

# Mathematical Theory to Compute Stochastic Cellular Processes

Keita Iida and Yoshitaka Kimura

**Abstract** A central challenge of gene expression analysis during the last few decades has been the characterization of the expression patterns experimentally and theoretically. Modern techniques on single-cell and -molecule resolution reveal that transcriptions and translations are stochastic in time and that clonal population of cells displays heterogeneity in the abundance of a given RNA and protein per cell. Hence, to take into account a cell-to-cell variability, we consider a stochastic model of transcription and the chemical master equation. Our stochastic analysis and Monte-Carlo simulation show that the limiting distribution of mRNA copy number can be expressed by a Poisson-beta distribution. The distribution represents the four different types of expression patters, which are typically found in various experimental profiles.

**Keywords** Gene expression · Single-cell analysis · Biological noise · Markov process · Master equation · Poisson-beta distribution

## 1 Introduction

Gene expression analysis is being used to investigate the functions of gene products (RNA and protein), to improve our understanding of various aspects of cellular function and disease, and to facilitate drug development [1]. Expression analysis has ever revealed the key regulators for various cell differentiations, which may help scientists establish novel cells [2]. However, little is known about the regulatory mechanisms of dynamic gene expressions. Although many biological processes, such as transcription factors binding, chromatin remodeling and cell cycle, have been reported as the important factors, a systematic understanding of unidirectional

---

K. Iida · Y. Kimura (✉)  
Tohoku University, 1-1 Seiryomachi, Aoba-ku, Sendai, Miyagi 980-8574, Japan  
e-mail: ykimura@med.tohoku.ac.jp

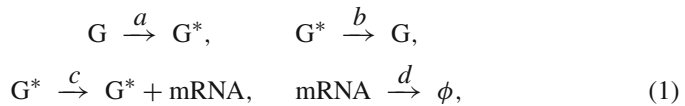
K. Iida  
e-mail: kiida1984@med.tohoku.ac.jp

cell differentiations remains to be acquired. Today systems biology gives us a novel methodology to systematically understand the complex intracellular dynamics [3].

Modern techniques on single-cell and -molecule resolution reveal that transcriptions and translations are stochastic in time and that clonal population of cells displays heterogeneity in the abundance of a given RNA and protein per cell [4–7]. Thus, expression analysis based on probability statistics becomes an indispensable tool today [8], which may shed light on the classical biological knowledge [9]. In the present article, we mathematically investigate a variety of expression patterns by analyzing a simple model of transcription. Then, we discuss a reduction of the model equation, which is the key step to make gene regulatory networks [10].

## 2 Mathematical Model

Based on previous papers [11], we consider the following mathematical model for a single gene expression induced by a transcription factor:



where  $G$  and  $G^*$  denote the genes being ‘off’ and ‘on’ states, respectively, and  $\phi$  the degraded mRNA. Here, we assume that the transcription event can only occur under the ‘on’ state [12]. The parameters  $a$  and  $b$  are the probabilities per unit time of the promoter switching from inactive to active and active to inactive, respectively,  $c$  and  $d$  are the probabilities per unit time of transcription and mRNA degradation, respectively.

We assume that the time evolution of the mRNA copy number is modeled by a simple Markov process in continuous time, of which the state space is defined as

$$S = \{(i, n) \mid i \in \{0, 1\}, n \in \mathbb{Z}_{\geq 0}\},$$

where  $i = 0$  and  $1$  are for ‘off’ and ‘on’ states, respectively, and  $n$  denotes the mRNA copy number in the system. Let  $P_n^{(i)}(t)$  be the probability of having  $(i, n)$  state at time  $t$ , which obeys the following master equation:

$$\frac{d}{dt} \begin{bmatrix} P_n^{(0)} \\ P_n^{(1)} \end{bmatrix} = \mathbf{A} \begin{bmatrix} P_n^{(0)} \\ P_n^{(1)} \end{bmatrix} + \mathbf{C} \begin{bmatrix} P_{n-1}^{(0)} - P_n^{(0)} \\ P_{n-1}^{(1)} - P_n^{(1)} \end{bmatrix} + \mathbf{D} \begin{bmatrix} (n+1)P_{n+1}^{(0)} - nP_n^{(0)} \\ (n+1)P_{n+1}^{(1)} - nP_n^{(1)} \end{bmatrix}, \quad (2)$$

where  $A = \begin{bmatrix} -a & b \\ a & -b \end{bmatrix}$ ,  $C = \begin{bmatrix} 0 & 0 \\ 0 & c \end{bmatrix}$  and  $D = \begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix}$ . The initial condition is

$$(P_n^{(0)}(0), P_n^{(1)}(0)) = (\delta_{0,n}, 0), \tag{3}$$

where  $\delta$  is the Kronecker delta.

