma)} \int_0^1 e^{-\mu t} t^{\lambda+m-1} (1-t)^{\gamma-1} \left(\frac{\mu t}{m+1} - 1 \right) dt \quad (4.4)$$

implying that $P(m + 1) - P(m) < 0$ for any $m > \tilde{m} = \mu - 1 > 0$. On the other hand, we have $P(1) - P(0) > 0$. Therefore, if we set $m_0 = \min\{m | P(m + 1) - P(m) < 0, m > \tilde{m}\}$, then we know that $P(m + 1) - P(m) \geq 0$ for $0 \leq m \leq m_0$ whereas $P(m + 1) - P(m) < 0$ for $m > m_0$. This finished the proof of theorem 2.

The reader can refer to the numerical result in Figure 4, where a representative mode of mRNA distribution is shown. This figure verifies that our theoretical prediction is in accordance with the numerical result.

Theorem 3 For the case of $P(1) < P(0)$, we have: (1) If $\lambda \geq 1$ and $\mu\lambda \leq \lambda + \gamma$, or if $\lambda < 1$ and $\mu \leq \min(\gamma + 1 - \lambda + 2\sqrt{(1-\lambda)\gamma}, 1 + \gamma/\lambda)$, $P(m)$ is unimodal; (2) If $\mu \geq \lambda + \gamma + 2$ and $2(\gamma + 1 - 2\lambda)\mu \leq (\lambda + \gamma - 1)^2 - 1$, $P(m)$ is bimodal.

The proof of this theorem is put on Appendix E. Here we show numerical results, referring to Figure 5.

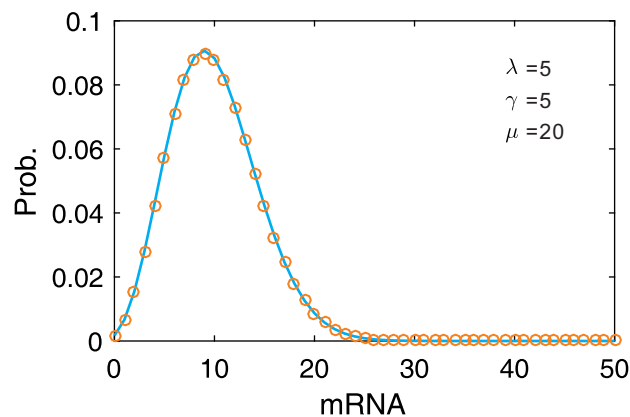


Figure 4. A representative mode of mRNA distribution in case of $P(1) > P(0)$. Empty circles are obtained by numerical realizations with Gillespie simulation algorithm [60]. Numerical simulation verifies that the mRNA distribution is only unimodal for any parameter values in this case.

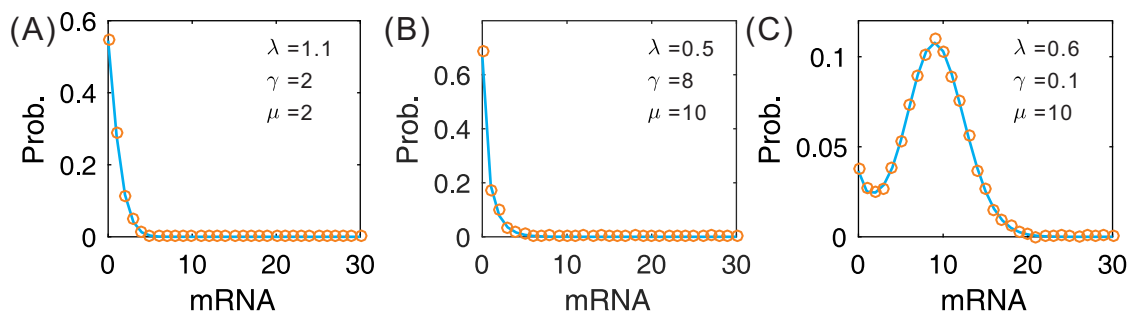


Figure 5. Three representative modes of mRNA distribution in the case of $P(1) < P(0)$, where (A) and (B) correspond to unimodal distribution whereas (C) corresponds to bimodal distribution. Empty circles are obtained by numerical realizations with Gillespie simulation algorithm [60].

4.2. Case of multiple-state gene model

Here, we only analyze a simple case: the promoter has one active state and two inactive states. Denote by λ_2 and λ_2' the transition rates between two OFF states, by λ_1 and γ_1 the transition rates from one OFF state to ON state and from ON state to the OFF state respectively, and by λ_2 and γ_2 the transition rates from the other OFF state to ON state and from ON state to the OFF state, respectively. According to the previous analysis, we know that the mRNA number follows a hypergeometric following distribution given by

$$P(m) = \frac{\mu^m (\beta_1)_m (\beta_2)_m}{m! (\alpha_1)_m (\alpha_2)_m} {}_2F_2(m + \beta_1, m + \beta_2; m + \alpha_1, m + \alpha_2; -\mu), m = 0, 1, 2, \dots, \quad (4.5)$$

where

$$\alpha_{1,2} = \frac{1}{2} \left(\gamma_1 + \gamma_2 + \lambda_1 + \lambda_2 + \lambda'_2 + \lambda_3 \pm \sqrt{(\gamma_1 + \gamma_2 + \lambda_1 - \lambda_2 - \lambda'_2 - \lambda_3)^2 + 4(\lambda_1 - \lambda_3)(\lambda_2 - \gamma_2)} \right),$$

$$\beta_{1,2} = \frac{1}{2} \left(\lambda_1 + \lambda_2 + \lambda'_2 + \lambda_3 \pm \sqrt{(\lambda_1 + \lambda_2 - \lambda'_2 - \lambda_3)^2 + 4\lambda_2\lambda'_2} \right).$$

Note that all the parameters have been normalized by the degradation rate δ , and α_i and β_i , $i = 1, 2$, may be complex. In the following, we only consider they are real.

In order to find parametric regions for unimodality and bimodality, we introduce the following lemma.

Lemma 1 *If $b \geq a$ and $c \geq d$, then we have the Kummer-type transform of the form*

$${}_2F_2(a, b; d, c; x) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{xt} t^{a-1} (1-t)^{b-a-1} {}_1F_1(c-d, c; -xt) dt. \quad (4.6)$$

Note that for any m , we have

$$\frac{P(m+1)}{P(m)} = \frac{\mu}{m+1} \frac{(\beta_1+m)(\beta_2+m)}{(\alpha_1+m)(\alpha_2+m)} \frac{{}_2F_2(m+1+\beta_1, m+1+\beta_2; m+1+\alpha_1, m+1+\alpha_2; -\mu)}{{}_2F_2(m+\beta_1, m+\beta_2; m+\alpha_1, m+\alpha_2; -\mu)}, \quad (4.7)$$

where two confluent functions are given according to Eq (4.6). Using Lemma 2, it is not difficult to prove the following two lemmas.

Lemma 2 *For any nonnegative integer m , then we have the inequality*

$$P(m+1) < \frac{\mu}{m+1} \frac{(\beta_1+m)(\beta_2+m)}{(\alpha_1+m)(\alpha_2+m)} P(m) \quad (4.8)$$

Lemma 3 *For any m , we have the following estimation*

$$P(m+1) > \xi_2 \frac{\mu}{m+1} \frac{(\beta_1+m)(\beta_2+m)}{(\alpha_1+m)(\alpha_2+m)} P(m) \quad (4.9)$$

where $\xi_2 = \left(1 - \frac{\mu}{4(m+\alpha_2)}\right) \left(1 - \frac{\mu}{4(m+\beta_2)}\right)$.

Similar to the case of the common two-state model, we may distinguish two cases to find parametric regions for unimodality and bimodality in a gene model with one active state and two inactive states. That is, we consider two cases separately: $P(1) > P(0)$ and $P(1) < P(0)$. The key equation to be solved is that under what conditions of reaction rates, the difference $P(m+1) - P(m)$ is positive or negative. When one wants to estimate the sign of this difference, the above three lemmas are useful.

