Krynica, 14th–18th September 2010

PROBABILITY DENSITY FUNCTIONS IN BISTABLE GENE ACTIVATION MODEL WITH TWO TYPES OF NOISE

Joanna Jaruszewicz[1], **Paweł Żuk**[2], **Tomasz Lipniacki** [1]

[1]Institute of Fundamental Technological Research Polish Academy of Sciences
Pawińskiego 5B, 02-106 Warszawa
[2] College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences
University of Warsaw
[1]jjarusz@ippt.gov.pl, tlipnia@ippt.gov.pl, [2]pzuk@ippt.gov.pl

ABSTRACT

The aim of this study is to demonstrate that in dynamical systems with underlying bistability the type of noise qualitatively influences the stationary probability distribution (SPD). Specifically, we consider a simplified model of gene expression with the nonlinear positive feedback, which in the deterministic approximation has two stable steady state solutions. Two types of noise are considered; transcriptional - due to the limited number of protein molecules, and gene switching noise - due to gene activation and inactivation. In the limit of zero noise, the SPD generically concentrates in the decreasing vicinity of one of the two stable steady states. We demonstrated that for a range of parameters the SPD corresponding to the system with transcriptional noise only concentrates around a different steady state than SPD corresponding to the system with gene switching noise only.

INTRODUCTION

Intracellular regulatory processes can be described in terms of stochastic dynamical systems. The stochasticity arises due to the limited number of reacting molecules such as DNA, mRNA or proteins. In many cases for the sake of simplicity the stochasticity is neglected and the deterministic approximations of the stochastic systems are considered. It is thus important to know what is the correspondence between the stochastic system and its deterministic approximation at least in the zero noise limit. In this study we focus on the correspondence between the stable steady state solutions of the deterministic approximation and stationary probability distributions (SPD) of stochastic system with underlying bistability.

In particular, we consider a simple model of gene expression with two types of noise: transcriptional - due to the limited number of protein molecules, and gene switching noise - due to gene activation and inactivation. We assume that the gene is regulated by its own product, which leads to non-linear positive feedback and introduces bistability (biological meaning and basic concepts for modelling switch-like behavior are shown in [3,4,9,10]). The model defines a time-continuous Markov process. In the limit of zero noise the corresponding SPD distribution concentrates in the decreasing vicinity of one of the two stable steady states. We aim to demonstrate that the relative magnitudes of the two considered noises dictate in which of the two stable steady states the SPD concentrates.

MODEL

We consider the stochastic model of gene expression with autoregulation in the so called Kepler-Elston approximation, which assumes that the protein is synthesized directly from the gene [7]. Such an approximation is justified when the mRNA degradation rate is larger than the gene activation and inactivation rates. We assume that the gene may be in only one of two states – active or inactive. The protein is synthesized with the constant rate Q when the gene is active and is degraded with rate r; we chose time units in which $r = 1$. The autoregulation arises when gene activation and inactivation rates ($c(Y)$ and $b(Y)$) depend on the level of synthesized protein Y. Here, we focus on the case with positive feedback and assume that $c(Y) = c_0 + c_2 Y^2/Q^2$, $b(Y) = b_0$ and $c_0, c_2, b_0 > 0$. The gene autoregulation model is illustrated on Fig. (1).

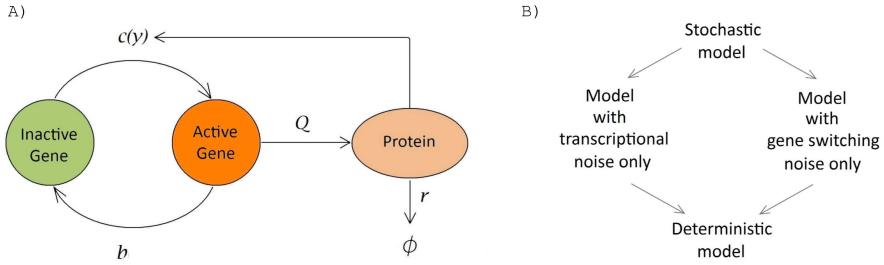

Figure 1. Model of gene expression with autoregulation. A) Cartoon of the model B) Exact (stochastic) model and its three approximations.

