Krynica, 14th–18th September 2010

PROBABILITY DENSITY FUNCTIONS IN BISTABLE KINASE ACTIVATION MODEL

Paweł Żuk[1]**,Tomasz Lipniacki** [2]

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

ABSTRACT

We consider a kinase auto-activation model in which the number of activated kinases follows the time-continuous Markov process. In the deterministic approximation the process is described by the single non-linear ordinary differential equation, which may have two stable steady states. We found that for sufficiently large number of kinases, the stationary probability distribution given by the Markov process concentrates in the vicinity of the two stable steady states of the deterministic approximation. However, if the number of kinases diverges to the infinity (zero noise limit), the stationary probability distribution concentrates (generically) in only one of the two steady states.

INTRODUCTION

Kinases are the key proteins mediating intracellular signalling. They are activated by receptors or other kinases due to phosphorylation at serine, theorine or tyrosine residues, and inactivated (dephosphorylated) by the enzymes called phosphatases. Relevant to the considered model, kinases can form dimers and can be auto-activated - *i.e.* activated by the same kinase species.

In this study we consider a simple kinase activation model with bistable behavior and analyze the three "variants" of probability distribution (PD):

(1) stationary probability distribution (SPD),
(2) cumulative probability distribution (CPD) corresponding to the Markov process with given initial condition $K(0)$ lasting from time 0 to T.
(3) population probability distribution (PPD), which is the CPD but with the random initial condition.

We analyze the correspondence between these three variants of the PD, focusing on the convergence of the SPD in the "zero noise limit". The zero noise limit will be analyzed numerically by introduction of a class of models with a growing number of kinase molecules.

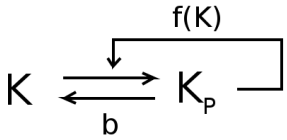

Figure 1. Kinase activation model with auto-regulation.

MODEL

The kinase activation model is illustrated on Fig. (1).

The model defines a time continuous Markov process which has $N + 1$ states, were N is the total number of kinases. States will be indexed by the number of active kinases molecules K. We assume that kinases are inactivated with constant rate $b > 0$ and activated with rate $f(K, N)$. Thus the transition propensities are

$$\begin{cases} K \to K + 1 & f(K, N)(N - K), \\ K \to K - 1 & bK. \end{cases} \tag{1}$$

We will consider two cases: without and with auto-regulation. In the former we set $f(K, N) = c > 0$, in the latter we set

$$f(K, N) = c + \frac{c_2 K^2}{N^2}, \qquad c_2 > 0. \tag{2}$$

The first case corresponds to kinase activation by other kinase species present at a constant level. In the second case the auto-activation by the same kinase species is considered. The quadratic dependence may result either when the activatory unit of the kinase is a dimer or when the double phosphorylation is needed to activate the kinase - both situations are fairly common in cell signalling [1,2]. We scale the activation coefficient with the total number of kinases N in order to have the same deterministic approximation of systems with different N; and so, in the deterministic limit, the system can be described by the nonlinear ordinary differential equation:

$$\frac{dK}{dt} = (N - K)f(N, K) - bK, \tag{3}$$

which in scaled variable $k = K/N$ takes the form independent on N,

$$\frac{dk}{dt} = (1 - k)f(k) - bk. \tag{4}$$

The system without auto-regulation, $f(k) = c$, has the unique steady state $k_0 = c/(c + b)$. In the auto-regulatory case, $f(k) = c + c_2 k^2$, fixed points are real roots of the third order polynomial

$$W = -c_2 k^3 + c_2 k^2 - (c + b)k + c = 0. \tag{5}$$

Here, we focus on the bistable case when W has 3 real roots such that $0 < k_1 < k_2 < k_3 < 1$. Steady states k_1 and k_3 are stable, while k_2 is unstable. Due to the fact that W has the same coefficient at third and second power, its roots satisfy

$$k_1 + k_2 + k_3 = 1. \tag{6}$$

For further analysis we use roots k_1, k_2, k_3, rather than the original coefficients b, c, c_2, which may be recovered from roots, using Vieta formulas, by the following relations:

$$c = \frac{bk_1k_2k_3}{(k_1 + k_2)(1 + k_1k_2 - k_1 - k_2)}, \qquad c_2 = \frac{b}{(k_1 + k_2)(1 + k_1k_2 - k_1 - k_2)}. \tag{7}$$

Let us notice, that if c and c_2 are expressed as a functions of b, the last coefficient determines only the time scale of the process $\tau = 1/b$. Due to relation (6) the (k_1, k_2, k_3) parameter space may be reduced to domain $D = (k_1, k_2)$ in which $k_1 < k_2$ and $1 - k_1 - k_2 = k_3 > k_2$, see Fig. (5).