By analysis, one is able to know that $P(m)$ is unimodal under some conditions of reaction rates whereas it is bimodal under other conditions of reaction rates. For example, we can rigorously prove the following result: For $P(1) < P(0)$, if $\mu\beta_1\beta_2 < \alpha_1\alpha_2$ and $\mu f_2(\tilde{x}) \leq 1$ are simultaneously satisfied, where $f_2(x) = [(\beta_1+x)(\beta_2+x)]/[(x+1)(\alpha_1+x)(\alpha_2+x)]$ and \tilde{x} is the extreme point of $f_2(x)$, $P(m)$ is unimodal. In fact, first note that $\mu\beta_1\beta_2 < \alpha_1\alpha_2$ can guarantee $P(1) < P(0)$ due to Eq (4.8). Second, for function $f_2(x)$, we know by calculation that there is a \tilde{x} such that $f_2(x)$ reaches its maximum at $x = \tilde{x}$, referring to the inset of Figure 6(A). Apparently, if $\mu f_2(\tilde{x}) \leq 1$, $P(m+1) < P(m)$ for all m , indicating that $P(m)$ monotonically decreases with regard to m . Therefore, $P(m)$ is unimodal. The numerical simulation has verified the correctness of our analysis, referring to Figure 6(A).

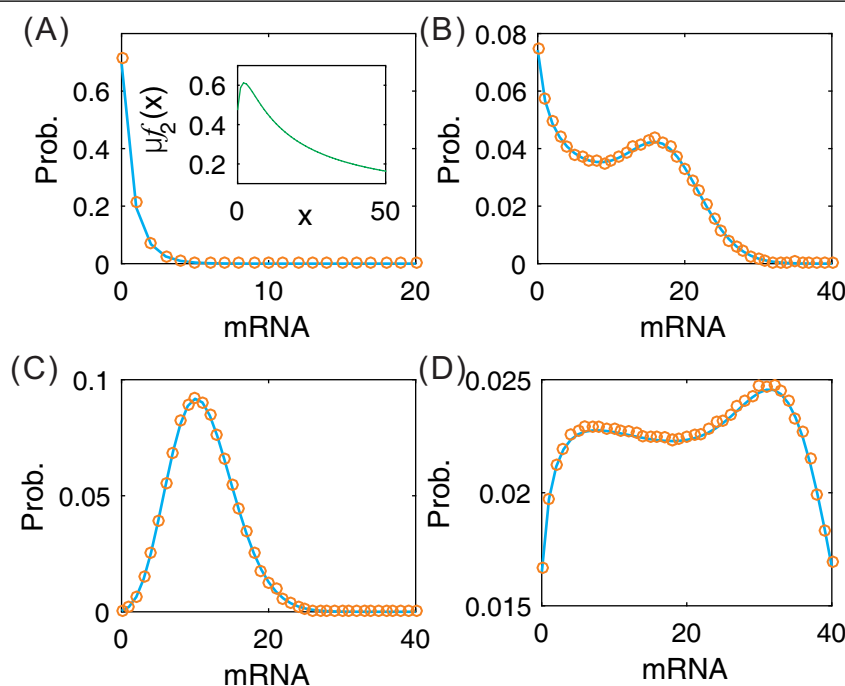


Figure 6. Representative modes of mRNA distribution in a model of stochastic transcription with the promoter having one ON state and two OFF states: (A) and (B) $P(1) < P(0)$; (C) and (D) $P(1) > P(0)$. Parameter values are set as: (A) $\gamma_1 = 10, \gamma_2 = 0, \lambda_2 = \lambda_3 = 1, \lambda_1 = \lambda'_2 = 0, \mu = 10$; (B) $\gamma_1 = 0.4, \gamma_2 = 0, \lambda_2 = \lambda_3 = 1, \lambda_1 = \lambda'_2 = 0, \mu = 20$; (C) $\gamma_1 = 4, \gamma_2 = 0, \lambda_2 = \lambda_3 = 10, \lambda_1 = \lambda'_2 = 0, \mu = 20$; (D) $\gamma_1 = 0.6, \gamma_2 = 0, \lambda_2 = \lambda_3 = 1.5, \lambda_1 = \lambda'_2 = 0, \mu = 40$. In all the cases, $\delta = 1$. Empty circles are obtained by numerical realizations with Gillespie simulation algorithm [60].

5. Conclusion and discussion

Transcription is a core process in life, but how upstream promoter kinetics affects downstream transcription dynamics remains to be not fully understood. Here, we have analyzed two representative yet general models of stochastic transcription. For the queuing model of transcription with general waiting-time distributions, we have derived a general formula for calculating raw moments of any order. In turn we have proposed an efficient method—the CM method, to reconstruct the mRNA distribution, where CMs are actually linear combinations of raw moments but the former has better properties than the latter, e.g., CMs converge to zero as their orders tend to infinity. For the multistate model of transcription with arbitrarily complex promoter structure, we have derived exact steady-state mRNA distributions, which are expressed by generalized confluent hypergeometric functions. Moreover, we have found that there are conjugate relationships among these complex distributions. For both models, we have shown that the mRNA distribution is in general sup-Poissonian. In addition, we have analyzed robust regions for unimodality and bimodality for two specific yet representative models. By combining numerical simulation, we have found parametric regions for how unimodal and bimodal expressions are robustly shaped. Importantly, our analysis methods and

results can save time for investigate gene expression behavior than numerical simulations. Our analysis lays a foundation for quantitative analysis of stochastic gene transcription whereas our results would be helpful for designing biologically robust gene regulatory circuits.

Although our analyzed models of stochastic transcription are general, i.e., they include many previously-studied models as their particular cases, we still neglect some other processes underlying transcription, such as alternative splicing [35], DNA methylation [18], RNA nuclear retention [61], cell division [62], negative feedback [19, 32] and spatial heterogeneity [63, 64]. These processes or factors can also impact the transcription level or transcription efficiency, but how this impact is mathematically analyzed and how exact analytical results are derived would need to develop new mathematical models and new analytical methods. In particular, when extracellular dynamic and stochastic regulatory signals are considered, the theoretical or numerical analysis would be challenging since the corresponding chemical master equation will become a non-Markovian system in this case. In addition, we have given conservative parametric regions for robust unimodal and bimodal expressions by analyzing complex mRNA distributions in two simple models of stochastic transcription, but a gene model with multiple active states may exhibit various modes of distribution, including unimodality, bimodality and multimodality [25]. Theoretically proving this point or giving exact parametric regions for multimodality seems difficult.

We point out that our first model of stochastic transcription belongs to the so-called $GI^X/G/\infty$ system in queueing theory [50, 54]. Using this theory, one can also derive formulae for calculating raw moments in a gene model of stochastic transcription where the mRNA lifetime follows a general distribution rather than an exponential distribution considered here, besides that ON and OFF times follow general distributions. In addition, we point out that it looks like a big difference between the two gene models studied above, but we can establish the relationship between them by calculating the mRNA noise. This is because that (3.16) always holds for both models. These models are all convergent, referring to some discussions in [66], and the corresponding mRNA distributions exhibit discontinuities [68], which can easily be verified from the expressions of the distributions reconstructed from our CMs.

Until now, the study of stochastic transcription focuses on the following two classes of models: (1) The first class is the so-called promoter models where the promoter has arbitrary activity states (either ON or OFF) and there are random transitions among these states, e.g., the model studied in this paper or referring to Figure 7(A); (2) The second class is the so-called queuing models, where the waiting times, e.g., from OFF to ON states, from ON to OFF states, from DNA to mRNA, or for a multistep process of mRNA degradation, follow a distribution. Note that this class of models can further be classified as more specified models by considering combinations of the mentioned cases, e.g., the state-waiting model in Figure 7(B) and the transcription-waiting model in Figure 7(C). These two classes of models were partially studied [25, 56, 66, 68]) and some interesting results were also obtained [25, 52, 53, 56, 66, 68]. In principle, our queuing model studied in this paper includes the queuing models in the existing literature as its special cases, and our results obtained in the paper can partially reproduce some previous results, at least those for the first two moments obtained under the same assumed conditions. Although we have established partial connections between the two classes of models, deeper connections are worth further study.

