Bayesian estimation of switching rates for blinking quantum emitters

By

Jemy Geordy

A thesis submitted to Macquarie University

for the degree of Master of Research

Department of Physics

November 2018

MACQUARIE
University
SYDNEY·AUSTRALIA

Typeset in LaTeX 2_ε.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Jemy Geordy

Acknowledgements

I would like to thank my supervisors A/Prof Alexei Gilchrist and A/Prof Thomas Volz for their guidance and support throughout this project. I also would like to thank Dr. Lachlan J Rogers for his mentoring and fruitful collaboration during this research study. I would like to thank all members in the Quantum Materials and Application (QMAPP) group for their support. I also would like to thank Dr. Cameron M Rogers for the collaboration. I would like to thank Macquarie University for providing International Macquarie University Research Excellence Scholarship (IMQRES MRES) and other facilities during the period of my masters study. I also would like to thank EQUS for the support. Finally, I would like to thank my love Sarath and my family for their encouragement and support over the years.

Abstract

Quantum emitters such as quantum dots or colour centres in diamond have a range of interesting applications ranging from quantum sensing to biomedical imaging to being the active material in the newest generation of screen technologies. However, the particular environment a quantum emitter is exposed to can lead to intermittency of its fluorescence known as blinking. Understanding blinking dynamics and its causes is essential for optimizing quantum emitters for technological applications. A first step in this process is the correct analysis and inference of underlying blinking rates which characterize the internal switching process of the quantum emitter from a dark to a bright state and vice versa. This thesis develops a methodology for inferring these rates from data using Bayesian analysis. It treats the underlying blinking process as a hidden Markov chain. Both discrete and continuous-time Markov models are developed and applied in order to infer switching rates from fluorescence time-series having a physically realistic range from very slow to very fast blinking. The hidden-chain Markov model developed here could find applications in other areas such as finance and computational biology.

Contents

List of Figures

1

Introduction

Single photon emitters or quantum emitters are fundamental ingredients for many quantum technological applications. As the name indicates, single photon emitters are capable of emitting a single photon per excitation pulse [1]. Only true quantum emitters ensure the correct experimental realisation of quantum cryptographic protocols and quantum computation schemes [2]. Over the years, many different physical implementations of single photon emitters have been developed [1]. Starting originally from trapped atoms and ions, the first stable single-photon emitter in the solid state at room temperature with enormous practical implications was successfully demonstrated in diamond [3]. Today, the most prominent single-photon sources in the solid state include colour centres in diamond [4, 5] and quantum dots [6, 7], with many new candidates in other material systems being heavily explored [8]. One of the predominant requirements of a useful single-photon emitter is its stability in emission. That means, the emitter should be free from *blinking* (periods of different intensities) or *bleaching* (permanent loss of the ability to fluoresce) [8]. However, many of the current

quantum light sources exhibit this unsought blinking phenomena [9], depending on the particular mesoscopic environment they are embedded in. Blinking will naturally limit their potential for applications. This thesis develops powerful new techniques to analyse blinking data for quantum emitter systems. Accurate analysis provides a tool to better understand the physical processes responsible for blinking, paving the way for novel engineering of quantum emitters to eliminate blinking.

1.1 Background and motivation

There are mainly two mechanisms reported in the literature that explain the physical mechanism behind the blinking phenomenon. These are photo-induced blinking [9] and charge-state induced blinking [10, 11]. As a concrete recent example of blinking, the behaviour of gallium nitride (GaN) single-photon emitters was studied with varying laser excitation power in Ref [9]. The normally stable emission from GaN started to exhibit blinking when the excitation power was increased over a certain threshold. The authors also found that the onset of blinking was accompanied by a significant change in photon statistics. As a possible explanation, the authors invoked the light-induced activation of a new trap state that can provide a new non-radiative pathway to the ground state. An example for a charge-state induced blinking behaviour was reported in Ref [10]. They detected two types of blinking in their data. One they attributed to the change in charge state happening in the core of QDs and the other one to the fluctuations in the charge state of electron-accepting surface states. In a study of nitrogen-vacancy (NV) centres in nanodiamonds [12], blinking was observed once the ND size was reduced to 5nm and below. A possible explanation there again is the closeness of NV centres to electron-accepting surface states. In any case it is clear that a good understanding of the underlying physical mechanism of the process is needed to eliminate blinking from these systems and make them useful for applications. The key insight into the physical mechanism typically comes from a thorough analysis of the switching rates. However, these rates are often the hardest information to extract from the raw time series of counts [13]. This thesis focuses on developing novel methods for reliably inferring the rates of switching events from blinking data for a better insight into blinking mechanisms.

Blinking typically leads to step-like switches in the fluorescence time trace, as illustrated in Figure 1.1(a). The most common method used to differentiate between the *on* and *off* states

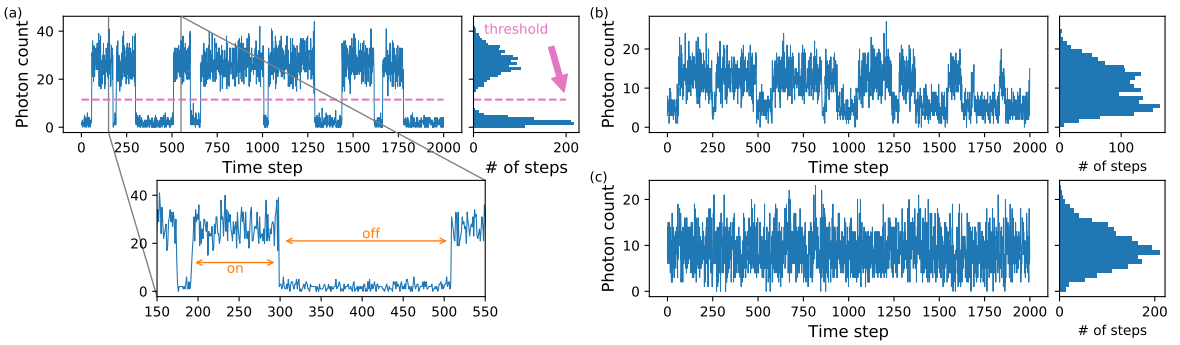


Figure 1.1: (a) Simulated blinking time-trace with clear *on* and *off* states. The doted line represents a threshold intensity level that differentiates between the *on* and *off* states. (b) Simulating a blinking time trace with the same switching rates as in (a) but reduced signal-to-background ratio results in obvious blinking but no applicable threshold. (c) When the switching rates between *on* and *off* are high relative to the data sampling rate then it is hard even to see that the emitter is blinking.

is threshold analysis [13–16]. A plausible threshold is illustrated in Figure 1.1(a). However, the choice of threshold intensity and sampling interval can significantly influence the statistics of the *on* and *off* states [17, 18]. These are often essentially arbitrary choices. What is more, this method is useful only if the *on* and *off* states are clearly differentiated from the blinking time series as for the simulated data in Figure 1.1(a). The use of the threshold technique to analyse the *on* and *off* states becomes difficult for emitters fluorescing at low intensity levels or when there is a high background noise as depicted in Figure 1.1(b). In both these cases brief *on* and *off* intervals are difficult to distinguish from the noise spikes. A similar problem arises if the blinking switches at rates high relative to the sampling rate as shown in Figure 1.1(c). This thesis presents a method based on Bayesian statistical inference to find the blinking rates from photon-count data series. The method works for *blinking* quantum light emitters independent of their brightness or speed of switching. Bayesian analysis is a powerful probabilistic modelling tool that allows inference of unknown parameters from noisy data. This method does not require a threshold intensity level to differentiate between the *on* and *off* states.

The observed photon counts that are accumulated over the sampling interval obscure the underlying process of a blinking quantum light emitter. In this thesis we have developed three different models for a blinking quantum light source. These models are based on hidden discrete and continuous-time Markov processes [19]. A Markov process is a memoryless stochastic process in which the current state is sufficient to determine the future dynamics [20, 21]. The simplest model considered is a blinking emitter with very low switching

rates. In this case, the possibility of switching within a sampling interval can be assumed to negligible. Such an emitter is well modelled using a discrete-time Markov process, with state transition happening only at the sampling boundaries. The second model proposed considers the possibility of multiple switching events inside the sampling interval or equivalently, that the sampling interval spans several time steps of the model. The last model is the extreme case of where the emitter can switch states at any point in time. This case is modelled with a continuous-time Markov process. The Bayesian inference on these models then allows us to determine the underlying switching rates of the state of the emitter hidden in the discretely-sampled detection counts. Consequently we can tackle switching rates ranging anywhere from low rates to very high rates with the appropriate model and inference.

1.2 Outline of the thesis

Chapter 2 briefly outlines blinking quantum light emitters. This chapter also explains discrete and continuous-time Markov processes required to model a *blinking* process. **Chapter 3** is an introduction to Bayesian statistics and it introduces the essential tools required in the following chapters. **Chapter 4** analyses the two models based on discrete time Markov processes. The first model is the simplest blinking process where the emitter is assumed to stay either *on* or *off* during the *entire* sampling interval, and switching of state can occur only at the sampling boundaries. This model is applied to published experimental data as an illustration of its utility. The second model considers subintervals inside a sampling interval to account for faster rates. **Chapter 5** analyses the continuous switching of a blinking emitter using a continuous time Markov process. In this chapter we also analyse the error in the single step model when the switching rates are high. **Chapter 6** concludes the thesis, presenting some of the extensions that may grow out of this work.

2

Blinking Quantum Emitters

Quantum light emitters received a lot of attention in the scientific community for their promising applications in various fields, such as quantum technology and medical applications. As introduced in Chapter 1, *blinking* imposes severe limitations on the usefulness of quantum light emitters. The occurrence of dark intervals can greatly affect the overall number of emitted photons plus the photon statistics of the emitted light. Though these states are not directly observable, they are correlated with the photon counts that are measured.

It is necessary to understand the underlying physical mechanism of the blinking emitters in order to make them stable for engineering systems with desired properties. Investigators have put forward a variety of possible explanations behind blinking phenomena based on the characteristics of the observed blinking. This thesis particularly focuses on the analysis and understanding of *rates* of the switching events in a blinking emitter using simulated blinking time traces, which is necessary for understanding the blinking process more deeply.

This chapter is organized in sections as follows: Section 2.1 gives a short review of blinking in quantum emitters and focuses on the characterisation of blinking from observed

photon counts. It also briefly outlines the conventional analysis method used to separate the *on* and *off* states in experimental blinking data, Section 2.2 discusses discrete and continuous-time Markov processes which are essential for understanding the coming chapters while modelling of physical blinking processes.

2.1 Photoluminescence intermittency

The term photoluminescence (PL) refers to the emission of light from matter when it is excited with electromagnetic radiation. The random fluctuation in the emission states between *bright* and *dark* state on continuous excitation in a quantum emitter is referred as blinking. In fact, it is a common phenomenon that has been observed with most of the current quantum light emitters [9], including quantum dots [22–25], diamond colour centers [12, 13, 26–28] and single molecules [29, 30]. The potential applications of these quantum light sources are limited by the PL blinking. A detailed characterisation of the blinking mechanism is needed for understanding the underlying physics of the blinking system in particular.