RESULTS

The SPD P_K corresponding to considered Markov process may be calculated analytically as follows. From Eq. (1) we get

$$\begin{cases} \frac{dP_0}{dt} = & bP_1 - Nf(0,N)P_0, \\ \frac{dP_K}{dt} = & bP_{K+1} + f(K-1,N)(N+1-K)P_{K-1} - \\ & -(b + f(K,N)(N-K))P_K \qquad \text{for} \quad 0 < K < N, \\ \frac{dP_N}{dt} = & f(N-1,N)P_{N-1} - bP_N. \end{cases} \tag{8}$$

Now, for $\frac{dP_K}{dt} = 0$ we can calculate P_K using the recurrence formula

$$\begin{cases} P_1 = Nf(0,N)P_0/b, \\ P_{K+1} = ((b + f(K,N)(N-K))P_K - f(K-1,N)(N+1-K)P_{K-1})/b \quad \text{for} \quad 0 < K < N. \\ P_N = \frac{f(N-1,N)}{b} P_{N-1} \end{cases} \tag{9}$$

Since the recurrence (9) is linear with respect to P_0, we can set $P_0 = 1$, calculate all the P_K, and then normalize them by dividing by $\sum P_K$. Although, we have analytical recurrence formula, we were able to analyze it only numerically.

To analyze time dependent CPD we performed the Gillespie algorithm simulations of the system (1) with varied initial conditions $K(0)$ and collect total time $t(K)$ which system spends in each state K. Then, we calculate $P_K(T, K(0)) = t(K)/T$, where T is the total simulation time. Since the system is ergodic for the fixed N the CPD $P_K(T, K(0))$ must converge to the SPD P_K for all initial conditions $K(0)$. As shown in Fig. (2), in the case without auto-regulation, *i.e.* when the system is monostable, this convergence is relatively fast even for big N. For $N = 100$ and $N = 5000$ the CPD for $T = 2000\tau$ is almost indistinguishable from the SPD which concentrates around Nk_0, where, recall, $k_0 = c/(c + b)$ is the unique fixed point for the system without auto-regulation.

The much different situation is when system is bistable. In this case, the required time T at which CPD $P_K(T, K(0))$ approaches the stationary probability P_K sharply grows with the system size N. As shown in Fig. (3), for $N = 2000$ and $T = 1.5 \times 10^5\tau$ the CPD $P_K(T, K(0))$ obtained in Gillespie simulation is not sensitive to the initial condition $K(0)$ and is very close to the SPD P_K calculated by the recurrence formula. However, for the same simulation time T and $N = 5000$, $P_K(T, K(0) = 0)$ differs from $P_K(T, K(0) = N)$ and they both differ from the SPD. For the initial condition $K(0) = 0$ the CPD concentrates around k_1N, and for $K(0) = N$ it concentrates around k_3N. This shows that although the system remains ergodic for all finite N, the characteristic "communication time" between states k_1N and k_3N sharply grows with N.

The recurrence (9) calculation were performed using double precision in C++ and GNU Multiprecision Library, which enables us for analysis of systems of size up to $N = 20000$ kinases, however for bigger N this method fails due to the huge differences between P_K, for example for $N = 18000$, in the case depicted in Fig. (4) (second column) $P_{K1}/P_0 \sim 10^{900}$ where $K1 = \mathrm{round}(k_1N)$.

Performing numerous calculations for different set of roots k_1, k_2 we found that for large N the probability distribution generically concentrates around only one of the two stable states. This is illustrated in Fig. (4), where the SDP is calculated for three values of N and three sets of roots $S_L = \{\overline{k_1} - \epsilon, \overline{k_2} + \epsilon\}$, $S_0 = \{\overline{k_1}, \overline{k_2}\}$, $S_R = \{\overline{k_1} + \epsilon, \overline{k_2} - \epsilon\}$. The set $S_0 = \{\overline{k_1} = 0.1782, \overline{k_2} = 0.3218\}$ is chosen so that for $N = 18000$ the probability density splits almost equally into two

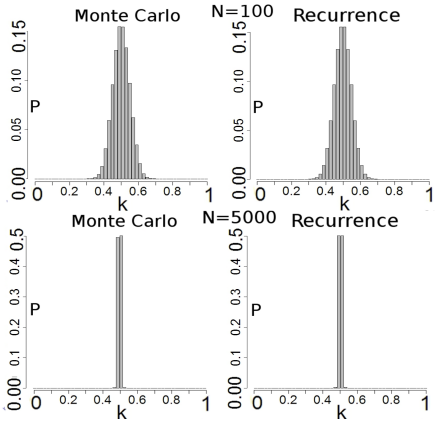

Figure 2. Probability distribution in monostable case with steady state in $k_0 = 0.5$, which refers to $c = b$. Left column: cumulative probability distribution obtained in Gillespie simulations for $T = 2000\tau$ and initial conditions $K(0) = 0$. Right column: steady state probability distribution calculated by the recurrence Eq. (9). Upper row $N = 100$, lower row $N = 5000$