Finally, it is worth mentioning that the CM method that we simply introduced here would have broad applications in modeling, analysis and simulation of biochemical reaction networks including complex

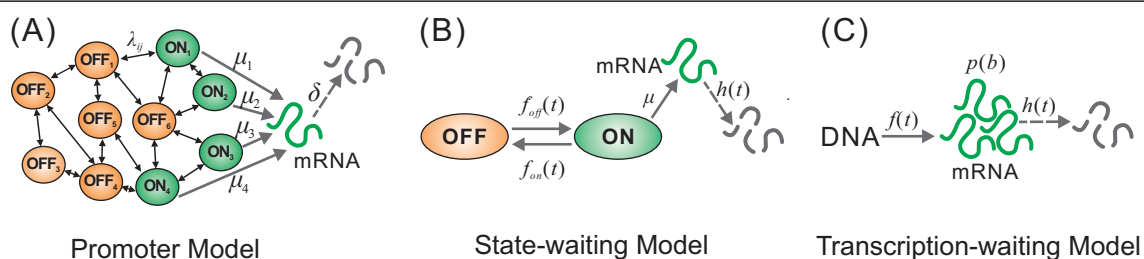


Figure 7. Schematic diagrams for three representative classes of gene models: (A) promoter model, where the promoter contains many (ON and OFF) states and there are transitions among these states, and the mRNA degradation rate is constant; (B) state-waiting model, where the promoter contains one ON state and one OFF state but the times that the gene dwell at these two states follow non-exponential distributions, the mRNAs degrade in a multistep manner (i.e., the corresponding times follow a non-exponential distribution), and the transcription rate is constant; (C) transcription-waiting model, where the waiting times from DNA to mRNA follow a general distribution, the mRNAs are produced in a bursty manner, and the mRNAs degrade in a multistep manner.

gene regulatory networks. In fact, apart from two remarkable advantages mentioned in this paper, i.e., CMs tend to zero as their orders go to infinity; CMs can be used to reconstruct the corresponding distribution, there are other advantages, e.g., according to the definition of CMs, we can derive the time evolution equations for CMs based on the chemical master equation, which are a linear equation group with constant coefficients. Thus, one can use CM equations to study how upstream promoter dynamics affect downstream expression dynamics, and may reveal dynamical mechanisms behind it, which would be difficult by directly analyzing the chemical master equation.

Acknowledgments

This work was partially supported by Grant Nos. 91530320, 91230204, 2014CB964703, 11475273, and 11631005.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. L. Cai, N. Friedman and X. S. Xie, Stochastic protein expression in individual cells at the single molecule level, *Nature*, **440** (2006), 358–362.
2. J. R. Chubb and T. B. Liverpool, Bursts and pulses: insights from single cell studies into transcriptional mechanisms, *Curr. Opin. Genet. Dev.*, **20** (2010), 478–84.
3. D. R. Larson, What do expression dynamics tell us about the mechanism of transcription? *Curr. Opin. Genet. Dev.*, **21** (2011), 591–599.

4. A. Sanchez and I. Golding, Genetic determinants and cellular constraints in noisy gene expression, *Science*, **342** (2013), 1188–1193.
5. J. Yu, J. Xiao, X. J. Ren, et al., Probing gene expression in live cells, one protein molecule at a time, *Science*, **311** (2006), 1600–1603.
6. W. J. Blake, M. Karn, C. R. Cantor, et al., Noise in eukaryotic gene expression, *Nature*, **422** (2003), 633–637.
7. J. R. Chubb, T. Trcek, S. M. Shenoy, et al., Transcriptional pulsing of a developmental gene, *Curr. Biol.*, **16** (2006), 1018–1025.
8. I. Golding, J. Paulsson, S. M. Zawilski, et al., Real-time kinetics of gene activity in individual bacteria, *Cell*, **123** (2005), 1025–1036.
9. A. Raj, C. S. Peskin, D. Tranchina, et al., Stochastic mRNA synthesis in mammalian cells, *PLoS Biol.*, **4** (2006), e309.
10. H. Boeger, J. Griesenbeck and R. D. Kornberg, Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription, *Cell*, **133** (2008), 716–726.
11. C. V. Harper, B. Finkenstadt, D. J. Woodcock, et al., Dynamic analysis of stochastic transcription cycles, *PLoS Biol.*, **9** (2011), e1000607.
12. C. Mao, C. R. Brown, S. F. Dong, et al., Quantitative analysis of the transcription control mechanism, *Mol. Syst. Biol.*, **6** (2010), 431.
13. L. Mariani, E. G. Schulz, E. Falkovskaia, et al., Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression, *Mol. Syst. Biol.*, **6** (2010), 359.
14. K. Miller-Jensen, S. S. Dey, D. V. Schaffer, et al., Varying virulence: epigenetic control of expression noise and disease processes, *Trends Biotechnol.*, **29** (2011), 517–525.
15. J. Raser and E. O’Shea, Control of stochasticity in eukaryotic gene expression, *Science*, **304** (2003), 1811–1814.
16. D. M. Suter, N. Molina, D. Gatfield, et al., Mammalian genes are transcribed with widely different bursting kinetics, *Science*, **332** (2011), 472–474.
17. G. Hager, C. Elbi, T. A. Johnson, et al., Chromatin dynamics and the evolution of alternate promoter states, *Chromosome Res.*, **14** (2006), 107–116.
18. T. J. Stasevich and J. G. McNally, Assembly of the transcription machinery: ordered and stable, random and dynamic, or both? *Chromosoma*, **120** (2011), 533–545.
19. L. Weinberger, Y. Voichek, I. Tirosh, et al., Expression noise and acetylation profiles distinguish HDAC functions, *Mol. Cell*, **47** (2012), 193–202.
20. A. Schwabe, K. N. Rybakova and F. J. Bruggeman, Transcription stochasticity of complex gene regulation models, *Biophys. J.*, **103** (2012), 1152–1161.
21. W. J. Blake, G. Balázsi, M. A. Kohanski, et al., Phenotypic consequences of promoter-mediated transcriptional noise, *Mol. Cell*, **6** (2006), 853–865.
22. M. Dobrzynski and F. J. Bruggeman, Elongation dynamics shape bursty transcription and translation, *Proc. Natl. Acad. Sci. USA*, **106** (2009), 2583–2588.

