Statistics of transitions in single molecule kinetics

Irina V. Gopich[a] and Attila Szabo
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

(Received 16 July 2002; accepted 1 October 2002)

When a system is monitored for a finite time T, the number of transitions between two states is random. In this Note, we obtain the probability distribution of the number of transitions that occur in a fixed time for a system whose dynamics can be described by an arbitrary kinetic scheme. In particular, for long observation times, we show that the variance in the number of transitions for all transitions can be obtained by inverting a modified rate matrix. Our interest in this problem arose in connection with the analysis of single molecule Förster energy transfer studies of protein folding,[1–3] where the number of photons emitted by the donor and acceptor is monitored. If the transition between two states is associated with the emission of a photon, then the probability distribution of the number of photons emitted during a fixed time is related to the distribution of the number of transitions. Our results should prove useful not only for single molecule fluorescence spectroscopy but also in other contexts. For example, in the analysis of the dynamic instability of microtubule growth, the statistics of the transitions between growing and shrinking phases is of interest. In fact the approach we use here is a generalization of the technique employed by Bicout and Rubin[4] to analyze a simple model of microtubule growth.

Consider a general kinetic scheme where the rate constant of the $i\to j$ transition is K_{ji}. Let $G_{ij}(t)$ be the Green's function or propagator which is the probability of finding the system in state i at time t provided initially it was in state j. The matrix of such probabilities satisfies the rate equation:

$$\frac{d}{dt}\mathbf{G}(t)=\mathbf{K}\mathbf{G}(t) \tag{1}$$

with initial condition $\mathbf{G}(0)=\mathbf{I}$, where $\mathbf{I}$ is the unit matrix. The subsequent analysis is most easily performed in the Laplace domain ($\hat{f}(s)=\int_0^\infty f(t)\exp(-st)dt$) where

$$\hat{\mathbf{G}}(s)=(s\mathbf{I}-\mathbf{K})^{-1}. \tag{2}$$

In terms of this function, we now find $P_{\beta\alpha}(N,T)$ [or equivalently its Laplace transform $\hat{P}_{\beta\alpha}(N,s)$], which is the probability that the system, starting in state α and ending up in state β at time T, underwent N $i\to j$ transitions. For notational simplicity, we indicate only the initial and final states. To obtain this quantity, it is convenient to introduce[4] a counting parameter λ by multiplying the off diagonal element K_{ji} of the rate matrix by λ. The corresponding Green's function can then be written as

$$\hat{\mathbf{F}}(\lambda,s)=(s\mathbf{I}-\mathbf{K}'-\lambda\mathbf{V})^{-1}, \tag{3}$$

where $\mathbf{V}$ is a matrix whose only nonzero element is K_{ji} and $\mathbf{K}'=\mathbf{K}-\mathbf{V}$. It is interesting to note that $\mathbf{K}'$ describes a kinetic scheme in which state i decays irreversibly (i.e., not to state j) with rate constant K_{ji}. Clearly, $\hat{\mathbf{F}}(1,s)=\hat{\mathbf{G}}(s)$. For $\lambda=0$, $\hat{\mathbf{F}}(0,s)=\hat{\mathbf{G}}'(s)$ where $\hat{\mathbf{G}}'(s)=(s\mathbf{I}-\mathbf{K}')^{-1}$. $\hat{G}'_{\beta\alpha}$ is just the probability of going from state α to state β in time T *without* making an $i\to j$ transition. It can be shown that it is related to the propagator corresponding to the original kinetic scheme [i.e., Eq. (2)] by

$$\hat{G}'_{\beta\alpha}(s)=\hat{G}_{\beta\alpha}(s)-\frac{\hat{G}_{\beta j}(s)K_{ji}\hat{G}_{i\alpha}(s)}{1+K_{ji}\hat{G}_{ij}(s)}. \tag{4}$$

If the propagator $\hat{\mathbf{F}}(\lambda,s)$ is expanded in powers of λ, the coefficient of λ^N is the contribution of all trajectories that involve exactly N $i\to j$ transitions and thus

$$\hat{\mathbf{F}}(\lambda,s)=\hat{\mathbf{G}}'+\lambda\hat{\mathbf{G}}'\mathbf{V}\hat{\mathbf{G}}'+\lambda^2\hat{\mathbf{G}}'\mathbf{V}\hat{\mathbf{G}}'\mathbf{V}\hat{\mathbf{G}}'+\cdots=\sum_{N=0}^{\infty}\lambda^N\hat{\mathbf{P}}(N,s). \tag{5}$$

This kind of perturbation expansion is a useful tool for answering a variety of questions about kinetic schemes (e.g., Cao's analysis of the distribution of residence times).[5] By equating coefficients of λ^N, we have

$$\hat{P}_{\beta\alpha}(N,s)=K_{ji}^N\hat{G}'_{\beta j}(s)\hat{G}'^{N-1}_{ij}(s)\hat{G}'_{i\alpha}(s),\quad N\geqslant 1 \tag{6}$$

and $\hat{P}_{\beta\alpha}(0,s)=\hat{G}'_{\beta\alpha}(s)$ for $N=0$. Here $\hat{G}'_{\beta\alpha}(s)$ is given by Eq. (4). This is the formal solution to our problem.