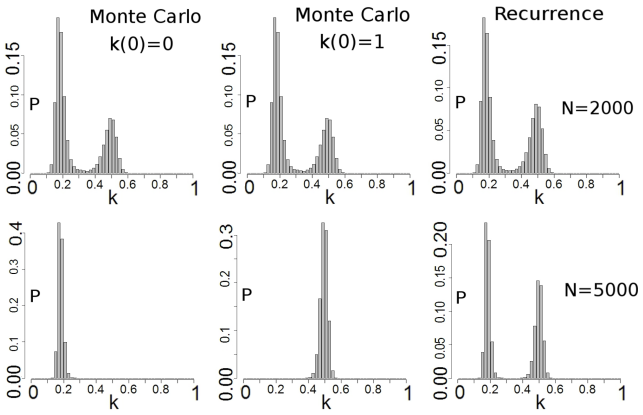

Figure 3. Probability distribution in bistable case with roots $k_1 = 0.1787$, $k_2 = 0.3213$, $k_3 = 0.5$. First and second column Monte Carlo simulations for $T = 1.5 \times 10^5$; Intial conditions for the first and second column are $K(0) = 0$ and $K(0) = N$ respectivelly. Last column steady state probality distribution calculated by the recurrence Eq. (9). Upper row $N = 2000$, lower row $N = 5000$.

basins of attraction. The "shift" $\epsilon = 0.0012$ is chosen in such a way that 90% of probability concentrates either in left, or right basin of attraction. This shows, that for large N, the values of k_1, k_2 have to be very precisely tuned in order to get SDP concentrated evenly around two stable steady states. In Fig. (5A) we show that, ϵ is a function of N and $\epsilon(N) \cong 1/N$ for large N. Since this analysis suggests that $\lim \epsilon(N) = 0$ for $N \to \infty$, we may draw the following hypothesis; The two-dimensional parameter space $D = (k_1, k_2)$ can be split by line $L = k_2(k_1)$ into two subdomains D_1, D_3, such that in limit of $N \to \infty$, for $\{k_1, k_2\} \in D_1$ SPD concentrates around k_1, while for $\{k_1, k_2\} \in D_3$ SPD concentrates around third steady state k_3. Our analysis, however, gives no suggestion about the convergence of SPD (for $N \to \infty$) for the points lying on the line L.

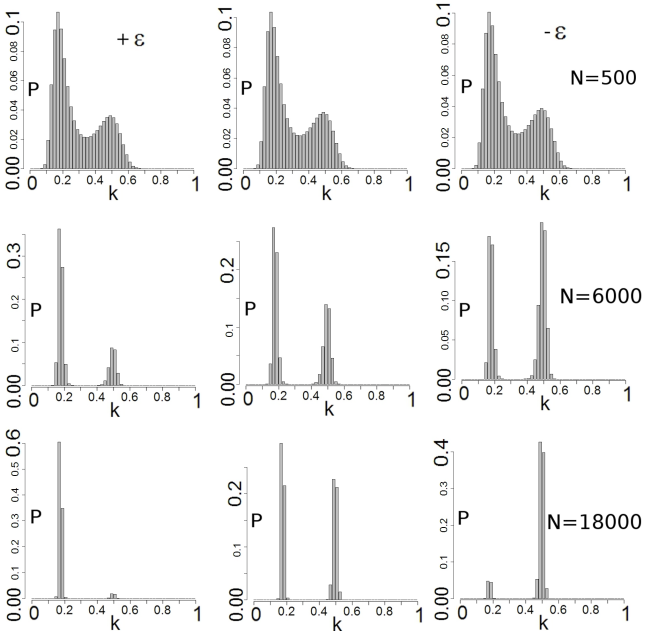

Figure 4. Stationary probability distributions calculated for three values of N: 500, 6000, 18000 (upper to lower row) and three different polynomials W of roots; middle column $\bar{k}_1 = 0.1782, \bar{k}_2 = 0.3218, \bar{k}_3 = 0.5$, right column: $\bar{k}_1 - \epsilon, \bar{k}_2 + \epsilon, \bar{k}_3$, left column: $\bar{k}_1 + \epsilon, \bar{k}_2 - \epsilon, \bar{k}_3$, where $\epsilon = 0.012$.