23. A. Sanchez, H. G. Garcia, D. Jones, et al., Effect of promoter architecture on the cell-to-cell variability in gene expression, *PLoS. Comput. Biol.*, **7** (2011), e1001100.
24. J. J. Zhang, L. N. Chen and T. S. Zhou, Analytical distribution and tunability of noise in a model of promoter progress, *Biophys. J.*, **102** (2012), 1247–1257.
25. J. J. Zhang and T. S. Zhou, Promoter-mediated transcriptional dynamics, *Biophys. J.*, **106** (2014), 479–488.
26. J. M. G. Vilar and L. Saiz, CplexA: a mathematica package to study macromolecular-assembly control of gene expression, *Bioinform.*, **26** (2010), 2060–2061.
27. G. Hornung, R. Bar-Ziv, D. Rosin, et al., Noise-mean relationship in mutated promoters, *Genome Res.*, **22** (2012), 2409–2417.
28. A. Halme, S. Bumgarner, C. Styles, et al., Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast, *Cell*, **116** (2004), 405–415.
29. L. M. Octavio, K. Gedeon and N. Maheshri, Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression, *PLoS. Genet.*, **5** (2009), e1000673.
30. L. Bintu, N. E. Buchler, H. G. Garcia, et al., Transcriptional regulation by the numbers: models, *Curr. Opin. Genet. Dev.*, **15** (2005), 116–124.
31. D. Huh and J. Paulsson, Non-genetic heterogeneity from stochastic partitioning at cell division., *Nat. Genet.*, **43** (2011), 95–100.
32. T. L. To and N. Maheshri, Noise can induce bimodality in positive transcriptional feedback loops without bistability, *Science*, **327** (2010), 1142–1145.
33. R. Skupsky, J. C. Burnett, J. E. Foley, et al., HIV promoter integration site primarily modulates transcriptional burst size rather than frequency, *PLoS. Comput. Biol.*, **6** (2010), e1000952.
34. L. S. Weinberger, J. C. Burnett, J. E. Toettcher, et al., Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity, *Cell*, **122** (2005), 169–182.
35. Q. L. Wang and T. S. Zhou, Alternative-splicing-mediated gene expression, *Phys. Rev. E.*, **89** (2014), 012713.
36. A. Ochab-Marcinek and M. Tabaka, Bimodal gene expression in noncooperative regulatory systems, *Proc. Natl. Acad. Sci. USA*, **107** (2010), 22096–22101.
37. N. Friedman, L. Cai and X. S. Xie, Linking stochastic dynamics to population- distribution: an analytical framework of gene expression, *Phys. Rev. Lett.*, **97** (2006), 168302.
38. P. J. Liu, Z. J. Yuan, L. F. Huang, et al., Roles of factorial noise in inducing bimodal gene expression, *Phys. Rev. E.*, **91** (2015), 062706.
39. V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, *Proc. Natl. Acad. Sci. USA*, **105** (2008), 17256–17261.
40. B. Bodenmiller, E. R. Zunder, R. Finck, et al., Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators, *Nat. Biotechnol.*, **30** (2012), 858.

41. N. Rimon and M. Schuldiner, Getting the whole picture: combining throughput with content in microscopy, *J. Cell. Sci.*, **124** (2011), 3743.
42. Y. Liu, A. Beyer and R. Aebersold, On the dependency of cellular protein levels on mRNA abundance, *Cell*, **165** (2016), 535–550.
43. J. J. Zhang, L. F. Huang and T. S. Zhou, A moment-convergence method for stochastic analysis of biochemical reaction networks, *J. Chem. Phys.*, **144** (2016), 018620.
44. L. J. Slater, *Confluent Hypergeometric Functions*, Cambridge University Press, Cambridge, 1960.
45. Y. He and E. Barkai, Super- and sub-Poissonian photon statistics for single molecule spectroscopy, *J. Chem. Phys.*, **122** (2005), 184703.
46. J. M. Pedraza and J. Paulsson, Effects of molecular memory and bursting on fluctuations in gene expression, *Science*, **319** (2008), 339–343.
47. L. F. Huang, Z. J. Yuan, P. J. Liu, et al., Feedback-induced counterintuitive correlations of gene expression noise with bursting kinetics, *Phys. Rev. E.*, **90** (2014), 052702.
48. J. Paulsson, Models of stochastic gene expression, *Phys. Life. Rev.*, **2** (2005), 157–175.
49. J. Peccoud and B. Ycart, Markovian modelling of gene product synthesis, *Theor. Popul. Biol.*, **48** (1995), 222–234.
50. J. Tao and R. V. Kulkarni, Intrinsic noise in stochastic models of gene expression with molecular memory and bursting, *Phys. Rev. Lett.*, **106** (2011), 058102.
51. T. B. Kepler and T. C. Elston, Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations, *Biophys. J.*, **81** (2001), 3116–3036.
52. B. J. Daigle, Jr. M. Soltani, L. R. Petzold, et al., Inferring single-cell gene expression mechanisms using stochastic simulation, *Bioinformatics*, **31** (2015), 1428–1435.
53. B. Zoller, D. Nicolas, N. Molina, et al., Structure of silent transcription intervals and noise characteristics of mammalian genes, *Mol. Syst. Biol.*, **11** (2015), 823.
54. L. Liu, B. R. K. Kashyap and J. G. C. Templeton, On the $G^X/G/\infty$ system, *J. Appl. Probab.*, **27** (1990), 671–683.
55. M. Thattai, Universal Poisson statistics of mRNAs with complex decay pathways, *Biophys. J.*, **110** (2016), 301–305.
56. A. R. Stinchcombe, C. S. Peskin and D. Tranchina, Population density approach for discrete mRNA distributions in generalized switching models for stochastic gene expression, *Phys. Rev. E.*, **85** (2012), 061919.
57. T. S. Zhou and J. J. Zhang, Analytical results for a multistate gene model, *SIAM. J. Appl. Math.*, **72** (2012), 789–818.
58. A. Sánchez, S. Choubey and J. Kondev, Stochastic models of transcription: From single molecules to single cells, *Methods*, **62** (2013), 13–25.
59. J. Garcia-Bernardo and M. J. Dunlop, Phenotypic diversity using bimodal and unimodal expression of stress response proteins, *Biophys. J.*, **110** (2016), 2278–2287.
60. D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J. Comput. Phys.*, **22** (1976), 403.

61. Q. L. Wang and T. S. Zhou, Dynamical analysis of mCAT2 gene models with CTN-RNA nuclear retention, *Phys. Biol.*, **12** (2015), 016010.
62. H. H. Wang, Z. J. Yuan, P. J. Liu, et al., Division time-based amplifiers for gene expression, *Mol. Biosyst.*, **11** (2015), 2417–2428.
63. M. Sturrock, P. J. Murray, A. Matzavinos, et al., Mean field analysis of a spatial stochastic model of a gene regulatory network, *J. Math. Biol.*, **71** (2015), 921–959.
64. M. Sturrock, S. Li and V. Shahrezaei, The influence of nuclear compartmentalisation on stochastic dynamics of self-repressing gene expression, *J. Theor. Biol.*, **424** (2017), 55–72.
65. N. Balakrishnan, N. L. Johnson and S. Kotz, A note on relationships between moments, central moments and cumulants from multivariate distributions, *Stat. Prob. Lett.* **39** (1998), 49–54.
66. N. Kumar, A. Singh and R. V. Kulkarni, Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models, *PLoS. Comput. Biol.*, **11** (2015), e1004292.
67. L. Takács, *An Introduction to queueing theory*, Oxford University Press, New York, 1962.
68. H. Xu, S. O. Skinner, A. M. Sokac, et al., Stochastic kinetics of nascent RNA, *Phys. Rev. Lett.*, **117** (2016), 128101.

Appendix

A. Derivation of (2.4) in the main text

Let $b_k^{(i)}$ be the convergent moments corresponding to the factorial distribution P_i . According to the CM definition, we have

$$b_k^{(i)}(t) = \sum_{m \geq k} \binom{m}{k} P_i(m, t), \text{ where } 1 \leq i \leq N, k = 0, 1, 2, \dots \quad (\text{A1})$$

Note that (1) in the main text is expressed in the component as

$$\begin{aligned} \frac{\partial}{\partial t} \begin{pmatrix} P_1(m; t) \\ P_2(m; t) \\ \vdots \\ P_N(m; t) \end{pmatrix} &= \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1N} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{NN} \end{pmatrix} \begin{pmatrix} P_1(m; t) \\ P_2(m; t) \\ \vdots \\ P_N(m; t) \end{pmatrix} \\ &+ \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_N \end{pmatrix} \left[\begin{pmatrix} P_1(m-1; t) \\ P_2(m-1; t) \\ \vdots \\ P_N(m-1; t) \end{pmatrix} - \begin{pmatrix} P_1(m; t) \\ P_2(m; t) \\ \vdots \\ P_N(m; t) \end{pmatrix} \right] \\ &+ \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_N \end{pmatrix} \left[\begin{pmatrix} P_1(m+1; t) \\ P_2(m+1; t) \\ \vdots \\ P_N(m+1; t) \end{pmatrix} - \begin{pmatrix} P_1(m; t) \\ P_2(m; t) \\ \vdots \\ P_N(m; t) \end{pmatrix} \right]. \end{aligned}$$