An understanding of the underlying physical mechanisms often comes from the dependence of switching rates on particular external parameters such as excitation intensity, wavelength, temperature, and polarisation. Yet, these rates are often the hardest to extract from the raw time series of photon counts. The presence of noise in the data makes it even worse and the task of extracting meaningful switching rates from blinking signals has been a challenge [13]. A universal physical mechanism that could possibly explain the blinking phenomena is still an open problem [9, 24, 31].

2.1.1 Characterisation of blinking data

Usually, in photon counting experiments the photons are counted over a given sampling interval of fixed width [17, 32]. That is, the photon counts that are registerd during the sampling interval get integrated to produce photon counts per time bin. The discrete nature of photons causes statistical fluctuations in their detection on short time scales. These kind of fluctuations in the registred photon counts are intrinsic and cannot be avoided. They are called *shot noise*. An uncorrelated light source with a constant intensity can be modelled using Poissonian statistics [33], with the probability of observing k photons in a sampling

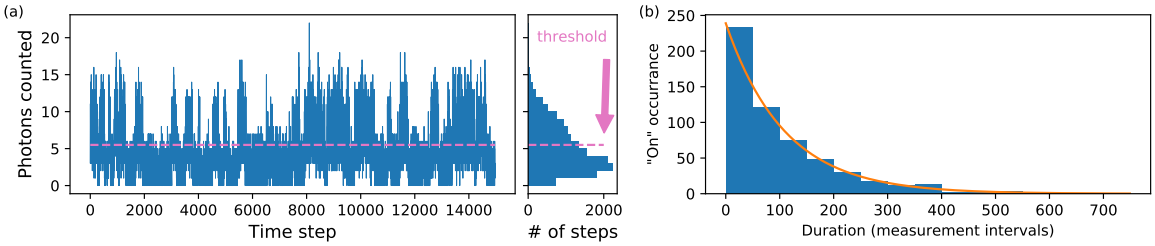


Figure 2.1: Experimental blinking data for a silicon-vacancy centre in nanodiamond material under 30nW excitation, reproduced from [13] with authors' consent. (a) Photon counts were recorded periodically at a certain sampling rate, here normalised to unity. The illustrated threshold was not sufficient, and the authors required multiple thresholds to identify "on" and "off" intervals. (b) Once on and off intervals are identified, the histogram of their durations yields a characteristic lifetime, and hence a blinking rate.

interval given by,

$$P_k = \frac{\bar{n}^k \, e^{-\bar{n}}}{k!}. \tag{2.1}$$

here, $\bar{n}$ is the average number of photons observed in the particular chosen sampling interval.

The simplest and generally accepted model for PL blinking is that there are only two states included — a bright (*on*) and a dark (*off*) state. The most common analysis method for extracting the *on* and *off* states from a blinking time series is threshold analysis [13–16]. Here, an ad-hoc threshold intensity $I_{threshold}$ is chosen to distinguish between *on* and *off* states, as illustrated in Figure 1.1(a). There are various procedures described in the literature to determine a sensible $I_{threshold}$ for a given blinking time trace. Examples include choosing $I_{threshold}$ to be the mid point between the peaks of the *on-off* intensities, and selecting a point about 2–3 times standard deviation higher than the average intensity of the *off* state from the blinking time trace [15, 16].

It can be difficult to find a suitable threshold for certain experimental data. A particularly illustrative example is provided by [13], in which blinking is reported for Silicon-vacancy centres in nanodiamonds. Figure 2.1 reproduces the blinking data summary from [13] with the authors' permission, renormalised to make the sampling interval unity for better comparison with the rest of this thesis. In their original analysis, the authors required multiple thresholds to differentiate the *on* and *off* states to find the switching rates.

Given a threshold, the emitter is taken to be in the *on* state if the intensity I_n in an n^{th} time bin, satisfies $I_n > I_{threshold}$ and it is *off* if $I_n \leq I_{threshold}$. Once the *on* and *off* states are distinguished, then the durations of *on* and *off* intervals can be found from the binned data series [15, 17], and the appropriate parameter distributions can be fitted to this data. This

thesis presents a more direct method for inferring the switching rates that does not rely on selecting an arbitrary threshold intensity.

There have been general proposals for the possible physical mechanisms behind blinking phenomena. For example, Efros and Rosen [34] proposed an explanation for the random fluctuations of a quantum dot (QD). The theory was based on the charge states found in QD's. When a QD is neutral, the electron-hole pair generated by the photo-excitation recombines radiatively leading to photoluminescence, converting to the *on* state of the system. However, photo-excitation can cause Auger ionisation as the electron leaves the QD and stays in surrounding acceptor-like states. This leaves the QD in a charged state. This charged QD influences the surrounding electron-hole pairs and thereby causes rapid Auger-like non radiative recombination [24]. Auger recombinations are faster compared with radiative recombinations. So this suppresses the photoluminescence in charged QDs. The QD remains in the dark state until it get neutralized again. This theory predicted the characteristic durations of *on* and *off* intervals to be exponential. It should be noted that some experiments found a power law behaviour, characterised using conventional methods [14, 24, 31]. It is clear that careful discrimination of exponential and power law dependence is only possible by careful and reliable analysis of the blinking data. The focus of this thesis itself is to develop a new theoretical approach to analyse *blinking* data more efficiently and accurately. We model blinking as a two state Markov process which naturally models a system whose time spent in each *on* and *off* state decays exponentially. The following section gives a detailed description of discrete and continuous-time Markov processes underlying our model of a *blinking* light emitter.

2.2 Markovian Models for blinking

A Markov process is a memoryless stochastic process, and as such makes minimal assumptions about the physical mechanism causing the blinking. The evolution of the two states from a blinking emitter can be modelled using such a process. In particular, for a slow switching process our study assumes a discrete-time Markov process. As an extension we also developed a continuous-time Markov process which will be valid for modelling blinking for slow to fast switching rates. In general, the state of the system is not directly observable, and therefore we talk about a hidden Markov model. This model will be described in the

sections to follow.

2.2.1 Discrete Time Markov Chain (DTMC)

The defining characteristic of a Markov chain is the Markov property, which says the state of the system in the future depends only on the current state of the system and is independent of the past states [20, 21]. This can be mathematically written as,

$$P(X_{n+1} = k|X_n = l_n, X_{n-1} = l_{n-1}, ...X_0 = l_0) = P(X_{n+1} = k|X_n = l_n), \tag{2.2}$$

here, X_n refers to current state at time n, X_{n+1} to the future state and X_0,X_1..X_{n-1} to past states of the system. k, l_n, l_{n-1}.. represents the discrete set of states in a state space S. For a time-homogeneous system,

$$P(X_{n+1} = k|X_n = l) = p_{lk} \quad \text{for all} \quad n \geq 0, l, k \in S. \tag{2.3}$$

p_{lk} is called the transition probability of Markov chain, which gives the probability to jump from a state l to k for all possible values of n. The evolution of the state can be represented by a transition matrix P where, $P_{lk} = [p_{lk}]$, which gives all possible transitions between the states in S. The two properties that a probability transition matrix satisfies are, $p_{lk} \geq 0$ and $\sum_{k \in S} p_{lk} = 1$ for all k in S, this is called a column stochastic matrix. Consider a system with initial state l with a probability distribution s_l. After n transitions the probability distribution of a state k can be found from,

$$s_k = \sum_{l \in S} s_l \, p_{lk}^{(n)}, \tag{2.4}$$

where, $p_{lk}^{(n)} = P(X_n = k|X_0 = l)$, which is the *n-step transition probability*.

The n-step transition probabilities satisfy an equation called **Chapman-Kolmogorov** equations given by [19, 21],

$$p_{lk}^{(n)} = \sum_{r \in S} p_{lr}^{(m)} \, p_{rk}^{(n-m)} \qquad l, k \in S \tag{2.5}$$

here, r is an intermediate state on the trajectory from l to k and occurs at a time $0 \leq m \leq n$. This can be written in matrix notation as $P^{(n)} = [p_{lk}^{(n)}]$ with $P^{(n)}$ as the *n-step transition probability matrix*, which satisfies the **Chapman-Kolmogorov equation**, $P^{(n)} = P^{(m)} P^{(n-m)}$. Here $P^{(n)} = P^n$ where n is the power of P.

As an example consider the relevant *on-off* blinking system. The system can be either in the *on* ($1 \equiv [0\ 1]^T$) or the *off* ($0 \equiv [1\ 0]^T$) state at a given time step. Let the transition

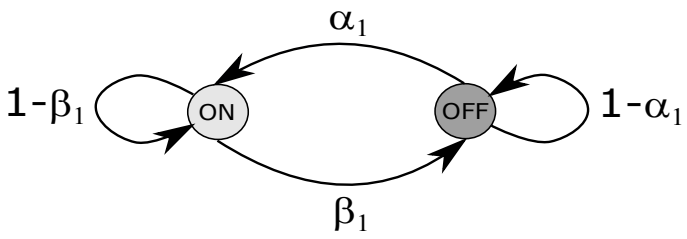


Figure 2.2: Transition graph for an *on-off* system with a DTMC

probabilities be $p_{01} = \alpha_1$, $p_{10} = \beta_1$. This two state Markov chain can be represented using a graph known as transition graph given in Figure 2.2. The nodes represent the possible states of the system and arrows represent the transition probabilities between the states. The transition matrix for the *on-off* system is given by,

$$P = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} 1-\alpha_1 & \beta_1 \\ \alpha_1 & 1-\beta_1 \end{pmatrix} \end{array}. \tag{2.6}$$

With the above matrix we can find the transition probability for the system to be in the state 0 or 1 in next time step given the current state, which is p_{lk}. If P has definite eigenvalues then the n-step transition matrix P^n can be found by first diagonalizing P. If that is the case then there exists a 2×2 matrix S such that, $P = S\,D\,S^{-1}$. Here, D is the diagonal matrix of P. The two eigenvalues of P are $\lambda_1 = 1$ and $\lambda_2 = (1 - \alpha_1 - \beta_1)$, so the terms are given by,

$$S = \begin{bmatrix} \frac{\alpha_1}{\beta_1} & -1 \\ 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 1-\alpha_1-\beta_1 \end{bmatrix}, \quad S^{-1} = \frac{1}{\alpha_1+\beta_1} \begin{bmatrix} \alpha_1 & \alpha_1 \\ -\alpha_1 & \beta_1 \end{bmatrix}. \tag{2.7}$$

The n-state transition matrix can be written by considering the convention, $P^n = S\,D^n\,S^{-1}$, that is,

$$P^n = \frac{1}{\alpha_1+\beta_1} \begin{bmatrix} \frac{\beta_1}{\alpha_1} & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1^n & 0 \\ 0 & (1-\alpha_1-\beta_1)^n \end{bmatrix} \begin{bmatrix} \alpha_1 & \alpha_1 \\ -\alpha_1 & \beta_1 \end{bmatrix}. \tag{2.8}$$

On matrix multiplication the above equation gives,

$$P^n = \frac{1}{\alpha_1+\beta_1} \begin{bmatrix} \beta_1 & \beta_1 \\ \alpha_1 & \alpha_1 \end{bmatrix} + \frac{(1-\alpha_1-\beta_1)^n}{\alpha_1+\beta_1} \begin{bmatrix} \alpha_1 & -\beta_1 \\ -\alpha_1 & \beta_1 \end{bmatrix}. \tag{2.9}$$

