Continuous Time Markov Chains

C.W. Seitz

September 7, 2023

1 Continous Time Markov Chains

1.1 Brief derivation of the master equation

Each row of the generator matrix G_i represents the a conditional probability distribution on the future state given the present state $P(\omega|X(t) = \omega_j)$:

$$\sum_j G_{ij} = \sum_j P(X(t+dt) = \omega_j|X(t) = \omega_i) = 1$$

The generator is not necessarily symmetric. Also, one should note that the columns G_j *do not* define a probability distribution and do not necessarily sum to unity. The columns of G have no meaning in this context, since we have defined the rows to represent a probability of the future given the present. The first order dynamics for a particular state ω_i is given by

$$P(\omega_i, t + dt) = P(\omega_i, t) + \mathcal{J}_i dt$$

where the probability current $\mathcal{J}_i$ must be the net probability flux into the state ω_i

$$\mathcal{J}_i = \sum_i G_{ij} P(\omega_j, t) - \sum_j G_{ij} P(\omega_i, t)$$

The first is a sum on a column (in) and the second a sum on a row (out). This can be simplified further by noticing that the normalization condition implies

$$G_{ij} = 1 - \sum_j G_{ij}(1 - \delta_{ij}) = 1 - \sum_j G_{ij} + \sum_j G_{ij}\delta_{ij}$$

Inserting this into the probability current gives

$$\mathcal{J}_i = \sum_i G_{ij} P(\omega_j, t) - \sum_j G_{ij} P(\omega_i, t)$$

$$= \sum_i \left(1 - \sum_j G_{ij} + \sum_j G_{ij} \delta_{ij} \right) P(\omega_j, t) - \sum_j G_{ij} P(\omega_i, t)$$

$$= |\Omega| - |\Omega| + \sum_i \sum_j G_{ij} P(\omega_j, t) \delta_{ij} - \sum_j G_{ij} P(\omega_i, t)$$

$$= \sum_j G_{ji} P(\omega_j, t) - G_{ij} P(\omega_i, t)$$

which is RHS of the master equation. Notice that the Kronecker delta effectively just swaps the index.

2 Markovian photoswitching dynamics

The number of molecules within the diffraction limit is $K\left(\frac{\lambda}{2\mathrm{NA}}\right)$. If α is the *detection probability*, then $\alpha K\left(\frac{\lambda}{2\mathrm{NA}}\right)$ are detected, on average. We want to minimize

$$\mathcal{L} = \alpha K \left(\frac{\lambda}{2\mathrm{NA}} \right) + \gamma \left(\Delta_{\mathrm{SR}} + \frac{2N}{\log(1-\alpha)} \right)^2$$

The second term contains $\frac{2N}{\log(1-\alpha)}$, which is the minimum number of frames needed to detect 99 percent of N molecules (which can be obtained from the geometric distribution). If we assume a two-state generator, then

$$P(t) = P(0) e^{Gt}$$

and $G\pi = 0$ gives $\pi = (\alpha, \beta) = \frac{1}{k}(k_{12}, k_{21})$ where $k = k_{12} + k_{21}$.

The true photo-switching behavior of the fluorophore is a continuous time stochastic phenomenon. However, an experimenter can only ever observe a discretized manifestation of this by imaging the fluorophore in a sequence of frames. These frames are regarded as a set of sequential exposures of the fluorophore and the observed discrete time signal indicates whether the fluorophore has been observed in a particular frame. It is the continuous time process on which we wish to draw inference based on the observed discrete-time process indicating whether the fluorophore was observed in a frame. In this section we first present the continuous time Markov model of the true (hidden) photo-switching behavior, and then derive the observed discrete time signal, together with key results on its statistical properties.