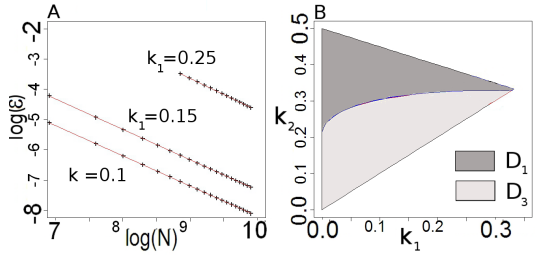

Figure 5. Left panel: numerically calculated $\epsilon(N)$ showed on log-log plot with the best fits $\log(\epsilon) = \alpha \log(N)$ that gave $\alpha = -1.00$ for $k_1 = 0.1, k_2 \approx 0.303$, $\alpha = -1.00$ for $k_1 = 0.15, k_2 \approx 0.316$ and $\alpha = -1.06$ for $k_1 = 0.25, k_2 \approx 0.321$; Right panel: the two-dimensional parameter space $D = (k_1, k_2)$ with line separating subdomains D_1 and D_3 for which stationary probability distribution concentrates in k_1 or k_3 in the limit of $N \to \infty$.

Finally, we investigated the population probability distribution $P_K(T)$ defined as the probability that the system of random initial condition ($P_K(0) = 1/N$) is in state K in time interval $(0, T)$. Analysis of $P_K(T)$ revealed existence of three time scales; (see Fig. (6))

- Short, where the PPD remains almost uniform
- Intermediate, where PPD is concentrated around stable steady states k_1 and k_3 in such a way that PPD concentrated around k_i is proportional to the basin of attraction of k_i
- Long, where PPD is close to the SPD.

The CPD and PPD for the intermediate time scales can be more relevant to the real biological process than the stationary distribution, as the time to reach SPD can be extremely long -

much longer than cell cycle. As showed in Fig. (6), in bistable system, the PPD for intermediate times is qualitatively different from SPD. As one may expect in monostable system (without auto-regulation) the convergence of PPD to SPD is much faster - data not shown.

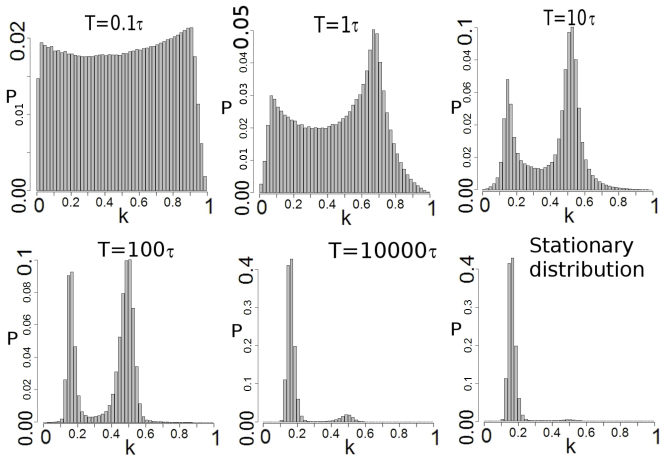

Figure 6. The population probability distribution $P_K(T)$ for steady states $k_1 = 0.162$, $k_2 = 0.338$, $k_3 = 0.5$, $N = 2000$ and 5 different times T compared to the stationary probability distribution given by the recurrence (9).

CONCLUSIONS

The correspondence between the saddle-node bifurcation diagram and probability density function have been analyzed recently by Song et al. [3]. The authors considered bistable galactose utilization network in S. cerevisiae, in which cell population splits into two subpopulations with high and low *Gal10* gene expression. In bistable range, the ratio of two subpopulations was found to be a function galactose level - considered as bifurcation parameter.

Here, based on numerical analysis we found that, generically, in the zero noise limit ($N \to \infty$) stationary probability distribution in bistable case concentrates in the one of the two stable steady states. We showed also, that in such a case, there is a significant difference between the stationary probability distribution and time dependent probability distributions: CPD and PPD, which – although in the limit of $T \to \infty$ converge to SPD – for intermediate times differ qualitatively from the latter.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Witold Bednorz for discussion.
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] N.I. Markevich, J.B. Hoek, and B.N. Kholodenko: *Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades*, J. Cell. Biol. **164** (2004), 353–359.

[2] A.M. Gilfillan and J. Rivera: *The tyrosine kinase network regulating mast cell activation*, Immunological Reviews **228** (2009), 149–163.

[3] C. Song, H. Phenix, V. Abedi, M. Scott, B.P. Ingalls, et al.: *Estimating the Stochastic Bifurcation Structure of Cellular Networks*, PLoS Comput. Biol. **6(3)** (2010), e1000699. doi:10.1371/journal.pcbi.1000699.