Consider one of the eigenvalues of the transition matrix $\lambda_2 = (1 - \alpha_1 - \beta_1)$. Since, $0 \leq \alpha_1, \beta_1 \leq 1$, the limits of λ_2 are $-1 \leq \lambda_2 \leq 1$. The two cases when $\alpha_1 = \beta_1 = 0$ and $\alpha_1 = \beta_1 = 1$ are not interesting because the former condition satisfies a steady state and the later indicates an indefinite swapping of states. Thus $|\lambda_2| < 1$, which implies

$(1-\alpha_1-\beta_1)^n \longrightarrow 0$ as $n \to \infty$. Hence the steady-state limit of the transition probabilities are given by the first term of the Equation 2.9,

$$\lim_{n\to\infty} P^n = \frac{1}{\alpha_1+\beta_1}\begin{bmatrix}\beta_1 & \beta_1\\ \alpha_1 & \alpha_1\end{bmatrix}. \tag{2.10}$$

Here, the two elements in the rows are identical, which implies that the system is in the *on* state with a probability of $\alpha_1/(\alpha_1+\beta_1)$ and in *off* state with a probability of $\beta_1/(\alpha_1+\beta_1)$ regardless of the initial state of the system. That is,

$$\frac{1}{\alpha_1+\beta_1}\begin{bmatrix}\beta_1 & \beta_1\\ \alpha_1 & \alpha_1\end{bmatrix}\begin{bmatrix}q\\ 1-q\end{bmatrix} = \frac{1}{\alpha_1+\beta_1}\begin{bmatrix}\beta_1\\ \alpha_1\end{bmatrix}, \tag{2.11}$$

where, $[q \quad 1-q]^T$ represents the state of the system.

2.2.2 Continuous Time Markov Chain (CTMC)

Continuous Time Markov Chains (CTMC) also have the memoryless property as that the discrete version (DTMC) does. However unlike DTMC, the state of the CTMC is explicitly a function of time. We write the state as $X(t)$ where $t \geq 0$. In general, the state space of CTMC is finite or countably infinite set S. The Markov property for CTMC at each time point for any pair of states $i, j \in S$ can be written as,

$$P(X(t+t') = j \mid X(t') = i, X(u)) = P(X(t+t') = j \mid X(t') = i) = p_{ij}(t), \tag{2.12}$$

where, $t, t' \geq 0$ and $0 \leq u \leq t'$.

Here, $p_{ij}(t)$ is the transition probability from state i to state j after time t. There probabilities can be assembled into a transition matrix $P(t)$ with $[P(t)]_{ij} = p_{ij}(t)$. The transition from one state to another depends on the transition probabilities. We are considering a time-homogeneous process here. For a time-homogeneous process the probability of transition from one state to another depends only on the time difference between the two state. Hence,

$$p_{ij}(t) = P(X(t+t') = j \mid X(t') = i) = P(X(t) = j \mid X(0) = i). \tag{2.13}$$

The transition matrix $P(t)$ has the following properties,

1. If the system is at a particular state initially, a $t = 0$ it must remain in the state, *i.e,* $P(0) = I$

2. The transition probabilities from a state to all other states sum to one, $\sum_{j\in S} p_{ij}(t) = 1$

3. $P(t)$ satisfies the Chapman-Kolmogorov equation, $P(t+t') = P(t)\,P(t')$

The system spends a random duration of time in each state before jumping to another state. The occurrences of jumps only depends on the transition probabilities but not the time spent in a particular state. It turns out that there is a unique continuous distribution to describe the time spent in each state that has the memoryless property and that is an exponential distribution. So the time spent in a state i can be written with holding-time parameter ν_i as $e^{-\nu_i t}$. The time spent in state i (*i.e,* the lifetime of i) decreases with increasing ν_i. Hence ν_i is the total exit rate out of state i. (for $\nu_i = 0$, the system cannot transition to another state.) The holding time parameter ν_i for a state i can be written as the sum of all the rates to other states j, ν_{ij}:

$$\nu_i = \sum_{j\neq i} \nu_{ij} \qquad i \in S \tag{2.14}$$

All these transition rates ν_{ij} can be collected as a matrix Q matrix called an infinitesimal generator or simply generator matrix, given by,

$$[Q]_{ij} = \nu_{ij}, \tag{2.15}$$

where, $i, j \in S$ and $\nu_{ii} = -\nu_i$.

This generator matrix is analogous to the one-step transition matrix of DTMC. An important property associated with the generator matrix is that elements in a row sum to zero,

$$\sum_{j\in S} \nu_{ij} = 0. \tag{2.16}$$

The relation between the generator matrix Q and the transition matrix $P(t)$ is given by the Kolmogorov forward and backward equations. Here $P'(t)$ is the derivative of the matrix elements $p'_{ij}(t)$.

Backward equation: $\frac{d}{dt}P(t) = P'(t) = Q\,P(t)$

Forward equation: $\frac{d}{dt}P(t) = P'(t) = P(t)\,Q$

We can then write the forward equation as,

$$p'_{ij}(t) = \sum_{k\in S} p_{ik}(t)\,\nu_{kj}. \tag{2.17}$$

As a specialisation relevant to the blinking analysis, consider a two state system, say an *on-off* system shown in Figure 2.3. Let r_α ad r_β be the holding time parameter for the *on*

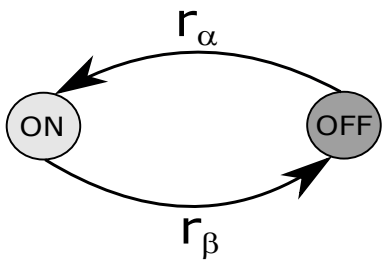


Figure 2.3: Transition graph for an *on-off* system for a CTMC

and *off* states or since there is only one other state to switch to, the rate of switching *on* and switching *off* respectively. As shown in Figure 2.4 each *on* period ends in the *off* state, the time spent in the *on* state can be written as $\exp(-r_\beta\, t)$. Similarly the time spent in the *off* state can be written as $\exp(-r_\alpha\, t)$. The generator matrix Q for the CTMC with $(X(t), t \geq 0)$ is,

$$Q = \begin{bmatrix} -r_\alpha & r_\alpha \\ r_\beta & -r_\beta \end{bmatrix}. \tag{2.18}$$

We can now apply the Kolmogorov differential equations to the two-state model illustrated in Figure 2.3. Here only the forward equations are considered and are given below,

$$p'_{00}(t) = -r_\alpha p_{00}(t) + r_\beta p_{01}(t) \qquad p_{00}(0) = 1 \tag{2.19}$$

$$p'_{01}(t) = r_\alpha p_{00}(t) - r_\beta p_{01}(t) \qquad p_{01}(0) = 0 \tag{2.20}$$

$$p'_{10}(t) = -r_\alpha p_{10}(t) + r_\beta p_{11}(t) \qquad p_{10}(0) = 0 \tag{2.21}$$

$$p'_{11}(t) = r_\alpha p_{10}(t) - r_\beta p_{11}(t) \qquad p_{11}(0) = 1 \tag{2.22}$$

From the property of transition probabilities, transition from a state to all other states sum to one,

$$p_{00}(t) + p_{01}(t) = P(X(t) = 0 \text{ or } 1 | X(0) = 0) = 1, \tag{2.23}$$

$$p_{10}(t) + p_{11}(t) = P(X(t) = 0 \text{ or } 1 | X(0) = 1) = 1. \tag{2.24}$$

Substituting for p_{01} and p_{10} in the forward equations we can determine the transition probabilities $p_{ab}(t)$ from state a to state b in interval Δt:

$$p_{00}(\Delta t) = \frac{1}{r_\alpha + r_\beta}\left(r_\beta + r_\alpha e^{-(r_\alpha + r_\beta)\Delta t}\right), \tag{2.25}$$

$$p_{01}(\Delta t) = \frac{r_\alpha}{r_\alpha + r_\beta}\left(1 - e^{-(r_\alpha + r_\beta)\Delta t}\right), \tag{2.26}$$

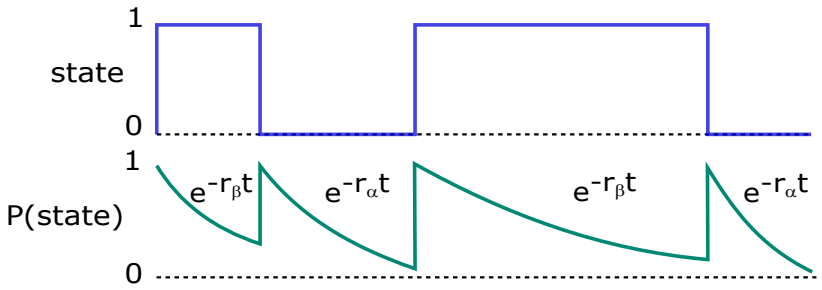


Figure 2.4: Exponentially distributed dwell times for each state.

$$p_{10}(\Delta t) = \frac{r_\beta}{r_\alpha + r_\beta}\left(1 - e^{-(r_\alpha + r_\beta)\Delta t}\right), \tag{2.27}$$

$$p_{11}(\Delta t) = \frac{1}{r_\alpha + r_\beta}\left(r_\alpha + r_\beta e^{-(r_\alpha + r_\beta)\Delta t}\right). \tag{2.28}$$

Relation between transition probability and rate of switching

Let α_1 and β_1 be the transition probabilities between states in a DTMC model. The probability densities of switching *on* and *off* can be written in terms of an exponential distribution as given below,

$$\rho_{\text{on}}(t) = r_\alpha\, e^{-r_\alpha t}, \tag{2.29}$$

$$\rho_{\text{off}}(t) = r_\beta\, e^{-r_\beta t}, \tag{2.30}$$

where, r_α and r_β are the switching *on* and switching *off* rates respectively. From the above equations the probability of staying in *on* and *off* states is,

$$P(\text{not swiching off in the inerval } [0,t]) = 1 - \int_0^t r_\beta\, e^{-r_\beta t'}\, dt' = e^{-r_\beta t} \tag{2.31}$$

$$P(\text{not swiching on in the inerval } [0,t]) = 1 - \int_0^t r_\alpha\, e^{-r_\alpha t'}\, dt' = e^{-r_\alpha t} \tag{2.32}$$

The relationship between the transition probabilities and the rates of switching can be easily obtained from the above two equations are,

$$\alpha_1 = 1 - e^{-r_\alpha t}, \tag{2.33}$$

$$\beta_1 = 1 - e^{-r_\beta t}. \tag{2.34}$$

It is assumed that no switching events are occurring during the interval t.

The basic tools explained in this chapter are used in the construction of blinking models in the following chapters.

3

Logic and Probability theory

In real life we often deal with situations or facts that are not perfectly predictable, such as the question if it will rain tomorrow at 4 pm or not. Incomplete information makes it hard to predict events with certainty. However, what we can do is make a statement of how plausible the event would be. In the same way, scientific studies more often than not deal with parameters that are not directly accessible. With limited information, an exact prediction of the unknown parameters is not possible. However, finding a region of plausibility for the unknown parameters is a viable approach. Probability theory in connection with logical reasoning provides an appropriate tool for quantifying *plausibility*, and is called *Bayesian probabilistic inference*. There are two main key attractive features in Bayesian inference. The first one, is its ability to incorporate prior information for a problem under consideration. The second one, is the possibility of marginalising unwanted parameters from the problem, or the converse, adding additional parameters to an inference. These two make the Bayesian inference a powerful method for inference even in the presence of noise. The key task in the thesis is to infer unknown switching rates from noisy *blinking* data. In such situations

Bayesian inference can provide a robust picture about parameter uncertainties. This chapter deals with tools of probability theory that are necessary for the following chapters.