• **Exact stochastic model.** The model defines a time continuous Markov process with an infinite number of states described by two random variables: gene state $S(t) \in \{0, 1\}$ and number of protein molecules $Y(t) \in \mathbb{N}$. Resulting transition propensities are

$$\begin{cases} S = 0 \to S = 1 & c(Y), \\ S = 1 \to S = 0 & b_0, \\ Y = n \to Y = n + 1 & QS, \\ Y = n \to Y = n - 1 & n. \end{cases} \tag{1}$$

The system can be described by a countable set of ordinary differential equations (Master's Equations). Let g_n denote the probability that $\{S, Y\} = \{1, n\}$ and h_n denote the probability that $\{S, Y\} = \{0, n\}$, then

$$\begin{cases} \frac{dg_n}{dt} = Q(g_{n-1} - g_n) + (n+1)g_{n+1} - ng_n + c(n)h_n - b_0 g_n, & \text{for} \quad n > 0 \\ \frac{dh_n}{dt} = (n+1)h_{n+1} - nh_n - c(n)h_n + b_0 g_n, \\ \frac{dg_0}{dt} = -Qg_0 + g_1 + c(0)h_0 - b_0 g_0, \\ \frac{dh_o}{dt} = h_1 - c(0)h_0 + b_0 g_0. \end{cases} \tag{2}$$

The above system in stationary case can be solved using a moment generation function for $c(n) = c_0 + c_1 y$ using the method proposed in [6], however in our case $c(n) = c_0 + c_2 y^2$ it leads to third order ordinary differential equation, we failed to solve. We will thus estimate marginal SPD $f_n = g_n + h_n$ corresponding to exact model by Monte Carlo simulations of the system (1).

• **Model with gene switching noise only (Continuous model).** Such a model is a good approximation, when the characteristic number of protein molecules is very large. In such case we may consider $y = Y/Q$ as a continuous variable which follows

$$\frac{dy}{dt} = S - y, \tag{3}$$

where S, as in the exact model, is given by the process (1). System (1)-(3) defines time continuous piece-wise deterministic Markov process (see [2]). In this approximation $g_n(t)$, $h_n(t)$ are replaced by the continuous probability density functions $g(y,t)$, $h(y,t)$, that satisfy

$$\frac{\partial g}{\partial t} - \frac{\partial}{\partial y}(yg) = b_0 h - c(y)g, \tag{4}$$

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial y}((1-y)h) = -b_0 h + c\,(y)\,g. \tag{5}$$

The above system, as shown in [5], has the following stationary solution

$$g(y) = Exp\left[\int_0^y \left(\frac{-b_0}{(1-s)} + \frac{c(y)-1}{s}\right) ds\right], \qquad h(y) = \frac{yg(y)}{(1-y)}. \tag{6}$$

For $c(y) = c_0 + c_2 y^2$ marginal SPD $f(y)$ takes the form:

$$f(y) = g(y) + h(y) = Ce^{\frac{1}{2}c_2 y^2} y^{c_0-1}(1-y)^{b_0-1}, \tag{7}$$

where C is such that $\int_0^1 f(y) = 1$.

• **Model with transcriptional noise only (Discrete model).** This approximation is justified when transition rates $c(n)$ and b_0 are much larger than unity. In such case S may be replaced (see [1]) by its expected value $S = S(n) = c(n)/(c(n) + b_0)$. In this approximation the stationary marginal probabilities F_n are given by recurrence providing the analytical solution:

$$\begin{cases} F_1 = \frac{F_0 Q c_0}{br}, \\ F_{n+1} = \frac{1}{n+1}(F_n(n + \frac{Q}{r}S(n)) - F_{n-1}\frac{Q}{r}S(n-1)) \quad \text{for} \quad n > 0. \end{cases} \tag{8}$$

Since the recurrence (8) is linear with respect to F_0, we can set $F_0 = 1$, calculate all the F_n, and then normalize them by dividing by $\sum F_n$. Having analytical recurrence formula, for rational parameters we are able to calculate the precise value for each element. The coefficient $\epsilon = 1/Q$ will be considered as a measure of transcriptional noise.