Also note that for any fixed $i : 1 \leq i \leq N$, we have

$$\begin{aligned} \sum_{m \geq k} \binom{m}{k} [P_i(m-1; t) - P_i(m; t)] &= \sum_{m+1 \geq k} \binom{m+1}{k} P_i(m; t) - \sum_{m \geq k} \binom{m}{k} P_i(m; t) \\ &= \binom{k}{k} P_i(k-1; t) + \sum_{m \geq k} \left[\binom{m+1}{k} - \binom{m}{k} \right] P_i(m; t) \\ &= P_i(k-1; t) + \sum_{m \geq k} \binom{m}{k-1} P_i(m; t) \\ &= \sum_{m \geq k} \binom{m}{k} P_i(m; t) = b_k^{(i)}(t), \end{aligned}$$

$$\begin{aligned} \sum_{m \geq k} \binom{m}{k} [P_i(m+1; t) - P_i(m; t)] &= \sum_{m-1 \geq k} \binom{m-1}{k} P_i(m; t) - \sum_{m \geq k} \binom{m}{k} P_i(m; t) \\ &= \sum_{m \geq k+1} \left[\binom{m-1}{k} \right] P_i(m; t) - \binom{k}{k} P_i(k; t) \\ &= - \sum_{m \geq k+1} \binom{m-1}{k-1} P_i(k; t) \\ &= -k \sum_{m \geq k} \binom{m}{k} P_i(m; t) = -k b_k^{(i)}(t). \end{aligned}$$

Therefore, the summing over all the possible m being greater than or equal to k in (A1) yields

$$\begin{aligned} \frac{\partial}{\partial t} \begin{pmatrix} b_k^{(1)}(t) \\ b_k^{(2)}(t) \\ \vdots \\ b_k^{(N)}(t) \end{pmatrix} &= \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1N} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{NN} \end{pmatrix} \begin{pmatrix} b_k^{(1)}(t) \\ b_k^{(2)}(t) \\ \vdots \\ b_k^{(N)}(t) \end{pmatrix} \\ &+ \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_N \end{pmatrix} \begin{pmatrix} b_{k-1}^{(1)}(t) \\ b_{k-1}^{(2)}(t) \\ \vdots \\ b_{k-1}^{(N)}(t) \end{pmatrix} - k \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_N \end{pmatrix} \begin{pmatrix} b_k^{(1)}(t) \\ b_k^{(2)}(t) \\ \vdots \\ b_k^{(N)}(t) \end{pmatrix} \end{aligned}$$

Assume that all the degradation rates are equal and denote by δ the common degradation rate. Using δ to normalize all the parameters and setting $\tau = \delta t$, we can obtain

$$\frac{d\mathbf{b}_k}{d\tau} = \mathbf{W}\mathbf{b}_k + \Lambda\mathbf{b}_{k-1} - k\mathbf{b}_k, \quad k = 1, 2, \dots, \quad (\text{A2})$$

where $\mathbf{b}_k = (b_k^{(1)}(t), b_k^{(2)}(t), \dots, b_k^{(N)}(t))^T$ is a column vector. (A2) is just (4) in the main text.

B. Details for the proof of proposition 1 in the main text

At steady state, it follows from (4) in the main text or (A2) that

$$(k\mathbf{I} - \mathbf{W})\mathbf{b}_k = \Lambda\mathbf{b}_{k-1}, \quad k = 1, 2, \dots \quad (\text{B1})$$

Note that the \mathbf{W} matrix describing the promoter structure is an M-matrix (i.e., the sum of elements in every row is equal to zero, implying that the determinant of \mathbf{W} is equal to zero, or $\det(\mathbf{W}) = 0$). Therefore, $(k\mathbf{I} - \mathbf{W})$ is a reversible matrix with the inverse given by $(k\mathbf{I} - \mathbf{W})^{-1} = \frac{1}{\det(k\mathbf{I} - \mathbf{W})}(k\mathbf{I} - \mathbf{W})^*$ in which $(i\mathbf{I} - \mathbf{W})^*$ and $\det(i\mathbf{I} - \mathbf{W})$ are the adjacency matrix and the determinant of matrix $(i\mathbf{I} - \mathbf{W})$, respectively. Left multiplying $(k\mathbf{I} - \mathbf{W})^{-1}$ on both sides of (B1) yields

$$\mathbf{b}_k = \frac{1}{\det(k\mathbf{I} - \mathbf{W})}(\mathbf{H} - \mathbf{W})^* \Lambda \mathbf{b}_{k-1}, \quad k = 1, 2, \dots \quad (\text{B2})$$

from which we have

$$\mathbf{b}_k = \left[\frac{(k\mathbf{I} - \mathbf{W})^* \Lambda}{\det(k\mathbf{I} - \mathbf{W})} \right] \mathbf{b}_{k-1} = \left[\prod_{i=k}^1 \frac{(i\mathbf{I} - \mathbf{W})^* \Lambda}{\det(i\mathbf{I} - \mathbf{W})} \right] \mathbf{b}_0, \quad k = 1, 2, \dots \quad (\text{B3})$$

If we denote by $\mathbf{u}_N = (1, 1, \dots, 1)$ a N -dimensional row vector, $\mathbf{u}_N \mathbf{b}_k = \sum_{i=1}^N b_k^{(i)} = b_k$ holds. Left multiplying \mathbf{u}_N on both sides of (B3) yields

$$b_k = \mathbf{u}_N \left[\prod_{i=k}^1 \frac{(i\mathbf{I} - \mathbf{W})^* \Lambda}{\det(i\mathbf{I} - \mathbf{W})} \right] \mathbf{b}_0, \quad k = 1, 2, \dots, \quad (\text{B4})$$

where $\mathbf{b}_0 = (b_0^{(1)}, b_0^{(2)}, \dots, b_0^{(N)})^T$ is a column vector.

Next, we derive the expressions of $b_0^{(i)}$ ($1 \leq i \leq N$). The results are

$$b_0^{(j)} = \prod_{i=1}^{N-1} \frac{\beta_i^{(j)}}{\alpha_i}, \quad 1 \leq j \leq N, \quad (\text{B5})$$

where $-\alpha_1, -\alpha_2, \dots, -\alpha_{N-1}$ are nonzero characteristic values of matrix \mathbf{W} whereas $-\beta_1^{(j)}, -\beta_2^{(j)}, \dots, -\beta_{N-1}^{(j)}$ are those of \mathbf{W}_j , the minor matrix of \mathbf{W} after crossing out the j^{th} row and j^{th} column of the entry λ_{ij} of \mathbf{W} . In fact, $\det(\mathbf{W}) = 0$ implies $\mathbf{W}\mathbf{W}^* = \det(\mathbf{W})\mathbf{I} = \mathbf{0}$. As a direct consequence of Laplace's formula for the determinant of matrix \mathbf{W} , we have

$$\mathbf{W} \begin{pmatrix} \det(\mathbf{W}_1) \\ \vdots \\ \det(\mathbf{W}_N) \end{pmatrix} = \mathbf{0}, \quad (\text{B6})$$

where

$$\det(\mathbf{W}_k) = (-1)^{N-1} \prod_{i=1}^{N-1} \beta_i^{(k)}, \quad 1 \leq k \leq N \quad (\text{B7})$$

since $-\beta_1^{(j)}, -\beta_2^{(j)}, \dots, -\beta_{N-1}^{(j)}$ are the characteristic values of matrix \mathbf{W}_j . Note that the null space of \mathbf{W} is one-dimensional due to $\text{rank}(\mathbf{W}) = N - 1$, and that (B1) still holds for $k = 0$ (i.e., $\mathbf{W}\mathbf{b}_0 = \mathbf{0}$). Therefore, from the combination of (B6) and $\mathbf{W}\mathbf{b}_0 = \mathbf{0}$, we can express \mathbf{b}_0 as