This chapter is organized in three sections: Section 3.1 introduces the concept of plausible reasoning and its aspects. Section 3.2 introduces Bayesian inference and describes two useful tools for plausible inference, Bayes' rule and marginalisation. Section 3.3 clarifies parameter independencies using a graphical model called a Bayesian network.

3.1 Deductive and Plausible reasoning

Deductive reasoning is a traditional way of determining if a statement is true or false. It uses a series of premises (propositions) to draw conclusions about an event. The use of deductive logic is applicable to situations which are certain. Let us consider the following example:

Premises	**Conclusion**
If A is true, then B is true (major premise) B is false (minor premise)	Therefore A is false

We can summarize this structure with Boolean algebra as $A = AB$, where, the product of propositions AB represents logical AND. It can also be written in terms of implication operation $A \Longrightarrow AB$. Here, if B is true, then A can be either true or false. If B is false then A is also false.

In real situations the information available is often not enough for deductive reasoning. In these situations it is often desirable to argue about the plausibility of conclusion. The plausibility of a conclusion changes as we get more information about the system. For example in many situations people intuitively run arguments like the following:

Premises	**Conclusion**
If A is true, then B is true B is true	A become *more* plausible

In scientific study, we often have to find the plausibility of unknown parameters given experimental observations and prior knowledge. A 'calculus' of plausible reasoning is useful

in such situations. Such a ‘calculus of plausibility’ follows from some conditions known as desiderata.

3.1.1 Desiderata for plausible reasoning

It is a surprising fact that the entire calculus can be pinned down on just three of these desiderata [35, 36],

1. **Degrees of plausibility are represented using real numbers** : If proposition A is more plausible than B, then a greater number is assigned to A than B. Assigning plausibility to a proposition A also depends on whether it is conditioned on any other proposition (say C). This conditional plausibility can be expressed symbolically as, A|C. This is called the “conditional plausibility of A is true given C is true" or simply “A given C". If there is another proposition B|C, which is less plausible than A|C, then B|C will be smaller than $A|C$ ($B|C < A|C$).

2. **Qualitative agreement with common sense** : The number representing the plausibility of a proposition must increase continuously and monotonically as we get more evidence that supports the truth of the proposition. A larger number assigned to a proposition represents its greater plausibility. Also in the limit of certainty it should agree with logic.

3. **Consistency** :There are three sub-parts and each stands for a common colloquial meaning of the word consistency. They are (i) Every possible way that a conclusion can be reasoned out must give the same plausibility, (ii) All information relevant to the question should be taken into consideration, (iii) Equal plausibility are assigned to equal states of knowledge.

3.1.2 Rules for manipulating plausibility

Arguing from the desiderata it is possible to arrive at the rules of plausibility reasoning [35, 36] and they are,

- $0 \leq P(A|B) \leq 1$
- **product rule**: $p\,(A\ B \mid C) = p\,(A \mid B\,C)\ p\,(B \mid C)$

- **sum rule**: $p\,(A + B \mid C) = p\,(A \mid C) + p\,(B \mid C) - p\,(A\,B \mid C)$

where $P(A|B)$ denotes the plausibility of proposition A given the assumption proposition B is true.

The function p defines the degree of plausibility of a proposition. These rules are clearly the same as those obeyed by probability theory [35, 37]. So we arrive at probability theory as the 'calculus of plausibility'. These foundations of probability theory are known as *Bayesian probability theory* or more specifically *extended logic*. Inference based on Bayesian probability theory gives a clear picture about parameters from noisy data. The following sections give a detailed discussion of Bayesian inference.

3.2 Bayesian Inference

The primary aim of the present study is to infer the unknown switching rates from blinking data. Utilizing Bayesian inference as the main tool. The probability distribution of unknown model parameters that are being calculated are called posterior probability distribution. A simple rule that is utilized to estimate the posterior probability distribution of the model parameters is called *Bayes' rule*, which can be directly derived from the above-described sum and product rules. *Bayes' rule* lets us switch the parameter being infered and the conditional parameter. Bayesian inference makes use of *Bayes' rule* to update the probability distributions of model parameters as new information becomes available. A detailed account of the process to be utilized in this study is given in the sub-sections. A more detailed description on Bayesian Inference and its scope can be found in literature [35, 36].

3.2.1 Bayes' rule and Marginalisation

The two main pillars of Bayesian inference are *Bayes' rule* and *marginalisation*. These can be deduced from the **sum** and **product** rules as follow:

Bayes' rule: Bayes' rule is a very useful tool in data analysis. The ultimate reason for its usefulness is the fact that it swaps the 'givens' with the propositions. Bayes' rule follows directly from the product rule and is given by:

$$p\,(A \mid B\,C) = \frac{p\,(A \mid C)\; p(B \mid A\,C)}{p\,(B \mid C)}. \tag{3.1}$$

The above equation can be easily obtained by equating the expansion of $p(A\ B \mid C)$ and $p(B\ A \mid C)$ sing the product rule. Now replace A, B and C in Equation 3.1 with propositions describing the model parameters (M), experimental data (D) and prior information (I) respectively. Then we can use Bayes' rule to calculate the plausibility of the model parameters, given the experimental data and the prior information:

$$p(M \mid D\ I) = \frac{p(M \mid I)\ p(D \mid M\ I)}{p(D \mid I)}, \tag{3.2}$$

here, $p(M \mid D\ I)$ is the posterior probability distribution of M, given D and I, $p(M \mid I)$ is the prior probability distribution of the M before observing any data, $p(D \mid M\ I)$ is the likelihood, *i.e.* probability of observing D, given M and I are true, $p(D \mid I)$ is the prior predictive probability for D. Here $p(D \mid I)$ is the normalisation factor which ensures $p(M \mid D\ I) = 1$. Hence, Equation 3.2 can be also written as $p(M \mid D\ I) \propto p(M \mid I)\ p(D \mid M\ I)$.

Marginalisation: Marginalisation is a procedure that helps eliminating the unwanted parameters from a model, *i.e.* it handles the nuisance parameters emerging in the specific model underlying. Nuisance parameters are occasionally important for analysis but not interesting for the final conclusions. The marginalisation equation is obtained from the sum and product rules and reads as follows:

$$p(A \mid C) = \int_{-\infty}^{+\infty} p(A\ K \mid C)\ dK. \tag{3.3}$$

Suppose if we are trying to find the probability of the proposition A and we know there is another parameter K on which A depends. The $p(A \mid C)$ is calculated irrespective of the true value of K by considering all possible values $K_1, K_2, ..K_n$ that K can take. We can write the probability of A as,

$$p(A \mid C) = \sum_{i=1}^{n} p(A\ K_i \mid C). \tag{3.4}$$

In real situations, our focus would be on finding the probability of the model parameter (M). But, often we need to deal with nuisance parameters appearing in the analysis (for example background noise in the signal, say γ). Provided experimental data (D) and background information (I), the marginalisation procedure can be used to get rid of γ. That is,

$$p(M \mid D\ I) = \int_{\gamma_1}^{\gamma_n} p(M\ \gamma \mid D\ I)\ d\gamma. \tag{3.5}$$

where the summation in Equation 3.4 has been changed to integration for continuous γ.

3.2.2 Prior probability distribution

The prior probability distribution describes the distribution of the model parameters before observing some data. Basically there are two kind of priors, one is called an informative prior and other is non-informative prior.

As the name indicates, the informative prior directly influences the posterior distribution. For example the prior could contain information from previous studies, prior information about the model etc. We can use the posterior distribution as prior, while we update the model with a new set of data.

The non-informative prior is often used when the prior knowledge about the model is limited. Then it is usual to assign equal probability for all parameters. A more sophisticated approach would be that a non-informative prior assumes certain symmetries that are then represented by the probability assignments. The simplest example would be that the state of knowledge is invariant under permutations of labels. This symmetry implies that we should assign equal probability to all labels.

3.2.3 Parameter estimation

Parameter estimation refers to the estimation of values of parameters of the model from the data [36]. In the blinking problem, the switching rates of a blinking system are the model parameters that we want to estimate. Bayes' theorem forms a posterior distribution of model parameters: $p(\theta|D,I)$. The posterior distribution can then be used to form summaries about the model parameters by using point estimates or interval estimates. The point estimate gives a single most probable value of the parameter. The thesis uses interval estimates for inference, called *credible regions*. With credible regions we can find a region containing the true value of the parameter with a specific probability from a posterior probability distribution [36]. For example, a 95% credible region contains the true value with 95% probability. The posterior density region R is then defined by the value C assigned the desired probability of the credible region such that,

$$\int_R p(\theta|D,I)\ d\theta = C. \tag{3.6}$$

3.3 Bayesian networks

Bayesian networks (BNs) denote the graphical representation of variables and their relations in a probabilistic domain. Such graphical models make it easier to visually tackle the conditional dependencies and independencies between variables and describe their probabilistic distributions. BNs are directed acyclic graphs (DAGs), where each edge in such a graph has a direction and there are no cycles [38]. The nodes represent variables and edges reflect the immediate dependencies. An example of a BN is shown in Figure 3.1. A BN can be used to determine conditional independencies between the variables of a problem, using the property of d-separability, a concept that will be explained further below.

3.3.1 Independencies from Bayesian network

Consider a DAG given in Figure 3.1. Here, a 'direct casual influence' is assumed between variables that are connected with edges. As an example, *Rain (R)* is a direct cause of *Wet grass (G)*. Common notations for BNs associated with a variable X are as folows [38]:

- Parents(X) : The parents of X are a set of variables, which form direct edges to variable X. For example, in Figure 3.1 parents of variable G are R and S.
- Children(X) : The children of X are a set of variables to which edges from X are connected. For example, variable W is a child of variable G in Figure 3.1.
- Descendant(X) : These are the children of X and their children and grand children etc.
- Non-descendants(X) : Non-descendant are a set of variables excluding Parents(X) and Descendant(X). For example, in Figure 3.1, variables R and C are non-descendants of variable S.