If systems chosen from an equilibrium distribution are observed for time T or if a long trajectory is divided into bins of length T, the probability distribution of observing N $i\to j$ transitions in a bin is given by

$$\begin{aligned}\hat{P}(N,s)&=\sum_{\alpha,\beta}\hat{P}_{\beta\alpha}(N,s)p_\alpha^0\\&=\frac{K_{ji}^N p_i^0\hat{G}_{ij}^{N-1}(s)}{s^2(1+K_{ji}\hat{G}_{ij}(s))^{N+1}},\quad N\geqslant 1\\&=\frac{1}{s}\left(1-\frac{K_{ji}p_i^0}{s(1+K_{ji}\hat{G}_{ij}(s))}\right),\quad N=0,\end{aligned} \tag{7}$$

where we have used Eqs. (4) and (6). Here p_i^0 is the normalized steady state probability of state i (i.e., the solution of the $\mathbf{K}\mathbf{p}^0=0$). This result can be inverted into the time domain using, for example, the Stehfest algorithm.[6]

It is often useful to characterize a complicated probability distribution in terms of only its mean and variance. Let $\langle N(T)\rangle$ be the mean number of $i\rightarrow j$ transitions that occur during time T under steady state conditions. Then

$$\langle \hat{N}(s)\rangle = \sum_{N=1}^{\infty} N\hat{P}(N,s) = \sum_{\alpha,\beta} \frac{\partial}{\partial\lambda}\hat{F}_{\beta\alpha}(\lambda,s)|_{\lambda=1} p_\alpha^0. \quad (8)$$

The derivative at $\lambda=1$ is easily evaluated by noting that $\hat{\mathbf{F}}(\lambda,s)=(s\mathbf{I}-\mathbf{K}'-\lambda\mathbf{V})^{-1}=(s\mathbf{I}-\mathbf{K}-(\lambda-1)\mathbf{V})^{-1}$. In this way we recover the standard result:

$$\langle N(t)\rangle = K_{ji}p_i^0 T. \quad (9)$$

The variance, $\sigma_N^2(T)\equiv\langle N^2(T)\rangle-\langle N(T)\rangle^2$, can be found analytically by evaluating the second derivative of $\hat{\mathbf{F}}(\lambda,s)$ at $\lambda=1$:

$$\sigma_N^2(T)=\langle N(T)\rangle - 2(K_{ji}p_i^0)^2\int_0^T (T-t)\frac{\langle \delta n_i(t)\,\delta n_j(0)\rangle}{\langle \delta n_i \delta n_j\rangle}dt, \quad (10)$$

where $\delta n_i(t)$ is the deviation of the population of state i from steady state and $\langle \delta n_i(t)\,\delta n_j(0)\rangle/\langle \delta n_i\delta n_j\rangle = 1-G_{ij}(t)/p_i$ is the population cross correlation function. For long times the variance becomes linear in time:

$$\lim_{T\rightarrow\infty}\sigma_N^2(T)=K_{ji}p_i^0 T(1-2K_{ji}p_i^0\tau_{ij}), \quad (11)$$

where τ_{ij} is the relaxation time of the population cross relaxation function. This time is the element of matrix $\boldsymbol{\tau}$ which can be shown to be given by

$$\boldsymbol{\tau}=(\mathbf{KD}+\boldsymbol{\Phi})^{-1}-\mathbf{E}, \quad (12)$$

where we have defined $D_{ml}=p_m^0\delta_{ml}$, $\Phi_{ml}=p_m^0 p_l^0$ and matrix $\mathbf{E}$ has all unit elements. This result shows that it is possible to obtain the variance in the number of transitions for every transition at long times by a single matrix inversion.