$$\mathbf{b}_0 = c \begin{pmatrix} \det(\mathbf{W}_1) \\ \vdots \\ \det(\mathbf{W}_N) \end{pmatrix}, \quad (\text{B8})$$

where c is a constant. Also note that $b_0^{(i)}(t) = \sum_{m \geq 0} P_i(m; t)$ according to the CM definition and $\sum_{i=1}^N \sum_{m \geq 0} P_i(m; t) \equiv 1$ for any $t \geq 0$ due to the conservative condition of the total probability. Therefore, we have the constraint condition $\mathbf{u}_N \mathbf{b}_0 = 1$, which in combination with (B8) gives

$$c = \frac{1}{\sum_{k=1}^N \det(\mathbf{W}_k)}. \quad (\text{B9})$$

If we denote by $f_{\mathbf{w}}(x) = \det(x\mathbf{I}_N - \mathbf{W}) = x(x + \alpha_1) \cdots (x + \alpha_{N-1})$ the characteristic polynomial of matrix \mathbf{W} , it is easy to show that

$$\left. \frac{df_{\mathbf{w}}(x)}{dx} \right|_{x=0} = \prod_{i=1}^{N-1} \alpha_i. \quad (\text{B10})$$

On the other hand, Jacobi's formula gives

$$\begin{aligned} \left. \frac{df_{\mathbf{w}}(x)}{dx} \right|_{x=0} &= \left. \frac{d \det(x\mathbf{I}_N - \mathbf{W})}{dx} \right|_{x=0} = \text{tr} \left((x\mathbf{I}_N - \mathbf{W})^* \frac{d(x\mathbf{I}_N - \mathbf{W})}{dx} \right) \Big|_{x=0} \\ &= (-1)^{N-1} \sum_{k=1}^N \det(\mathbf{W}_k) = \sum_{k=1}^N \prod_{i=1}^{N-1} \beta_i^{(k)}, \end{aligned} \quad (\text{B11})$$

where "tr" represents the matrix trace. The combination of (B10) and (B11) gives the equality

$$\sum_{k=1}^N \prod_{i=1}^{N-1} \beta_i^{(k)} = \prod_{i=1}^{N-1} \alpha_i. \quad (\text{B12})$$

Thus, we have the relationship

$$\sum_{k=1}^N \det(\mathbf{W}_k) = (-1)^{N-1} \prod_{i=1}^{N-1} \alpha_i \quad (\text{B13})$$

due to (B7). In addition, note that $c = \frac{(-1)^{N-1}}{\prod_{i=1}^{N-1} \alpha_i}$ due to $c = \frac{1}{\sum_{k=1}^N \det(\mathbf{W}_k)}$, and $b_0^{(k)} = c \cdot \det(\mathbf{W}_k)$ due to

$$\mathbf{b}_0 = c \begin{pmatrix} \det(\mathbf{W}_1) \\ \vdots \\ \det(\mathbf{W}_N) \end{pmatrix},$$

as well as $\det(\mathbf{W}_k) = (-1)^{N-1} \prod_{i=1}^{N-1} \beta_i^{(k)}$ due to the fact that $-\beta_i^{(k)}$ ($1 \leq i \leq N$) are the characteristics of matrix \mathbf{W}_k . Therefore, we have

$$b_0^{(k)} = c \cdot \det(\mathbf{W}_k) = (-1)^{N-1} \left(\prod_{i=1}^{N-1} \alpha_i \right)^{-1} (-1)^{N-1} \prod_{i=1}^{N-1} \beta_i^{(k)} = \prod_{i=1}^{N-1} \frac{\beta_i^{(k)}}{\alpha_i}, \quad 1 \leq k \leq N. \quad (\text{B14})$$

To that end, the proof of Proposition 1 in the main text is finished.

C. Derivation of (3.16) and (3.17) in the main text

For clarity, here we consider only the case of constant degradation rate (δ). First, we derive the analytical expression for the mRNA mean. According to (33), (34) and $\rho_u = \frac{\mu}{\delta} (1 - e^{-\delta u})$ as well as

definitions in the main text, we can respectively show

$$\begin{aligned} \langle m \rangle_{init} &= W'_{init}(0) = \frac{\int_{t=0}^{\infty} \rho_t f_{on}(t) dt}{1 - \int_{t=0}^{\infty} \int_{s=0}^{\infty} H(t) f_{on}(t) f_{off}(s) dt ds} = \frac{\mu}{\delta} \frac{\int_{t=0}^{\infty} (1 - e^{-\delta t}) f_{on}(t) dt}{1 - \int_{t=0}^{\infty} \int_{s=0}^{\infty} e^{-\delta(t+s)} f_{on}(t) f_{off}(s) dt ds} \\ &= \frac{\mu}{\delta} \frac{1 - \int_{t=0}^{\infty} e^{-\delta t} f_{on}(t) dt}{1 - \int_{t=0}^{\infty} e^{-\delta t} f_{on}(t) dt \int_{s=0}^{\infty} e^{-\delta s} f_{off}(s) ds} = \frac{\mu}{\delta} \frac{1 - L_{on}(\delta)}{1 - L_{on}(\delta)L_{off}(\delta)}, \end{aligned} \quad (C1)$$

$$\begin{aligned} \langle m^2 \rangle_{init} &= W''_{init}(0) = \langle m \rangle_{init} \frac{\int_{t=0}^{\infty} [H(s+t)(1-H(s+t)+2\rho_t)] f_{on}(t) f_{off}(s) dt ds}{1 - L_{on}(2\delta)L_{off}(2\delta)} \\ &+ \frac{\int_{t=0}^{\infty} \int_{s=0}^{\infty} \rho_t (1 + \rho_t) f_{on}(t) f_{off}(s) dt ds}{1 - L_{on}(2\delta)L_{off}(2\delta)} \\ &= \langle m \rangle_{init} \frac{\int_{t=0}^{\infty} \int_{s=0}^{\infty} [e^{-\delta(ts)} (1 - e^{-\delta(ts)} + 2\rho_t)] f_{on}(t) f_{off}(s) dt ds}{1 - L_{on}(2\delta)L_{off}(2\delta)} + \frac{\int_{t=0}^{\infty} \rho_t (1 + \rho_t) f_{on}(t) dt}{1 - L_{on}(2\delta)L_{off}(2\delta)} \\ &= \langle m \rangle_{init} \frac{\left(1 + \frac{2\mu}{\delta}\right) L_{on}(\delta)L_{off}(\delta) - L_{on}(2\delta)L_{off}(2\delta) - 2\frac{\mu}{\delta} L_{on}(2\delta)L_{off}(\delta)}{1 - L_{on}(2\delta)L_{off}(2\delta)} \\ &+ \frac{\frac{\mu}{\delta} \left(1 + \frac{\mu}{\delta}\right) - \frac{\mu}{\delta} \left(1 + \frac{2\mu}{\delta}\right) L_{on}(\delta) + \frac{\mu^2}{\delta^2} L_{on}(2\delta)}{1 - L_{on}(2\delta)L_{off}(2\delta)}, \end{aligned} \quad (C2)$$

$$\langle m \rangle_s = \frac{d}{dz} W_{init} \left(\log \left[1 + e^{-\delta s} (e^z - 1) \right] \right) \Big|_{z=0} = W'_{init}(0) e^{-\delta s} \quad (C3)$$

$$\langle m^2 \rangle_s = \frac{d^2}{dz^2} W_{init} \left(\log \left[1 + e^{-\delta s} (e^z - 1) \right] \right) \Big|_{z=0} = W''_{init}(0) e^{-2\delta s} + W'_{init}(0) (e^{-\delta s} - e^{-2\delta s}),$$

$$\begin{aligned} \langle m \rangle_u &= \frac{d}{dz} \int_{s=0}^{\infty} W_{init} \left(\log \left(1 + e^{-\delta(s+u)} (e^z - 1) \right) \right) e^{\rho_u(e^z-1)} f_{off}(s) ds \Big|_{z=0} \\ &= W'_{init}(0) L_{off}(\delta) e^{-\delta u} + \rho_u W_{init}(0), \end{aligned}$$

$$\begin{aligned} \langle m^2 \rangle_u &= \frac{d^2}{dz^2} \int_{s=0}^{\infty} W_{init} \left(\log \left(1 + e^{-\delta(s+u)} (e^z - 1) \right) \right) e^{\rho_u(e^z-1)} f_{off}(s) ds \Big|_{z=0} \\ &= \rho_u (1 + \rho_u) + W''_{init}(0) e^{-2\delta u} L_{off}(2\delta) \\ &+ W'_{init}(0) \left[e^{-\delta u} L_{off}(\delta) + 2\rho_u e^{-\delta u} L_{off}(\delta) - e^{-2\delta u} L_{off}(2\delta) \right]. \end{aligned} \quad (C4)$$