We model the photoswitching behavior of a fluorophore as a continuous time Markov process. It is a stochastic process, which satisfies the Markov property

$$P(X_t | X_{t-1}, X_{t-2}, ..., X_{t-N}) = P(X_t | X_{t-1})$$

We consider a general model for $X(t)$ that can accommodate the numerous mechanisms of photo-switching utilized in standard SMLM approaches such as (F)PALM and (d)STORM. Specifically, this model consists of a photon emitting (On) state 1, $m+1$ non photon emitting (dark/temporary off) states $0_0, 0_1, \ldots, 0_m$, where $m \in \mathbb{Z} \geq 0$, and a photobleached (absorbing/permanently off) state 2. In order to accommodate for the $m = 0$ case when we have a single dark state, we use the notational convention that state $0_0 \equiv 0$. The model, allows for transitions from state 1 to the multiple dark states (from a photochemical perspective, these can include triplet, redox and quenched states). These dark states are typically accessed via the first dark state 0 (reached as a result of inter-system crossing of the excited $S1$ electron to the triplet $T1$ state). Further dark states 0_{i+1}, $i = 0, \ldots, m-1$, are accessible by previous dark states 0_i (by, for example, the successive additions of electrons forming radical anions (Van de Linde et al., 2010)). We allow the On state 1 to be accessible by any dark state and we consider the most general model in which the absorption state 2 is accessible from any combination of other states (Vogelsang et al., 2010; Van de Linde and Sauer, 2014; Ha and Tinnefeld, 2012).

To capture $P(X_t|X_{t-1})$ for all possible pairs X_t and X_{t-1}, we define a square generator matrix $G \in \mathcal{R}^{N \times N}$ where N denotes the number of states. As such, the elements of G represent the probability of a transition from a state ω_j to ω_i in an infinitesimal time interval

$$G_{ij} = \Pr\left(X(t+dt) = \omega_i, \mid X(t) = \omega_j\right)$$

Let the state space for the process $X(t)$ be $\Omega = \{0_0, 0_1, 0_2, 1, 2\}$. The generator matrix for such a process reads

$$G = \begin{pmatrix} \lambda_{00} & \lambda_{00_1} & 0 & \lambda_{01} & \mu_0 \\ 0 & \lambda_{0_10_1} & \lambda_{0_10_2} & \lambda_{0_11} & \mu_1 \\ 0 & 0 & \lambda_{0_20_2} & \lambda_{0_21} & \mu_2 \\ \lambda_{10} & 0 & 0 & \lambda_{11} & \mu_0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Of considerable practical interest is using the matrix G to determine the probability the system is in a particular state at a time t. This can be achieved by solving the master equation:

$$\frac{\partial P(\omega_i)}{\partial t} = \sum_j G_{ji} P(\omega_j, t) - G_{ij} P(\omega_i, t)$$

Our notational convention is the G_{ij} is a transition $i \rightarrow j$ while G_{ji} is $j \rightarrow i$. The above equation and its solution is very powerful in this context and in the broader context of non-equilibrium dynamics, so I will devote a short section to its derivation. The reader can safely skip this section if time is scarce.

2.1 Solving the master equation

The master equation is a first order differential equation and its solution is straightforward when the dimensionality of the state space is small. The solution is found easily by massaging the master equation into something that has a simple exponential solution. Define a matrix W s.t. $W_{ij} = T_{ij}$ and $W_{ii} = -\sum_{j \neq i} G_{ij}$. This operator acts on $P(\omega)$ and gives a vector of probability currents $\dot{P}(\omega, t) = \mathcal{J}(\boldsymbol{\omega}) = WP(\boldsymbol{\omega})$.

$$P(\boldsymbol{\omega}, t) = \exp(WP(\omega))$$

The matrix W for the 4-state system presented before reads

$$W = \begin{pmatrix} -\sigma_0 & \lambda_{00_1} & 0 & \lambda_{01} & \mu_0 \\ 0 & -\sigma_{0_1} & \lambda_{0_1 0_2} & \lambda_{0_1 1} & \mu_1 \\ 0 & 0 & -\sigma_{0_2} & \lambda_{0_2 1} & \mu_2 \\ \lambda_{10} & 0 & 0 & -\sigma_1 & \mu_0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where we have introduced shorthands $\sigma = -\sum_{j \neq i} G_{ij}$.