With this notation the general form of local independence statements can be written as,

$$\text{X} \perp \text{Non-descendants(X)} \,|\, \text{Parents(X)}.$$

This can be read as X is conditionally independent of its Non-descendants given its Parents (direct causes of X). The variables with no parents are said to be marginally independent of their Non-descendants. For example, from Figure 3.1 R and S have no parents, hence $R \perp S$ and $S \perp RC$. The other independencies in Figure 3.1 are $W \perp RCS \,|\, G$, $C \perp SGW \,|\, R$

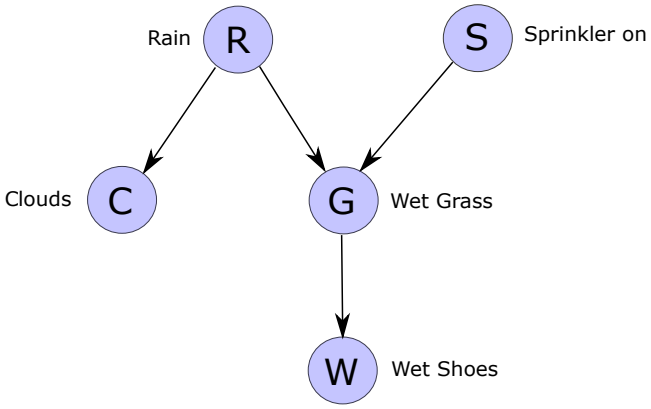


Figure 3.1: A Bayesian network: the nodes represents variables and their dependencies are displayed by the connected edges.

and $G \perp C \mid S\,R$. For example, if we are interested about wet shoes (W) and it is given that grass is wet (G), then all extra information about Rain (R) or sprinkler on (S) are not any more informative. Even they have influence on the wet shoes (W), the information about wet grass (G) will provide all the information needed.

3.3.2 Joint probability distribution from Bayesian network

BN are a graphical representation of a factorization of the joint probability distribution over all the parameters. The factors of the joint probability distribution are of the form of conditional probabilities, $p(X|\,\text{Parents}(X))$. The full joint probability distribution is given by the following chain rule of BNs [39, 40],

$$P(X_1, X_2..X_n) = \prod_i P(X_i \mid \text{Parents}(X)). \tag{3.7}$$

Using Equation 3.7, the joint probability distribution of Figure 3.1 can be written as,

$$P(R\,C\,S\,G\,W) = P(W \mid G)\,P(G \mid R\,S)\,P(C \mid R)\,P(R)\,P(S). \tag{3.8}$$

3.3.3 d-separation

d-separation denotes a method of testing weather the independencies between two variables A and C given B holds in a given BN, and goes beyond the local independencies mentioned earlier. Two variables A and C are said to be d-separated given B, if all nodes between A and C are blocked by B. The d-separation test can be used to find the indepedencies in a joint probability distribution from a BN. Hence with a d-separation test for a set of variables A, B,

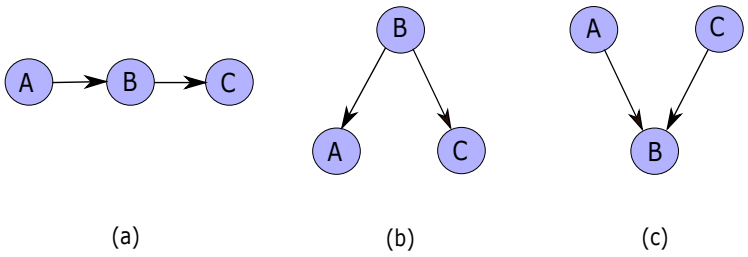


Figure 3.2: Bayesian networks showing three types of connections defining d-separation. (a) Indirect casual effect (b) Common cause. (c) Common effect.

C, one can write, $P(A \mid B\,C) = P(A \mid B)$, if all paths between A and C are blocked by B. The dependencies between A and C change as the information about B becomes available [41].

In the simple case of direct connection, A and C are directly connected through an edge (A $\rightarrow$ C). It is easy to see that there is a direct correlation between A and C regardless of any other evidence, say B.

In an indirect connection, two variables whose dependencies are considered are not connected with a single edge, but via other variables. For example consider a simple case with three variables, A, B and C, as illustrated in Figure 3.2. There are mainly three types of possible connections,

(a) Indirect casual effect (sequential value) with a sequential connection between A, B and C as shown in Figure 3.2 (a), the information can flow from A to C if B is unknown. However, the presence of evidence about B can block the information flow from A to C. Hence, $p(C|B,A) = p(C|B)$. The Joint probability distribution for the BNs can be written as,

$$p(A,B,C) = p(B|A)\ p(A)\ p(C|B). \tag{3.9}$$

(b) In the case of the common cause (divergent value), B is a parent, both A and C depend on B, and there is no direct connection between A and C as depicted in Figure 3.2 (b). Here A and C are conditionally independent given B. That is $p(A|B,C) = p(A|B)$. Also the symmetry of the independencies follow, $p(C|B,A) = p(C|B)$. With the BN, it is easy to write the joint probability distribution of variables,

$$p(A,B,C) = p(A|B)\ p(B)\ p(C|B). \tag{3.10}$$

(c) For the common effect (convergent value), A and C are parents of B as shown in Figure 3.2 (c). Unlike in the above cases, information cannot flow between A and C, if

B is not observed. That is, A can influence C only if B is known.

The conditional independencies obtained from the Bayesian network are used in the following chapters to find the independencies between model parameters for the blinking problem.

4

Discrete-time Markov Chain Models for a Blinking Emitter

A critical step in making an inference on the blinking rates from observed fluorescence data is the construction of a model of how the counts arise including all the sources of noise. Throughout the thesis, the blinking emitter is considered to have only two relevant states that affect the fluorescence levels: an *on* state and an *off* state. Further we assume that there is some mechanism that causes the emitter to switch between these states. In the present work, various models for blinking are constructed, but in general they will all be Markov processes. This chapter discusses modelling a blinking emitter with a Discrete-time Markov Chain (DTMC) and conducting the inference accordingly. In DTMC, a system is considered to be in a particular state for the entire interval of a time step.

This chapter is organized in sections, with Section 4.1 focuses on a simple model for the blinking emitter in which the state of the system fixed for the entire sampling interval and can only switch at the boundaries of the sampling interval. Section 4.2 consider the model

where the sampling interval spans multiple time steps.

4.1 Discrete-time Markov Chain - *Single-step Model*

As discussed in section 2.2.1, a two-state Markov chain can be used to model a blinking emitter. The key point is that the current state of the emitter is sufficient to determine the future dynamics, *i.e.* the process has no memory. For simplicity, this chapter starts out by considering the state to be fixed for the entire sampling interval. Switching events only occur at the boundaries of the sampling intervals.

4.1.1 DTMC single-step model description

Consider an emitter which can change its state only at the boundaries of sampling intervals. We denote the switch-on and switch-off probabilities as α_1 and β_1, respectively. Later, the subscript will represent the number of time steps in a given sampling interval. The graphical representation of the model is depicted in Figure 4.1.

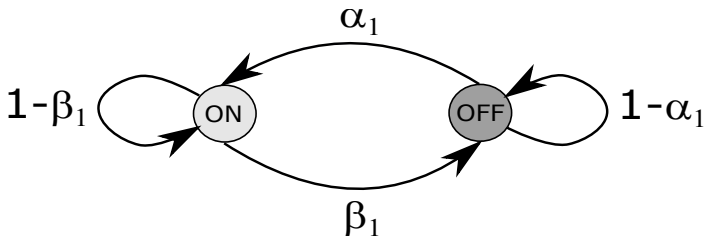


Figure 4.1: A discrete Markov process for state evolution of the blinking emitter with one time-step per sampling interval. α_1 and β_1 are the switching *probabilities* at each time step.

The blinking model is also parametrized by the fluorescence and background photon count rates λ and μ respectively. If the emitter is in an *off* state then the detector can still register counts due to noise processes in the detector itself or the associated counting electronics known as dark-counts. These counts are well-modelled by a Poissonian process with a rate μ. While fluorescing, it is assumed that the detector registers photon counts in a time scale which is much longer than the quantum regime. Then the photon counts are also well modelled by a Poissonian process (i.e. shot noise) with an overall rate λ that includes detector inefficiencies. The *observed* counts are then Poissonian with a rate of either μ or $\mu + \lambda$ since the combination of two independent Poissonian processes is again Poissonian.

Given the state s_{n-1} at the beginning of time-step n, and λ and μ, the probability of seeing c_n counts as given in Equation 2.1 is

$$P(c_n \mid s_{n-1}\, \lambda\, \mu\, I_1) = \begin{cases} \frac{e^{-\mu}\, \mu^{c_n}}{c_n!} & s_{n-1} = 0 \\ \frac{e^{-(\lambda+\mu)}\, (\lambda+\mu)^{c_n}}{c_n!} & s_{n-1} = 1 \end{cases} \tag{4.1}$$

where $s_{n-1} = 1$ represents the *on* state and $s_{n-1} = 0$ represents the *off* state of the system. I_1 tags the single-interval model as the background information. Clearly, if we knew the state of the emitter at each data point, it would be a simple matter to infer the switching rates. Unfortunately, the states are not directly observable, and worse still, because the states are not observed, the switching probabilities become dependent on the entire history of the count data. This makes the inference considerably more involved.

Bayesian network for the DTMC single-step model

We can summarize the dependencies amongst the variables by the Bayesian network (BN) [42] shown in Figure 4.2. In the BN, s_{n-1} represents the initial state of the emitter at the n^{th} sampling interval. For the single-step model I_1, the initial state extends for the *whole* duration of the interval.

The key variable independencies are the following

$$\alpha_1 \perp \beta_1 \perp \lambda \perp \mu \mid I_1, \tag{4.2}$$

$$s_n \perp s_u \mid s_{n-1}\, \Omega_1 \quad (u < n-1), \tag{4.3}$$

$$c_n \perp c_u\, s_v \mid s_{n-1}\, \Omega_1 \quad (u \neq n,\, v \neq n), \tag{4.4}$$

where $\Omega_1 = \alpha_1\, \beta_1\, \lambda\, \mu\, I_1$ for compactness and $a \perp b \mid c$ denotes a is indepenedent of b given c. We will make use of these independencies in the following section.

4.1.2 Bayesian Inference on the DTMC single-step model

Ultimately, the quantity that needs to be inferred is $P(\alpha_1\, \beta_1 \mid \vec{c}\, I_1)$, which is the probability distribution of α_1 and β_1 given photon counts $\vec{c}$ and background information about single-step model I_1. Using Bayes' rule, the posterior probability distribution of α_1 and β_1 can be written as,

$$P(\alpha_1\, \beta_1 \mid \vec{c}\, I_1) = \frac{P(\alpha_1\, \beta_1 \mid I_1)\, P(\vec{c} \mid \alpha_1\, \beta_1\, I_1)}{P(\vec{c} \mid I_1)}. \tag{4.5}$$

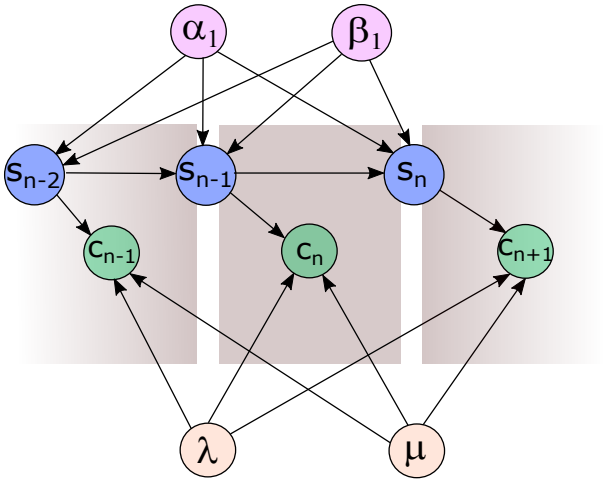


Figure 4.2: A partial Bayesian network representing the joint probability distribution of problem parameters. The pattern of nodes and connections inside the central shaded region representing sampling intervals, which are repeated for each data value. The parameters of interest are α_1, β_1, λ, and μ, given the observed counts c_n. Enough of the full network is drawn to be able to easily determine the variable independencies. In the case of a single step over the sampling time, s_{n-1} is the state over the entire sampling interval.