### 3 Analysis

The limiting distribution  $\bar{P}_n$  of the system (2) and (3) becomes as follows:

$$\bar{P}_n = \frac{\gamma^n}{n!} e^{-\gamma} \frac{(\alpha)_n}{(\alpha + \beta)_n} {}_1F_1(\beta, \alpha + \beta + n; \gamma), \tag{4}$$

where we define  $\alpha = a/d$ ,  $\beta = b/d$  and  $\gamma = c/d$ . Here,  $(\cdot)_n$  is the Pochhammer symbol and  ${}_1F_1$  is the Kummer function. As indicated in an earlier paper [13], (4) can be further simplified as follows:

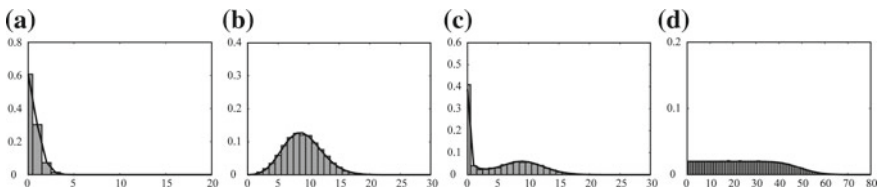
$$\bar{P}_n = \int_0^1 \frac{(\gamma x)^n}{n!} e^{-\gamma x} \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} dx, \tag{5}$$

where  $B$  is the beta function. The distribution (5), which is called a Poisson-beta distribution, shows that the transcription rate can be regarded as  $\gamma x$  in which  $x$  follows the beta distribution with parameters  $\alpha$  and  $\beta$ . Hence, in the long-time limit, the model (1) can be approximated by the following scheme:



where the stochastic variable  $X$  follows the beta distribution  $B(x; \alpha, \beta)$ .

Figure 1a–d shows the limiting distribution (4) with various parameter sets. As one can see in Fig. 1, the expression patterns widely change depending on the para-



**Fig. 1** The limiting distribution with respect to the mRNA copy number obtained from (2) and (3). The exact solutions (*bold line*) are obtained from (4) and numerical solutions (*filled bar graph*) from Monte-Carlo simulation with  $\Delta t = 0.1$ . The parameters  $(\alpha, \beta, \gamma)$  are **a** (50, 50, 1), **b** (1, 10, 10), **c** (0.1, 0.1, 10), **d** (1, 1, 50)

meters  $\alpha$ ,  $\beta$  and  $\gamma$ . From the analytical result (5), we found that the beta distribution produces the variation and the Poisson distribution guarantees the discreteness of mRNA molecules.

## 4 Conclusion

Mathematical models of gene regulation have been studied since 1960s [14–16]. However, the classical deterministic approaches based on the population-wide average methods, such as the statistical procedure and the modeling with ordinary differential equations, are not enough to understand cell-to-cell variability. To understand the mechanisms of cell-to-cell variation in gene expressions, we should consider intrinsic and extrinsic noises (‘biological noise’) when constructing a mathematical model [4]. In the present article, we considered a simple model of transcription with only two gene states (‘on’ and ‘off’) and investigated the probability distribution of mRNA copy number. We found that the limiting distribution can be described by the Poisson-beta distribution, which represents four different types of expression patterns Fig. 1. Thus, the classical model (1) can be approximated by the scheme (6) in the long-time limit.

**Acknowledgments** This work was supported by Tohoku University’s Juten Senryaku Program.

## References

1. Holstege, F.C.P., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., Young, R.A.: *Cell* **95**, 717–728 (1998)
2. Takahashi, K., Yamanaka, S.: *Cell* **126**, 663–676 (2006)
3. Kitano, H.: *Science* **295**, 1662–1664 (2002)
4. Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S.: *Science* **297**, 1183–1186 (2002)
5. Golding, I., Paulsson, J., Zawilski, S.M., Cox, E.C.: *Cell* **123**, 1025–1036 (2005)
6. Chubb, J.R., Trcek, T., Shenoy, S.M., Singer, R.H.: *Curr. Biol.* **16**, 1018–1025 (2006)
7. Sanchez, A., Golding, I.: *Science* **342**, 1188–1193 (2013)
8. Burrage, K., Hegland, M., Macnamara, S., Sidje, R.B.: *Markov Anniversary Meeting: an International Conference to Celebrate the 150th Anniversary of the Birth of A.A. Markov*, Bosc Books, pp. 21–38 (2006)
9. Selvarajoo, K.: *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4**, 385–399 (2012)
10. Hegland, M., Burden, C., Santoso, L., MacNamara, S., Booth, H.: *J. Comput. Appl. Math.* **205**, 708–724 (2007)
11. Peccoud, J., Ycart, B.: *Theor. Popul. Biol.* **48**, 222–234 (1995)
12. Golding, I., Cox, E.C.: *Curr. Biol.* **16**, R371–R373 (2006)
13. Kim, J.K., Marioni, J.C.: *Genome Biol.* **14**, 1–12 (2013)
14. Monod, J., Jacob, F.: *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401 (1961)
15. Simon, Z.: *J. Theor. Biol.* **8**, 258–263 (1965)
16. Griffith, J.S.: *J. Theor. Biol.* **20**, 202–208 (1968)