• **Deterministic model.** Such model arises when the transition rates $c(n)$ and b_0 are much larger than unity, and at the same time the characteristic number of protein molecules is very large. In such case the protein level y is given by a single ordinary differential equation

$$\frac{dy}{dt} = S(y) - y, \quad \text{where} \quad S(y) = \frac{c(y)}{c(y) + b_0}. \tag{9}$$

The stationary solutions of the equation (9) are the real roots of the third order polynomial

$$W = -c_2 y^3 + c_2 y^2 - (c_0 + b_0)y + c_0 = 0. \tag{10}$$

Here, we focus on the bistable case when W has 3 real roots such that $0 < y_1 < y_2 < y_3 < 1$. The bistability domain in $(c_0/b_0, c_2/b_0)$ parameter space is shown in Fig. 2. Steady states y_1 and y_3 are stable, while y_2 is unstable. Due to the fact that W has the same coefficient at third and second power, its roots satisfy $y_1 + y_2 + y_3 = 1$. The original coefficients b_0, c_0, c_2 may be recovered from the roots by the following relations:

$$c_0 = \frac{b_0 y_1 y_2 y_3}{y_1(y_2+y_3) + y_2 y_3(1-y_1)}, \qquad c_2 = \frac{b_0}{y_1(y_2+y_3) + y_2 y_3(1-y_1)}. \tag{11}$$

Due to relation $y_1 + y_2 + y_3 = 1$ the (y_1, y_2, y_3) parameter space may be reduced to domain $D = (y_1, y_2)$ in which $y_1 < y_2$ and $1 - y_1 - y_2 = y_3 > y_2$, see Fig. 3. The coefficient $\delta = 1/b_0$ will be considered as a measure of gene switching noise.

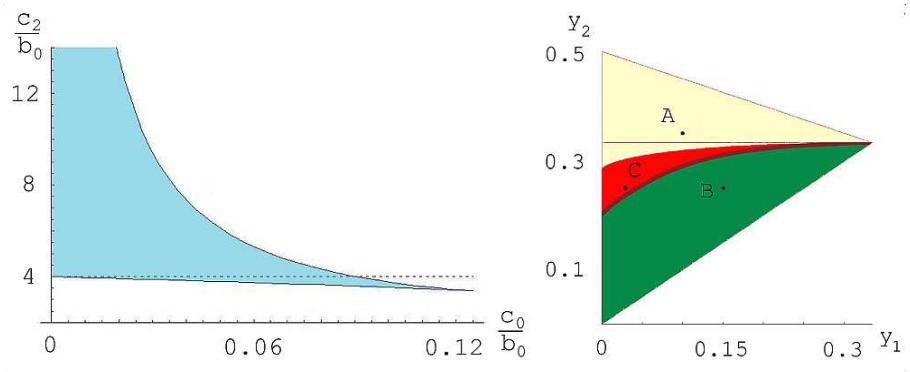

Figure 2. A) Bistability domain in $(c_0/b_0, c_2/b_0)$ parameter space. B) Bistability domain in (y_1, y_2) space. In the subdomain coloured yellow the SPD of both models concentrates around y_1, in green subdomain SPD of both models concentrates around y_3. In the red subdomain SPD of continuous model concentrates in y_3 while SPD of discrete model concentrates in y_1

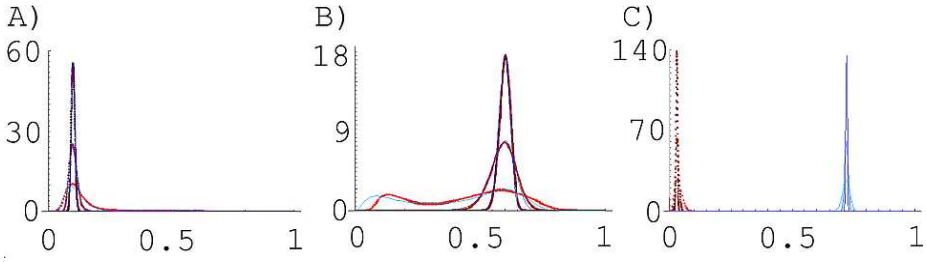

Figure 3. Stationary probability distributions; Discrete model: red points, Continuous model: blue lines. Panels A, B, C corresponds to points A, B, C, see Fig. 2B. Panel A: $\epsilon = 1/200, 1/1000, 1/2000$, $\delta = 1/150, 1/800, 1/4050$ Panel B: $\epsilon = 1/200$, $1/1000, 1/2000$, $\delta = 1/40, 1/180, 1/800$. Panel C: $\epsilon = 1/200, 1/1000, 1/2000$, $\delta = 1/600, 1/3000, 1/14500$.