The rates of fluorescence and background counts, λ and μ, can be added and then subsequently removed by marginalisation as discussed in Section 3.2.1,

$$P(\alpha_1\,\beta_1 \mid \vec{c}\,I_1) = \int d\lambda \int d\,\mu\, \frac{P(\alpha_1\,\beta_1 \mid I_1)\,P(\vec{c}\,\lambda\,\mu \mid \alpha_1\,\beta_1\,I_1)}{P(\vec{c} \mid I_1)} \tag{4.6}$$

$$= \int d\lambda \int d\,\mu\, \frac{P(\alpha_1\,\beta_1 \mid I_1)\,P(\vec{c} \mid \lambda\,\mu\,\alpha_1\,\beta_1\,I_1)\,P(\lambda\,\mu \mid \alpha_1\,\beta_1\,I_1)}{P(\vec{c} \mid I_1)}. \tag{4.7}$$

The above equation can be simplified using the product rule to give,

$$P(\alpha_1\,\beta_1 \mid \vec{c}\,I_1) = \int d\lambda \int d\mu\, \frac{P(\alpha_1\,\beta_1\,\lambda\,\mu|I_1)\,P(\vec{c} \mid \Omega_1)}{P(\vec{c} \mid I_1)}. \tag{4.8}$$

Given independencies in 4.2, we can write,

$$P(\alpha_1\,\beta_1\,\lambda\,\mu \mid I_1) = P(\alpha_1 \mid I_1)\,P(\beta_1 \mid I_1)\,P(\lambda \mid I_1)\,P(\mu \mid I_1), \tag{4.9}$$

these are the prior probability distributions of $\alpha_1, \beta_1, \lambda$ and μ, and can be taken as constant over some initial range so that,

$$P(\alpha_1\,\beta_1 \mid \vec{c}\,I_1) = \frac{1}{\mathcal{N}} \int d\lambda \int d\mu\, P(\vec{c} \mid \Omega_1), \tag{4.10}$$

where, $\mathcal{N} = P(\vec{c} \mid I_1)/P(\alpha_1\,\beta_1\,\lambda\,\mu \mid I_1)$ is the normalisation factor and has to be determined at the end.

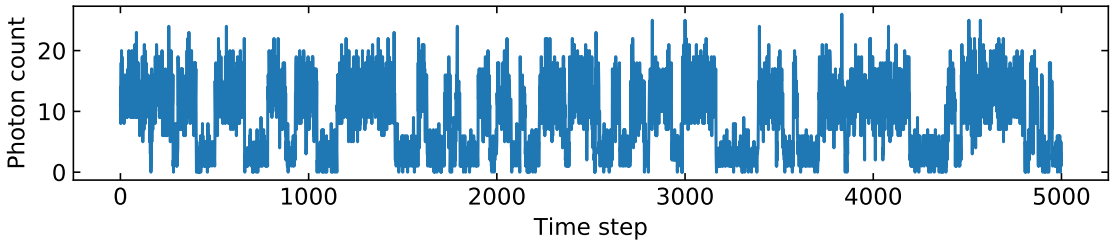


Figure 4.3: Simulated blinking time trace with very low switching rates. The *on* and *off* states are clear from the time trace.

If all the states $\vec{s} \equiv s_0\, s_1 \ldots s_{N-1}$ were known together with Ω_1, the probability of all the counts given Ω_1 ($P(\vec{c}\,|\,\Omega_1)$) would be easily determined. As before, we can use he same trick of adding parameters then marginalising over them. We therefore obtain,

$$P(\vec{c}\,|\,\Omega_1) = \sum_{\vec{s}} P(\vec{c}\,\vec{s}\,|\,\Omega_1) = \sum_{\vec{s}} P(\vec{c}\,|\,\vec{s}\,\Omega_1) P(\vec{s}\,|\Omega_1), \tag{4.11}$$

where $\sum_{\vec{s}} = \sum_{s_0} \ldots \sum_{s_{N-1}}$. Using the independencies in 4.4, $P(\vec{c}\,|\vec{s}\,\Omega_1)$ can be simplified to

$$P(\vec{c}\,|\,\vec{s}\,\Omega_1) = \prod_{n=1}^{N} P(c_n\,|\,s_{n-1}\,\Omega_1), \tag{4.12}$$

where each individual term is determined by Equation 4.1.

The remaining term $P(\vec{s}\,|\Omega_1)$ cannot be simply factorised over the states despite being a Markov chain, as observing Ω_1 introduces possible dependencies. However it can be expanded using the product rule and simplified by making use of the independencies in 4.3:

$$P(\vec{s}\,|\,\Omega_1) = \prod_{n=1}^{N} P(s_n\,|\,s_{n-1}\,\Omega_1)\, P(s_0\,|\,\Omega_1), \tag{4.13}$$

where the terms are expanded in temporal order repeatedly using the product rule. The first term in the RHS of the above equation is the transition probability of the Markov chain from a state s_{n-1} to a state s_n. The term $P(s_0\,|\,\Omega_1)$ is the probability of the initial state s_0 and can be taken as equal for both states *on* and *off*.

Finally combining the above equations, the inference becomes

$$\begin{aligned} P(\alpha_1\,\beta_1|\,\vec{c}\,I_1) = &\frac{1}{\mathcal{N}} \int d\lambda \int d\mu \sum_{\vec{s}} \prod_{n=1}^{N} P(c_n\,|\,s_{n-1}\,\Omega_1) \\ &\times P(s_n\,|\,s_{n-1}\,\Omega_1)\, P(s_0\,|\,\Omega_1). \end{aligned} \tag{4.14}$$

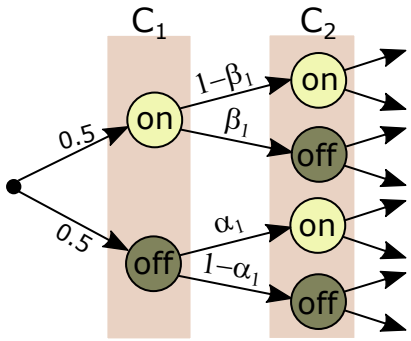


Figure 4.4: Number of plausible histones of the state as the counts are accumulated.

Equation 4.4 can then be used to write the joint distribution between c_n and s_n,

$$P(\alpha_1\,\beta_1|\vec{c}\,I_1) = \frac{1}{\mathcal{N}}\int d\lambda \int d\mu \sum_{\vec{s}}\prod_{n=1}^{N} P(c_n\,s_n|s_{n-1}\,\Omega_1)\times P(s_0|\Omega_1). \tag{4.15}$$

The problem with Equation 4.15 is that the sum over $\vec{s}$ contains 2^N terms each of which has N products. This will rapidly become intractable as the size of the data grows, as illustrated in Figure 4.4. Fortunately, it is possible to rewrite Equation 4.15 as a *single* term with N 2×2 matrix products. Consider the following matrix:

$$R_n = \begin{bmatrix} P(c_n\,s_n=0|s_{n-1}=0\,\Omega_1) & P(c_n\,s_n=0|s_{n-1}=1\,\Omega_1) \\ P(c_n\,s_n=1|s_{n-1}=0\,\Omega_1) & P(c_n\,s_n=1|s_{n-1}=1\,\Omega_1) \end{bmatrix}, \tag{4.16}$$

and vector

$$\vec{D}_0 = \begin{bmatrix} P(s_0=0|\Omega_1) \\ P(s_0=1|\Omega_1) \end{bmatrix}, \tag{4.17}$$

then

$$\sum_{\vec{s}}\prod_{n=1}^{N} P(c_n\,s_n|s_{n-1}\,\Omega_1)P(s_0|\Omega_1) = [1\;1]\prod_{n=1}^{N} R_n\vec{D}_0, \tag{4.18}$$

where the product on the RHS is in decreasing n to the right. The equivalence in Equation 4.18 is readily verified by expanding a few terms out. The matrix multiplication will sum over the columns of R_n which is a summation over the states s_{n-1}. Each successive multiplication sums over another state and the final multiplication by the row vector $[1\;1]$ will sum over s_N. With this equivalence, the inference reads

$$P(\alpha_1\,\beta_1\,|\,\vec{c}\,I_1) = \frac{1}{\mathcal{N}}\int d\lambda \int d\mu\,[1\;1]\prod_{n=1}^{N} R_n\vec{D}_0, \tag{4.19}$$

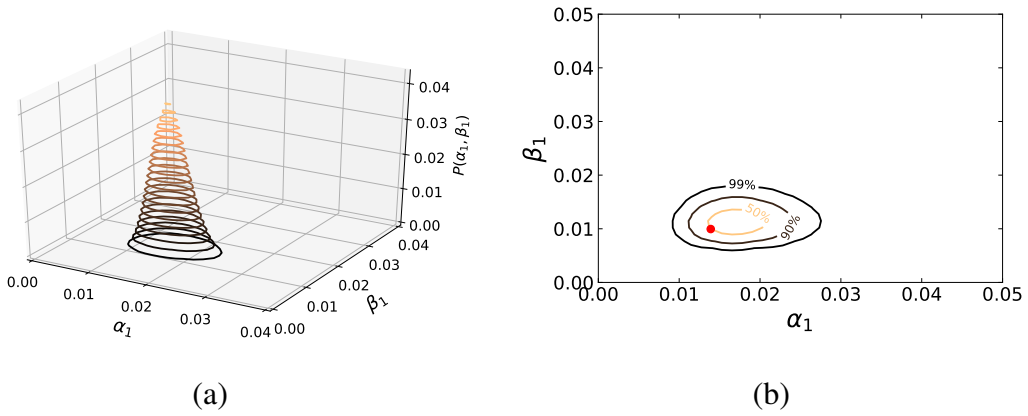


Figure 4.5: (a) Contour plot showing the posterior probability distribution for α_1 and β_1 obtained from the inference on the simulated time trace shown in Figure 4.3. (b) The credibility regions obtained from the inference, contain the 50%, 90%, and 99% credible regions. The red dot shown inside the credibility regions represents the true values of the parameters used in the simulation.

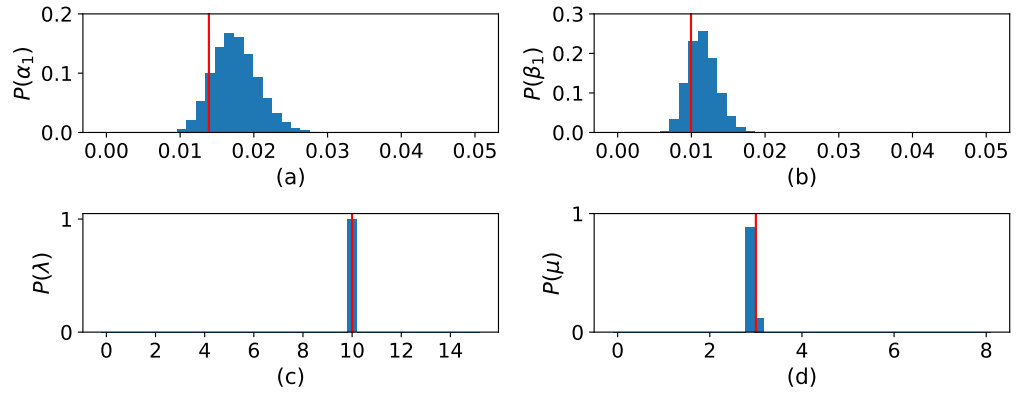


Figure 4.6: panels (a) and (b) shows the posterior probability distributions of α_1 and β_1 respectively. Panel (c) and (d) displays the marginal distributions of λ and μ respectively. The red line represents the true value used for simulation of the blinking time trace.

which is easily computable. Note that care must be taken to avoid underflow or overflow in calculating the matrix products.