At long times the probability distribution is Gaussian with the above mean and variance. We can obtain a better approximation that is useful for shorter times by using a two-state approximation for the Green's function:

$$G_{ij}(t)\approx p_i^0(1-e^{-t/\tau_{ij}}), \quad i\neq j, \quad (13)$$

where τ_{ij} is given by Eq. (12). This expression is exact at $t=0,\infty$ and has the exact "area." Using this in Eq. (7) in inverting the Laplace transform we find for $N\geqslant 1$,

$$P(N,T)=\frac{(1-\gamma^2)^N t^N e^{-t}}{(2\gamma)^N N!\sqrt{8\gamma t/\pi}}\{2\gamma(N+t)I_{N-1/2}(t) + (1+\gamma^2)tI_{N+1/2}(t)\}, \quad (14)$$

where $t=T/(2\tau_{ij})$, $\gamma=\sqrt{1-4K_{ji}p_i^0\tau_{ij}}$, and $I_n(z)$ are modified Bessel functions of the first kind. This result is exact for a two state system. As $T\rightarrow\infty$, it is exact for any system.

As the simplest example of the formalism consider the reaction $D\rightleftharpoons D^*$ where D (state 1) and D^* (state 2) are the ground and excited states of a fluorophore. The system is excited by continuous illumination with rate constant K_{21}. The rate constant for the decay (both radiative and nonradiative) of D^* is K_{12}. We are interested in the distribution of number of photons observed during time t. Suppose that only a fraction ϕ (which is the product of the quantum yield and detector efficiency) of the $D^*\rightarrow D$ transitions result in detectable photons. This can be easily taken into account in our general formalism by simply replacing K_{ji} by ϕK_{ji} in all our expressions in which K_{ji} appears explicitly ($\hat{\mathbf{G}}$ and $\boldsymbol{\tau}$ should not be modified). Thus in this case, the exact distribution for observing N photons in time t is given by Eq. (14) with $i=2$, $j=1$, $\tau_{21}=(K_{12}+K_{21})^{-1}$, $p_2^0=K_{21}/(K_{12}+K_{21})$, and $K_{12}\rightarrow\phi K_{12}$.

The above-given analysis can be readily extended to obtain the joint probability of observing N $i\rightarrow j$ and M $k\rightarrow l$ transitions during time T. This is of interest, for example, in Förster resonance energy transfer where the $i\rightarrow j$ and $k\rightarrow l$ transitions are associated with the emission of photons from donor and acceptor, respectively. To obtain the corresponding probability distribution one simply introduces two counting variables λ (for $i\rightarrow j$) and μ (for $k\rightarrow l$) and determines the coefficient of $\lambda^N\mu^M$ in the expansion of the resulting Green's function. The general result is a little too messy to be presented here, but a simple expression can be found for the cross correlation, $\sigma_{NM}^2(T)=\langle N(T)M(T)\rangle-\langle N(T)\rangle\langle M(T)\rangle$, under steady state conditions:

$$\sigma_{NM}^2(T)=-K_{ji}K_{lk}p_i^0 p_k^0\int_0^T dt\,(T-t) \times\left(\frac{\langle \delta n_i(t)\,\delta n_l(0)\rangle}{\langle \delta n_i\delta n_l\rangle}+\frac{\langle \delta n_k(t)\,\delta n_j(0)\rangle}{\langle \delta n_k\delta n_j\rangle}\right), \quad (15)$$

which for long times becomes

$$\lim_{T\rightarrow\infty}\sigma_{NM}^2(T)=-K_{ji}K_{lk}p_i^0 p_k^0 T(\tau_{il}+\tau_{kj}), \quad (16)$$

where τ_{ij} is given by Eq. (12).

In summary, we have expressed the probability distribution of observing N state-to-state transitions in a fixed time in terms of the matrix of conditional probabilities of going from one state to another. At long times this probability distribution is Gaussian with a mean and variance that depend only on the rate constant of transition of interest, the steady state probability of the initial state, and the average time required for the initial state to reach steady state when the system starts out in the final state. Finally, we should mention that although we presented explicit expressions only for discrete systems, our formalism can be trivially extended to treat Markovian systems in which some of the states are specified by continuous variables (e.g., systems described by reaction-diffusion equations).

We thank Sasha Berezhkovskii for very helpful discussions.

[a]On leave from the Institute of Chemical Kinetics and Combustion SB RAS, Novosibirsk, 630090, Russia; electronic mail: irinag@speck.niddk.nih.gov

[1] D. S. Talaga, W. L. Lau, H. Roder, J. Tang, Y. Jia, W. F. DeGrado, and R. M. Hochstrasser, Phys. Rev. A **97**, 13021 (2000).
[2] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss, Phys. Rev. A **97**, 5179 (2000).
[3] B. Schuler, E. Lipman, and W. A. Eaton, Nature (London) **419**, 743 (2002).
[4] D. J. Bicout and R. J. Rubin, Phys. Rev. E **59**, 913 (1999).
[5] J. Cao, Chem. Phys. Lett. **327**, 38 (2000).
[6] H. Stehfest, Commun. ACM **13**, 47 (1970).