RESULTS

To estimate SPD for exact model we performed long-run Monte Carlo simulations of the system (1) based on the Gillespie algorithm.

For the continuous model, based on the analytical solution, Eq. (7), we split domain $D = [0,1] \times [0,1] \ni (y_1, y_2)$ into two subdomains D_{A1}, D_{A3} (see Fig. 2B) such that $(y_1, y_2) \in D_{A1}$, $f(y)$ concentrates in the decreasing vicinity of y_1, while for $(y_1, y_2) \in D_{A3}$, $f(y)$ concentrates in the decreasing vicinity of y_3 in the limit of $\delta \to 0$. The line separating domains D_{A1}, D_{A3} is given in the implicit form:

$$(\frac{1-y_1}{y_1+y_2})(\frac{y_1}{1-y_1-y_2})^{p_1} e^{p_2} = 1, \tag{12}$$

where $p_1 = \dfrac{y_1 y_2 (1-y_1-y_2)}{(1-y_1)(1-y_2)(y_1+y_2)}$ and $p_2 = \dfrac{2y_1+y_2-1}{2(1-y_1)(y_1+y_2)}$.

For the discrete model, based on recurrence, Eq. (8), we split domain $D = [0,1] \times [0,1] \ni (y_1, y_2)$ into two subdomains D_{B1}, D_{B3} (see Fig. 2B) such that for $(y_1, y_2) \in D_{B1}$, $f(y)$ concentrates in the decreasing vicinity of y_1, while for $(y_1, y_2) \in D_{B3}$, $f(y)$ concentrates in the decreasing vicinity of y_3 in the limit of $\epsilon \to 0$. This result is based on numerics; we analyzed analytically calculated SPD for decreasing ϵ down to $\epsilon = 1/20000$.

Interestingly domains D_{A1}, D_{B1} are different, *i.e.* there exist non-empty domain $D_{AB} = D_{A3} \cap D_{B1}$. In domain D_{AB} SPD of continuous model concentrates around y_1, while in the discrete model SPD concentrates around y_3 (see Fig. (2B)). For further analysis we chose three sets of roots:

- $A = \{0.1, 0.35\} \in D_{A1} \cap D_{B1}$,
- $B = \{0.15, 0.25\} \in D_{A3} \cap D_{B3}$,
- $C = \{0.03, 0.25\} \in D_{AB}$;

Fig.(3) illustrates the SPD calculated for these points, for three values of ϵ and three values of δ.

Finally, we analyze the exact model with different ratio ϵ/δ, see Fig. 4. This analysis shows that in the exact model not only the parameters of the system, but also the relative magnitudes of two types of noise, influence the limit behavior for decreasing noise.