As a demonstration of the algorithm, for the data given in Figure 4.3 (the data is simulated with low rates of switch *on* (0.014) and switch *off* (0.01)), we infer α_1 and β_1 and obtain the posterior probability distribution in Figure 4.5 (a). To characterise the uncertainties in the inference we plot several credible regions of the posterior distribution in Figure 4.5 (b). Figure 4.6 shows the marginal distributions found from the inference. The red line in the figure indicates the true value used for simulation. From the figure it is clear that the inference gives peaks very close to the true values. Running a longer simulation will produce a tighter

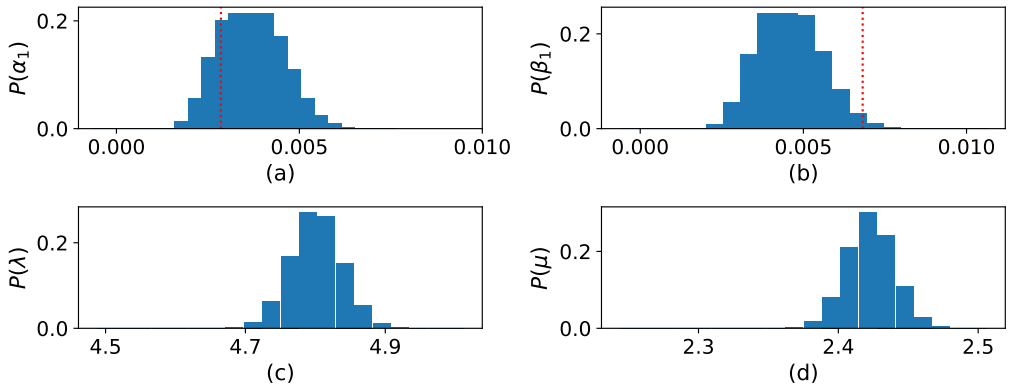


Figure 4.7: panels (a) and (b) shows the posterior probability distributions of α_1 and β_1 respectively inference done on the experimental data given in Figure 1.1 (c). The red dotted line is the values the authors found in the original paper using a double threshold technique. Panel (c) and (d) displays the marginal distributions of λ and μ respectively.

convergence.

Experimental blinking data for a silicon vacancy centre Figure 2.1 was also analysed and is plotted in Figure 4.7. In this case the red dotted line is the values that the authors found in using a double threshold technique for analysis (the data was obtained from the corresponding author). The inference gives similar values (although suggesting a smaller β_1) and also carries an indication of the uncertainty by providing the entire posterior distribution. This figure shows the applicability of the method discussed here in real situations. The single-interval model discussed here is most applicable of situations where the switching rates are low. This model fails for cases with high switching rates as we will see next.

4.2 Discrete-time Markov Chain - *Multi-step Model*

In the previous section 4.1, the emitter could only switch states at the boundaries of the sampling interval. Though this is a hidden Markov model, the inference was simple as the state is only obfuscated by the Poissonian distribution of counts. The model presented is expected to be valid if the switching rates per sampling interval are not too large. So that most sampling intervals are either wholly *on* or wholly *off*. If the switching rates get too large then this is clearly not the case as there is significant probability that the emitter switched during the sampling interval possibly more than once per sampling interval. This section successively approaches faster rates by considering the sampling interval as subdivided into

more and more sub-intervals during which the emitter is wholly *on* or *off*.

4.2.1 DTMC Multi-step model description

For an emitter it is possible to assume that the total number of photon counts observed in a sampling interval is the sum of all counts that are registered in each sub-interval within the sampling interval. This kind of system which switches its state within the sampling interval can also be modelled by a DTMC. Here we assume equally spaced sub-intervals within each sampling interval and the probability of transition of states between these sub-intervals are α_d (probability of switching *on*) and β_d (probability of switching *off*), as illustrated in Figure 4.8.

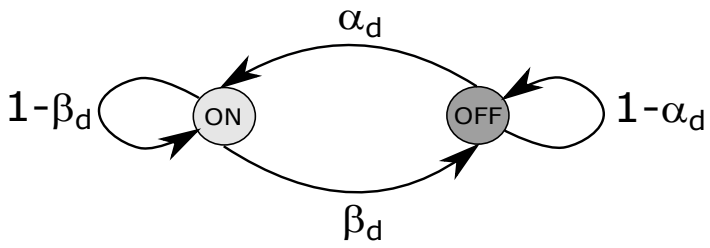


Figure 4.8: A discrete Markov process modelling the evolution of the state with d time-steps per detection interval. In this case, α_d and β_d are the switching *probabilities*.

The parameters of interest are rates of switch *on* (r_α) and switch *off* (r_β) per sampling interval. The switch *on* and switch *off* probabilities α_d and β_d can be written in terms of r_α and r_β as,

$$\alpha_d = 1 - e^{-r_\alpha T/d}, \tag{4.20}$$

$$\beta_d = 1 - e^{-r_\beta T/d}, \tag{4.21}$$

where T is the length of the sampling interval and d is the number of sub-intervals.

The counts arising in each of the sub-intervals can be considered as following a Poisson distribution with a rate $\lambda + \mu$ when the emitter is *on* and a rate μ when the emitter is *off*, as previously discussed in section 4.1. The Bayesian network for the DTMC multi-step model is drawn in Figure 4.9. The independencies arising from the BN can be written as,

$$r_\alpha \perp r_\beta \perp \lambda \perp \mu \,|\, I_d, \tag{4.22}$$

$$s_n \perp s_u \,|\, s_{n-1}\, \Omega_d \quad (u < n-1), \tag{4.23}$$

$$c_n \perp c_u \,|\, s_{n-1}\, s_n\, \Omega_d \quad (u \neq n), \tag{4.24}$$

where, $\Omega_d = r_\alpha\, r_\beta\, \lambda\, \mu\, I_d$ for simplicity.

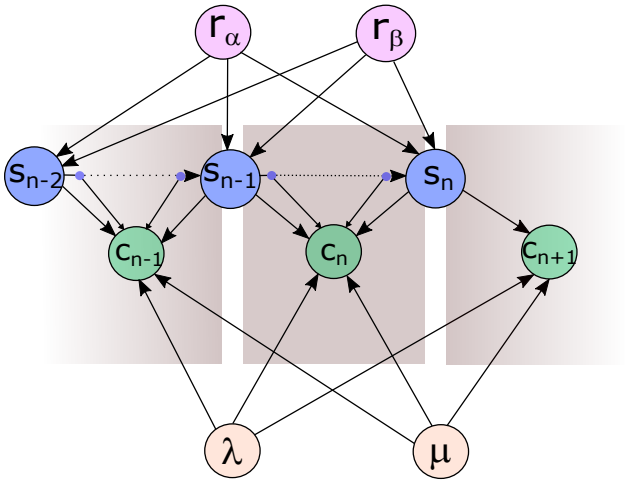


Figure 4.9: BN for the multi-step DTMC model. Each sampling interval consists of equally spaced sub-intervals, which can be *on* or *off* depending upon the transition probabilities α_d and β_d. The states s_{n-1} and s_n represent the state of the emitter at the boundaries of the n^{th} sampling interval.

4.2.2 Bayesian Inference on the DTMC multi-step model

Here we are inferring the rates of switch-on r_α and switch-off r_β given the data counts $\vec{c}$. As in the previous section 4.1.2, we use Bayes' rule to find the posterior probability distribution of r_α and r_β,

$$P(r_\alpha\, r_\beta \mid \vec{c}\, I_d) = \frac{P(r_\alpha\, r_\beta \mid I_d)\, P(\vec{c} \mid r_\alpha\, r_\beta\, I_d)}{P(\vec{c} \mid I_d)}, \tag{4.25}$$

where, I_d denotes the underlying d sub-interval model. The unknown fluorescence (λ) and background (μ) rates can be marginalized out as in section 4.1.2, and we obtain an expression analogous to Equation 4.10,

$$P(r_\alpha\, r_\beta \mid \vec{c}\, I_d) = \frac{1}{\mathcal{M}} \int d\lambda \int d\mu\, P(\vec{c} \mid \Omega_d), \tag{4.26}$$

where, $\mathcal{M} = P(\vec{c} \mid I_d)/P(r_\alpha\, r_\beta\, \lambda\, \mu \mid I_d)$ is the normalisation factor to be determined at the end. Further we note,

$$P(r_\alpha\, r_\beta \mid \vec{c}\, I_d) \propto P(\vec{c} \mid \Omega_d). \tag{4.27}$$

The probability of counts observed depends on the state of the emitter. Here, the states s_{n-1} and s_n are the boundary states at the n^{th} sampling interval, which again can be included and then marginalized out,

$$P(\vec{c} \mid \Omega_d) = \sum_{\vec{s}} P(\vec{c}\, \vec{s} \mid \Omega_d) = \sum_{\vec{s}} P(\vec{c} \mid \vec{s}\, \Omega_d)\, P(\vec{s} \mid \Omega_d). \tag{4.28}$$

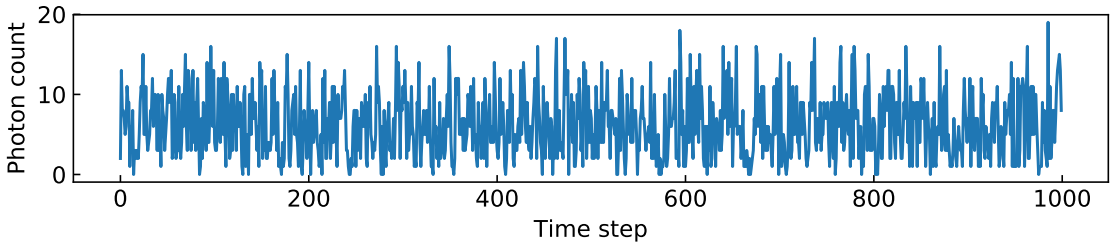


Figure 4.10: Simulated blinking time trace with high rate of switching between *on* and *off* states. This time trace is used to demonstrate the working of the multi-step model.

Using the independencies given in 4.23 and 4.24, we can write,

$$P(\vec{c} \mid \Omega_d) = \sum_{\vec{s}} \prod_{n=1}^{N} P(c_n \mid s_n\, s_{n-1} \Omega_d)\, P(s_n \mid s_{n-1}\, \Omega_d)\, P(s_0 \mid \Omega_d) \tag{4.29}$$

$$= \sum_{\vec{s}} \prod_{n=1}^{N} P(c_n\, s_n \mid s_{n-1}\, \Omega_d)\, P(s_0 \mid \Omega_d). \tag{4.30}$$

This equation can be converted into a matrix form identical to the one given in Equation 4.18, that is,

$$\sum_{\vec{s}} \prod_{n=1}^{N} P(c_n\, s_n \mid s_{n-1}\, \Omega_d) P(s_0 | \Omega_d) = [1\ 1] \prod_{n=1}^{N} R_n \vec{D}_0, \tag{4.31}$$

with R_n and $\vec{D}_0$ as in Equation 4.16 and Equation 4.17 respectively.