Substituting these expressions into (36) in the main text yields

$$\begin{aligned} (\langle \tau_{on} \rangle + \langle \tau_{off} \rangle) \langle m \rangle &= \int_{s=0}^{\infty} \langle m \rangle_s (1 - F(s)) ds + \int_{u=0}^{\infty} \langle m \rangle_u (1 - G(u)) du \\ &= \int_{s=0}^{\infty} W'_{init}(0) e^{-\delta s} \left(\int_{t=s}^{\infty} f_{off}(t) ds + \int_{u=0}^{\infty} [W'_{init}(0) L_{off}(\delta) e^{-\delta u} + \rho_u] \left(\int_{t=u}^{\infty} f_{on}(t) dt \right) du \right) ds \\ &= W'_{init}(0) \int_{s=0}^{\infty} e^{-\delta s} \left(\int_{t=s}^{\infty} f_{off}(t) dt \right) ds \\ &+ W'_{init}(0) L_{off}(\delta) \int_{u=0}^{\infty} e^{-\delta u} \left(\int_{t=u}^{\infty} f_{on}(t) dt \right) du + \int_{u=0}^{\infty} \rho_u \left(\int_{t=u}^{\infty} f_{on}(t) dt \right) du, \end{aligned}$$

where

$$\begin{aligned} \int_{s=0}^{\infty} e^{-\delta s} \int_s^{\infty} f_{off}(t) dt ds &= \frac{1}{\delta} - \frac{1}{\delta} L_{off}(\delta), \\ \int_{u=0}^{\infty} e^{-\delta u} \int_u^{\infty} f_{on}(t) dt ds &= \frac{1}{\delta} - \frac{1}{\delta} L_{on}(\delta), \\ \int_{u=0}^{\infty} \rho_u \left(\int_{t=u}^{\infty} f_{on}(t) dt \right) du &= \frac{\mu}{\delta} \int_{u=0}^{\infty} (1 - e^{-\delta u}) \int_u^{\infty} f_{on}(t) dt du \\ &= \frac{\mu}{\delta} \int_{u=0}^{\infty} \int_u^{\infty} f_{on}(t) dt du - \frac{\mu}{\delta} \int_{u=0}^{\infty} e^{-\delta u} \int_u^{\infty} f_{on}(t) dt du \\ &= \frac{\mu}{\delta} \langle \tau_{on} \rangle - \frac{\mu}{\delta^2} + \frac{\mu}{\delta^2} L_{on}(\delta). \end{aligned}$$

Therefore, we have

$$\begin{aligned} & \left(\langle \tau_{on} \rangle + \langle \tau_{off} \rangle \right) \langle m \rangle \\ &= W'_{init}(0) \left(\frac{1}{\delta} - \frac{1}{\delta} L_{off}(\delta) \right) + W'_{init}(0) L_{off}(\delta) \left(\frac{1}{\delta} - \frac{1}{\delta} L_{on}(\delta) \right) + \frac{\mu}{\delta} \langle \tau_{on} \rangle - \frac{\mu}{\delta^2} + \frac{\mu}{\delta^2} L_{on}(\delta) \quad (C5) \\ &= \frac{\mu}{\delta} \langle \tau_{on} \rangle - \frac{\mu}{\delta^2} + \frac{\mu}{\delta^2} L_{on}(\delta) + \frac{W'_{init}(0)}{\delta} \left(1 - L_{off}(\delta) L_{on}(\delta) \right) = \frac{\mu}{\delta} \langle \tau_{on} \rangle. \end{aligned}$$

This gives the analytical expression for the mean mRNA in the case that the on-state and off-state lifetimes follow general distributions, i.e., (3.16) in the main text.

Next, we derive the analytical expression for the mRNA noise intensity. The key is to give the second-order moment (m^2). Note that

$$\begin{aligned} \left(\langle \tau_{on} \rangle + \langle \tau_{off} \rangle \right) \langle m^2 \rangle &= \int_{s=0}^{\infty} \langle m^2 \rangle_s (1 - F(s)) ds + \int_{u=0}^{\infty} \langle m^2 \rangle_u (1 - G(u)) du \\ &= \int_{s=0}^{\infty} \left(\left[W''_{init}(0) e^{-2\delta s} + W'_{init}(0) (e^{-\delta s} - e^{-2\delta s}) \right] \int_{t=s}^{\infty} f_{off}(t) dt \right) ds \\ &+ \int_{u=0}^{\infty} \left(\left\{ \begin{aligned} & \rho_u (1 + \rho_u) + W''_{init}(0) e^{-2\delta u} L_{off}(2\delta) \\ & + W'_{init}(0) \left[e^{-\delta u} L_{off}(\delta) + 2\rho_u e^{-\delta u} L_{off}(\delta) - e^{-2\delta u} L_{off}(2\delta) \right] \end{aligned} \right\} \int_{t=u}^{\infty} f_{on}(t) dt \right) du \\ &= W''_{init}(0) \left\{ \int_{s=0}^{\infty} \left(e^{-2\delta s} \int_{t=s}^{\infty} f_{off}(t) dt \right) ds + L_{off}(2\delta) \int_{u=0}^{\infty} \left(e^{-2\delta u} \int_{t=u}^{\infty} f_{on}(t) dt \right) du \right\} \\ &+ W'_{init}(0) \left\{ \int_{s=0}^{\infty} \left((e^{-\delta s} - e^{-2\delta s}) \int_{t=s}^{\infty} f_{off}(t) dt \right) ds + L_{off}(\delta) \int_{t=0}^{\infty} \left(e^{-\delta u} \int_{t=u}^{\infty} f_{on}(t) dt \right) du \right. \\ &\left. + 2L_{off}(\delta) \int_{u=0}^{\infty} \left(\rho_u e^{-\delta u} \int_{t=u}^{\infty} f_{on}(t) dt \right) du - L_{off}(2\delta) \int_{u=0}^{\infty} \left(e^{-2\delta u} \int_{t=u}^{\infty} f_{on}(t) dt \right) du \right\} \\ &+ \int_{u=0}^{\infty} \left(\rho_u (1 + \rho_u) \int_{t=u}^{\infty} f_{on}(t) dt \right) du \end{aligned}$$

Simple calculation yields

$$\left(\langle \tau_{on} \rangle + \langle \tau_{off} \rangle \right) \langle m^2 \rangle = -\frac{\mu^2 (1 - L_{on}(\delta)) (1 - L_{off}(\delta))}{\delta^3 (1 - L_{on}(\delta) L_{off}(\delta))} + \frac{\mu}{\delta} \langle \tau_{on} \rangle + \frac{\mu^2}{\delta^2} \langle \tau_{on} \rangle \quad (C6)$$