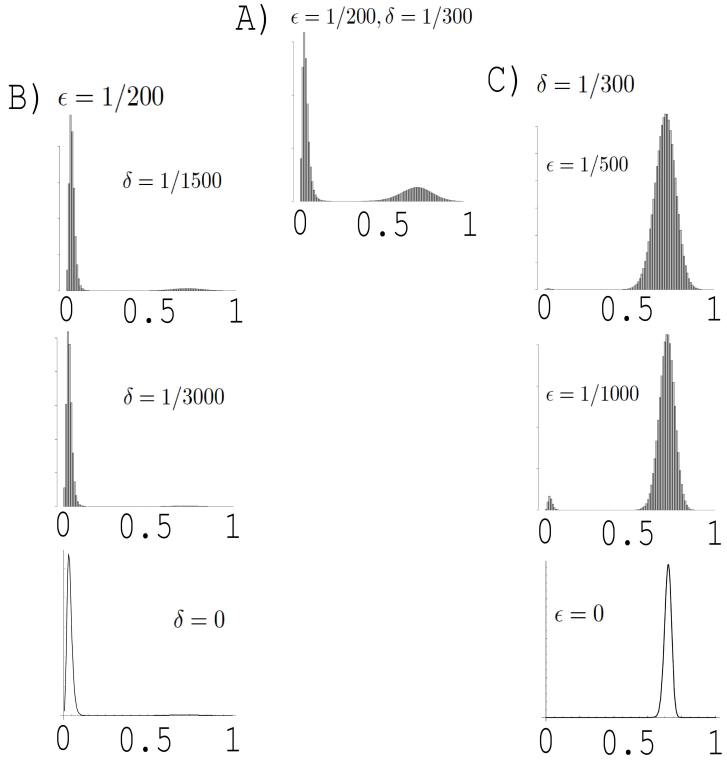

Figure 4. Approximation of SPD calculated for point C from the red subdomain, see Fig. 2B. Panel A: $\epsilon = 1/200$, $\delta = 1/300$. Panel B: constant ϵ and decreasing δ (gene switching noise) Panel C constant δ and decreasing ϵ (transcriptional noise) .

CONCLUSIONS

In this study we consider a stochastic model of gene expression with two types of noise; transcriptional - due to the limited number of protein molecules and gene-switching noise - due to gene activation and inactivation. Due to the nonlinear positive feedback the deterministic approximation of the considered model exhibits bistability. We analyzed two sub-models, each having only one type of noise to find that SPDs corresponding two both sub-models (as well as SPD corresponding to exact model with two types of noise) concentrate generically in the vicinity of one of the two stable stationary points of the limiting deterministic model. However, for a significant

range of parameters the two SPDs corresponding to the two sub-models are qualitatively different, in particular they may concentrate in the vicinities of different stationary points. In the exact model the choice of a stationary point around which the SPD concentrates is defined by the relative magnitude of two noises. Similar effect for stochastic dynamics in evolutionary games was found in [8], where the choice of Nash equilibria depends on the assumed type of noise.

Our finding demonstrates that in systems with underlying bistability, in the zero noise limit in which SPD generically converges to delta distribution in one of the stable points, the choice of a particular fixed point depends on the type of noise. Since all regulatory networks operating in cells with finite numbers of molecules have noise inherently embedded, an analysis of the exact stochastic process may be necessary to predict system's dynamics even in the qualitative way.

ACKNOWLEDGMENTS

This study was supported by Polish Committee for Scientific Research grant N N501 132936 and Foundation for Polish Science grant TEAM/2009-3/6.

REFERENCES

[1] A. Bobrowski: *From covergence of operator semigroups to gene expression, and back again*, Statistical Models in Biological Sciences, Banach Center Publications **80** (2008).

[2] A. Bobrowski, T. Lipniacki, K. Pichór, and R. Rudnicki: *Asymptotic behavior of distributions of mRNA and protein levels in a model of stochastic gene expression*, J. Math. Anal. Appl. **333** (2007), 753-769.

[3] A. Chatterjee, Y.N. Kaznessis, and W.-S. Hu: *Tweaking biological switches through a better understanding of bistability behaviour*, Current Opinion in Biotechnology **19** (2008), 475-481.

[4] J.L. Cherry and F.R. Adler: *How to make a Biological Switch*, J. Theor. Biol. **203** (2000), 117-133.

[5] B. Hat, P. Paszek, M. Kimmel, K. Piechór, and T. Lipniacki: *How the number of allels influences gene expression*, J. Stat. Phys. (2007).

[6] J.E.M. Hornos, D.Schultz, G.C.P. Innocenti, J. Wang, A.M. Walczak, J.N. Onuchic, and P.G. Wolynes: *Self-regulating gene: An exact solution*, Phys. Rev. E 72, 051907 (2005).

[7] T.B. Kepler and T.C. Elston: *Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations*, Biophys. J. **81** (2001), 3116-3136.

[8] J. Miękisz: *Equilibrium selection in evolutionary games with random matching of players*, J. Theor. Biol. **232** (2005), 47-53.

[9] A. Ochab-Marcinek: *Extrinsic noise passing through a Michaelis-Menten Reaction: A universal response of a genetic switch*, J. Theor. Biol. **263** (2010), 510-520.

[10] D. Siegal-Gaskins, E. Grotewold, and G.D. Smith: *Research article: The capacity for multistability in small gene regulatory networks*, BMC Sys. Biol. **3** (2009).