2.2 Modeling photoswitching dynamics

The imaging procedure requires taking a series of successive frames. Frame n is formed by taking an exposure over the time interval $[n\Delta, (n+1)\Delta)$, where $n \in \mathbb{Z} \geq 0$. The constant Δ corresponds to the exposure time for a single frame, also known as the frame length. We define the discrete time observed process $Y_n : n \in \mathbb{Z} \geq 0$, with state space $S_Y = 0, 1$, as $Y_n = 1$ if the fluorophore (characterized by $X(t)$) is observed in frame n and equal to 0 otherwise. For the fluorophore to be observed in the time interval $[n\Delta, (n+1)\Delta)$, it must be in the On state 1 for a minimum time of $\delta \in [0, \Delta)$. The value of δ is unknown and is a result of background noise and the imaging system's limited sensitivity. We note that if $X(t)$ exhibits multiple jumps to state 1 within a frame, then a sufficient condition for observing the fluorophore is that the total time spent in the On state exceeds δ. The $\delta = 0$ case is the idealistic scenario of a noiseless system and perfect sensitivity such that the fluorophore is detected if it enters the On state for any non-zero amount of time during the exposure time Δ.

$$Y_n = \mathbb{1}_{[\delta, \Delta)} \left(\int_{n\Delta}^{(n+1)\Delta} \mathbb{1}_1(X(t)) dt \right)$$

Further, the distribution on Δt will depend on the hidden state of the system before integration begins.

$$\Delta t = \int_{n\Delta}^{(n+1)\Delta} \mathbb{1}_1(X(t)) dt \quad \Delta t \sim P(\Delta t | X_t = i)$$

where $P(\Delta t | X_t = i)$ will be an exponential distribution with parameters dependent on the starting state. Following (Patel 2019), we write down the

joint probability of a transition from state i to state j during an interval Δ and the observation $Y_n = \ell$.

$$b_{ij,\Delta}^{\ell} = \mathbb{P}(Y_n = \ell, X(\Delta) = j | X(0) = i)$$
$$= \sum_k q_{ij}(k, \Delta) \xi_{ij}(\ell, k, \Delta)$$

where we have made the following definitions

$$q_{ij}(k, \Delta) = \mathbb{P}(N(\Delta) = k, X(\Delta) = j | X(0) = i)$$
$$\xi_{ij}(\ell, k, \Delta) = \mathbb{P}(Y_n = \ell | N(\Delta) = k, X(\Delta) = j, X(0) = i)$$

We focus first on $q_{ij}(k, \Delta)$, which represents the probability distribution over k transitions of type ij occuring in the interval Δ. This is analogous to considering the time-evolution of the probability mass over states in our solution of the master equation. However, here we instead find the time-evolution of the probability mass over the number of transitions k of a certain type q_{ij}, and then integrate out the variable k. As an example, the distribution over k for transition of type q_{i0} (from $i \to 0$), is determined by the distribution over the number of $i \to 1$ transitions and the rate of a $1 \to 0$ transition over a finite time interval Δt

$$q_{i0}(k, t + \Delta t) = (1 - \sigma_0 \Delta t) q_{i0}(k, t) + \lambda_{10} q_{i1}(k, t) \Delta t + \mathcal{O}(\Delta t)$$

The negative contribution represents the distribution over the number of transitions $i \to 0$ that have already occured by time t and will leave the 0 state in the interval Δt. The full coupled system of differential equations reads

A molecule is considered "detected" in principle if the measured ADU signal satisfies $\tilde{s} = \mu\tau \geq \delta$ where δ is a number of photons which satisfy a criterion on localization accuracy.

$$\alpha = \int_{\delta}^{\Delta} \left(\sum_{n=0}^{\infty} Q(N = n) \psi(\tau | n; \vec{k}) \right) d\tau \approx \mathop{\mathbb{E}}_{\tau \sim P(\tau)} (\mathbb{I}[\tau > \delta])$$

$P(\tau)$ is usually obtained by Monte Carlo simulation. This is useful for computing density measures and the total acqusition time:

$$D = \alpha K \left(\frac{\lambda}{2\mathrm{NA}} \right) \quad T = \left(\Delta_{SR} + \frac{2N}{\log(1 - \alpha)} \right)^2$$

For actually inferring k_1, k_2, we need a measure of distance between $P(\tilde{s})$ and $P(s|k_1, k_2)$ for many k_1, k_2 pairs. Luckily we only need to compute $P(s|k_1, k_2)$ once, and we can then perform a grid search