Using (C1) and according to the noise definition, we can obtain the analytical expression for the mRNA noise intensity

$$\begin{aligned} \eta_m^2 &= \frac{1}{\langle m \rangle} + \frac{\langle \tau_{off} \rangle}{\langle \tau_{on} \rangle} - \frac{\delta (\langle \tau_{off} \rangle + \langle \tau_{on} \rangle)}{\mu^2 \langle \tau_{on} \rangle^2} \frac{\int_{s=0}^{\infty} \int_{u=0}^{\infty} \rho_s \rho_u f_{off}(s) f_{on}(u) ds du}{1 - \int_{s=0}^{\infty} \int_{u=0}^{\infty} e^{-\delta(s+u)} f_{off}(s) f_{on}(u) ds du} \\ &= \frac{1}{\langle m \rangle} + \frac{\langle \tau_{off} \rangle}{\langle \tau_{on} \rangle} - \frac{\langle \tau_{off} \rangle + \langle \tau_{on} \rangle}{\delta \langle \tau_{on} \rangle^2} \frac{(1 - L_{on}(\delta))(1 - L_{off}(\delta))}{1 - L_{on}(\delta)L_{off}(\delta)} \end{aligned} \quad (C7)$$

D. Queuing model simulation algorithm

In the queuing model of stochastic transcription, we assume that the synthesis rate and degradation rate of mRNA are μ and δ respectively. And the times that the gene dwells at ON and OFF states are assumed to follow general distributions $f_{on}(t)$ and $f_{off}(t)$ respectively. Let $g(t)$ be gene state and $m(t)$ be the mRNA number in the cell at time t . If the gene dwells at ON state, $g(t) = 1$, otherwise $g(t) = 0$. We propose the following stochastic simulation algorithm to obtain exact numerical realizations of the mRNA synthesis and degradation process $m(t)$.

0. Initialize the time $t = t_0$ and the system's state $m(t) = m_0$.

1. if $g(t) = 1$

generate a waiting time τ_{on} according to probability $f_{on}(t)$

set $\tau_{on}^{end} = t + \tau_{on}$

repeat

update the mRNA's state $m(t)$ and based on Gillespie simulation algorithm [60] for reactions $\emptyset \rightarrow \text{mRNA}$ and $\text{mRNA} \rightarrow \emptyset$ whose reaction propensity functions are given by $a_1(m(t)) = \mu$ and $a_2(m(t)) = \delta m(t)$ respectively

until $t \geq \tau_{on}^{end}$

set $g(t) = 0$

else

generate a waiting time τ_{off} according to probability $f_{off}(t)$

set $\tau_{off}^{end} = t + \tau_{off}$

repeat

update the mRNA's state $m(t)$ and based on Gillespie simulation algorithm [60] for reaction $\text{mRNA} \rightarrow \emptyset$ whose reaction propensity function is given by $a_2(m(t)) = \delta m(t)$

until $t \geq \tau_{off}^{end}$

set $g(t) = 1$

end

2. Record trajectories $(t, g(t), m(t))$ as desired. Return to Step 1, or else end the simulation.

3. After discarding the early part of the trajectories (the burn-in phase), the remaining realizations are used to estimate the steady-state probability distribution and the related statistics.

E. Proof of Theorem 3 in the main text

Note that for any m , we have

$$P(m+1) - P(m) = \frac{\Gamma(\lambda+m)\mu^m e^{-\mu}}{m!\Gamma(\lambda+\gamma+m)} \frac{\Gamma(\lambda+\gamma)}{\Gamma(\lambda)} \times \left[\frac{\mu(\lambda+m)}{(m+1)(\lambda+\gamma+m)} {}_1F_1(\gamma, \lambda+\gamma+m+1; \mu) - {}_1F_1(\gamma, \lambda+\gamma+m; \mu) \right] \quad (E1)$$

and

$$F_1(\gamma, \lambda+\gamma+m+1; \mu) - F_1(\gamma, \lambda+\gamma+m; \mu) < 0 \quad (E2)$$

For clarity, we distinguish the following three cases.

First, if $\mu\lambda \leq \lambda + \gamma$, $\mu(\lambda+m) \leq (m+1)(\lambda+\gamma+m)$ holds for $m=0$, implying that $P(1) < P(0)$ is guaranteed.

Second, note that if $\lambda \geq 1$, the function $f_1(x) = \frac{\mu(\lambda+x)}{(x+1)(\lambda+\gamma+x)}$ has no extreme value with regard to x , implying that it is monotonically decreasing. Therefore, the inequality $u(\lambda+m) \leq (m+1)(\lambda+\gamma+m)$ holds for all m in this case. Thus, if $\lambda \geq 1$ and $\mu\lambda \leq \lambda + \gamma$, $P(m)$ is unimodal.

Third, if $\lambda < 1$, the function $f_1(x)$ reaches its maximum at the unique point $\tilde{x} = -\lambda + \sqrt{\gamma(1-\lambda)}$. Furthermore, if $u(\lambda+\tilde{x}) \leq (\tilde{x}+1)(\lambda+\gamma+\tilde{x})$, $f_1(\tilde{x}) \leq 1$, implying that $P(m+1) < P(m)$ holds for all m . Therefore, $P(m)$ is also unimodal in this case. However, $\mu(\lambda+\tilde{x}) \leq (\tilde{x}+1)(\lambda+\gamma+\tilde{x})$ is equivalent to $(\mu+\lambda-\gamma-1)\sqrt{(1-\lambda)\gamma} \leq 2(1-\lambda)\gamma$, which is further equivalent to $\mu \leq \gamma+1-\lambda+2\sqrt{(1-\lambda)\gamma}$. Thus, we know that if two inequalities $\lambda < 1$ and $\mu \leq \min(\gamma+1-\lambda+2\sqrt{(1-\lambda)\gamma}, 1+\gamma/\lambda)$ are simultaneously satisfied, $P(m)$ is unimodal, referring to Figure 4(A) and (B) in the main text.

Now, we turn to considering bimodality. It is not difficult to establish the inequality

$${}_1F_1(\gamma, \lambda+\gamma+m+1; \mu) > \xi_1 {}_1F_1(\gamma, \lambda+\gamma+m; \mu) \quad (E3)$$

for any m , where $\xi_1 = 1 - \frac{\mu}{(\sqrt{\lambda+m} + \sqrt{\lambda+m+1})^2}$. According to the expression of $P(m) = \frac{\Gamma(\lambda+m)e^{-\mu}\mu^m}{m!\Gamma(\lambda+\gamma+m)} \frac{\Gamma(\lambda+\gamma)}{\Gamma(\lambda)} {}_1F_1(\gamma, \lambda+\gamma+m; \mu)$, we have $\frac{P(m+1)}{P(m)} > \xi_2 \frac{(\lambda+m)\mu}{(m+1)(\lambda+\gamma+m)}$, where $\xi_2 = 1 - \frac{\mu}{4(\lambda+m)}$. Thus, in order to ensure that $P(m)$ is bimodal, we only need to find a $m_1 > 1$ such that $\left(1 - \frac{\mu}{4(\lambda+m_1)}\right) \frac{(\lambda+m_1)\mu}{(m_1+1)(\lambda+\gamma+m_1)} > 1$. However, this inequality can be rewritten as

$$4m_1^2 - 4(\mu - \lambda - \gamma - 1)m_1 + \mu^2 - 4\lambda\mu + 4(\lambda + \gamma) < 0 \quad (E4)$$

Therefore, if $\mu \geq \lambda + \gamma + 2$ and $(\mu - \lambda - \gamma - 1)^2 \geq \mu^2 - 4\lambda\mu + 4(\lambda + \gamma)$ (the latter is equivalent to $2(\gamma + 1 - 2\lambda)\mu \leq (\lambda + \gamma - 1)^2$) are simultaneously satisfied, we know that there is at least one such that

$$1 < m_1 < \frac{1}{2} \left(\mu - \lambda - \gamma - 1 + \sqrt{(\mu - \lambda - \gamma - 1)^2 - \mu^2 + 4\lambda\mu - 4(\lambda + \gamma)} \right) \quad (E5)$$



AIMS Press

©2019 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)