The states at the boundaries are not enough information to determine the true states of the emitter in a sampling interval from the observed counts. Since in the model considered here, there are d sub-intervals and each sub-interval with their own state. Again, including these states, then marginalization over the states of the sub-intervals in order to find the probability of count.

A recursive relation for the probability of counts in a sampling interval

In this section, s_n does not denote the state at the end of the nth measurement interval, but the state of the emitter at time n. Also, c_n denote the photons (signal + noise) emitted between times s_{n-1} and s_n. In the following derivation, the sub-state ζ_d is introduced, in order to identify a useful recurrence relationship for the DTMC multi-step photon-count distribution $P(c_n\, s_n | s_{n-1}\, \Omega_d)$. It is already noted that this distribution is independent of the photon counts in adjacent measurement windows, which is formally given by the independence condition

4.24. Consider a single sampling interval with boundary states s_0 and s_1. For a two sub-interval case the total sampled counts will be given by $c_1 = \kappa_1 + \kappa_2$, where κ_1 and κ_2 are the counts from the first and second sub-intervals respectively. There are only a discrete number of possibilities: either there are c_1 counts in the first sub-interval and zero in the second, or there are $c_1 - 1$ counts in the first sub-interval and 1 count in the second, and so forth. This gives rise to the expression

$$\begin{aligned}
&P(c_1\, s_1|s_0\, \Omega_d) \\
&\quad = P\left(\kappa_1 = c_1,\, \zeta_1|\zeta_0\, \Omega_d\right) P\left(\kappa_2 = 0,\, \zeta_2|\zeta_1\, \Omega_d\right) \\
&\quad + P\left(\kappa_1 = c_1 - 1,\, \zeta_1|\zeta_0\, \Omega_d\right) P\left(\kappa_2 = 1,\, \zeta_2|\zeta_1\, \Omega_d\right) \\
&\quad \cdots \\
&\quad + P\left(\kappa_1 = 0,\, \zeta_1|\zeta_0\, \Omega_d\right) P\left(\kappa_2 = c_1,\, \zeta_2|\zeta_1\, \Omega_d\right) \\
&\quad = \sum_{c_d=0}^{c_1} P\left(\kappa_1 = c_d,\, \zeta_1|\zeta_0\, \Omega_d\right) P\left(\kappa_2 = c_1 - c_d,\, \zeta_2|\zeta_1\, \Omega_d\right),
\end{aligned}$$

which is the *discrete convolution* of count probability distributions for each sub-interval. Since the only measurable quantity is the number of counts c_1 over this interval and we do not have knowledge of the state ζ, it is important to marginalise over the two possibilities of this intermediate state,

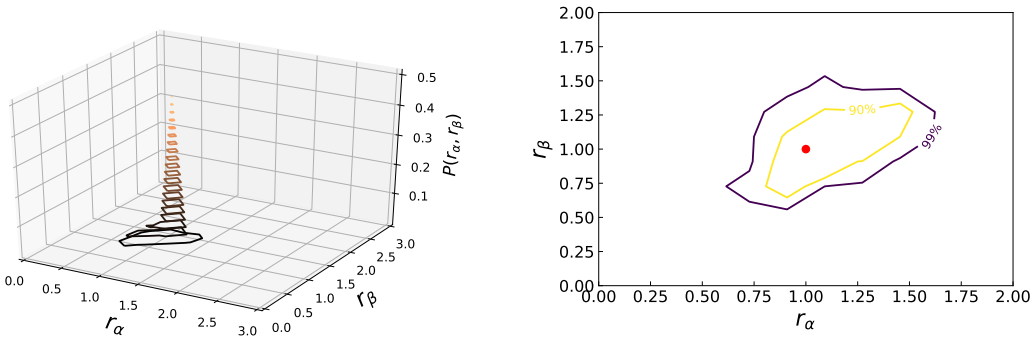
$$\begin{aligned}
&P(c_1\, s_1|s_0\, \Omega_d) = \\
&\quad \sum_{c_d=0}^{c_1} P\left(\kappa_1 = c_d,\, \zeta_1 = 0\,|\,\zeta_0\, \Omega_d\right) P\left(\kappa_2 = c_1 - c_d,\, \zeta_2\,|\,\zeta_1 = 0\, \Omega_d\right) \\
&+ \sum_{c_d=0}^{c_1} P\left(\kappa_1 = c_d,\, \zeta_1 = 1\,|\,\zeta_0\, \Omega_d\right) P\left(\kappa_2 = c_1 - c_d,\, \zeta_2\,|\,\zeta_1 = 1\, \Omega_d\right),
\end{aligned} \tag{4.32}$$

where, $\zeta_0 = s_0$ and $\zeta_2 = s_1$ and

$$P(\kappa_d\, \zeta_d\,|\,\zeta_{d-1}\, \Omega_d) = P(\kappa_d\,|\,\zeta_d\, \zeta_{d-1}\, \Omega_d)\, P(\zeta_d\,|\,\zeta_{d-1}\, \Omega_d).$$

When switching probabilities are high, however, the emitter state will be (nearly always) constant over sufficiently small time window. Then an integer d is chosen such that the distribution $P(\kappa_d\, \zeta_d\,|\,\zeta_{d-1}\, \Omega_d)$ is well-approximated by

$$P(\kappa_d\, \zeta_d\,|\,\zeta_{d-1}\, \Omega_d) = \begin{cases}
\mathcal{P} \times (1 - \alpha_d) & \zeta_{d-1} = 0, \zeta_d = 0 \\
\mathcal{P} \times (1 - \beta_d) & \zeta_{d-1} = 1, \zeta_d = 1 \\
\mathcal{P} \times \alpha_d & \zeta_{d-1} = 0, \zeta_d = 1 \\
\mathcal{P} \times \beta_d & \zeta_{d-1} = 1, \zeta_d = 0
\end{cases}$$



(a) (b)

Figure 4.11: (a) Contour plot showing the posterior probability distribution for r_α and r_β inferred from the simulated data given in Figure Figure 4.10. (b) Credible regions of the switching rates inferred from the simulated data. The contours lines indicate the 50%, 90%, and 99% credible regions. The true value is shown as red dot.

here, $\mathcal{P}$ denotes the corresponding Poissonian distribution of counts.

Here, for the calculation of $P(c_n\, s_n \mid s_{n-1}\, \Omega_d)$, the requirement is to keep track of the boundary states s_{n-1} and s_n. Hence there will be four possible combinations of initial and final states for a sampling interval: $(0 \to 0)$, $(1 \to 0)$, $(0 \to 1)$, $(1 \to 1)$. The calculation of $P(c_n\, s_n \mid s_{n-1}\, \Omega_d)$ starts from the first two sub-intervals. Then each step of the calculation accumulates the next sub-interval, one after another, up-to total of d sub-intervals. Hence the procedure is to initially assume two sub-intervals, and find the new distribution of possible counts c_1 resulting from the first two sub-intervals. The distribution is calculated by finding the sum of all possible combinations of counts that could have originated from the first two subintervals as given in Equation 4.32. For simplicity call the new distribution from first two sub-interval as $P_{2\zeta}$,

$$P_{2\zeta} = \sum_{i=0}^{c_1} P((\kappa_1 = i)\, \zeta_1 \mid \zeta_0\, \Omega_d)\, P((\kappa_2 = c_1 - i)\, \zeta_1 \mid \zeta_0\, \Omega_d). \tag{4.33}$$

Next, take the third sub-interval into account and consider the first two sub-intervals as a single block with count distribution $P_{2\zeta}$. Then calculate the possible ways that the counts c_1 could have been detected from the first three subintervals,

$$P_{3\zeta} = \sum_{i=0}^{c_1} P_{2\zeta}(\kappa_{2\zeta} = i)\, P_{3\zeta}((\kappa_3 = c_1 - i)\, \zeta_3 \mid \zeta_2\, \Omega_d). \tag{4.34}$$

Now consider these three as a single distribution. The process continues up to sub-interval d to find $P_{d\zeta}$, which can then be substituted back into Equation 4.31 to get the

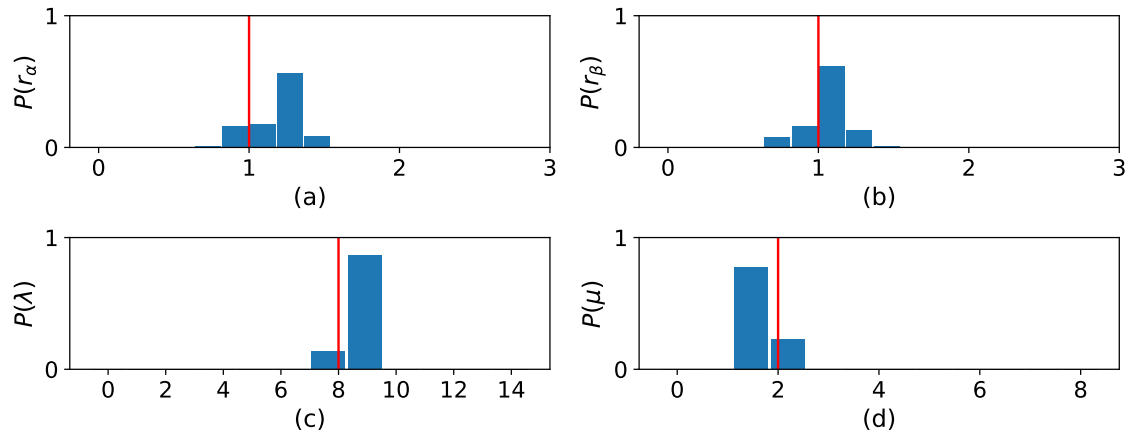


Figure 4.12: Marginal distribution: (a) and (b) shows the posterior probability distributions of r_α and r_β respectively. (c) and (d) gives the marginal distributions of the fluorescence and background rates λ and μ respectively. The red lines represents the true value used for simulation.

posterior probability distributions of r_α and r_β. Hence, the equation for inference can be written as,

$$P(r_\alpha\, r_\beta \,|\, \vec{c}\, I_d) = \frac{1}{\mathcal{M}} \int d\lambda \int d\mu\, [1\ 1] \prod_{n=1}^{N} R_n \vec{D}_0, \tag{4.35}$$

where the normalising constant is given by $\mathcal{M} = P(r_\alpha\, r_\beta\, \lambda\, \mu \,|\, I_d)/P(\vec{c} \,|\, I_d)$.

Figure 4.11 (a) shows the contour plot of the posterior distribution inferred for the simulated data given in Figure 4.10. The data is simulated with a high rate of $r_\alpha = 1$ and $r_\beta = 1$. The Figure 4.11 (b) shows the credible regions of r_α and r_β for the data given in Figure 4.10. The image can be made smoother by accumulating more number points for inference while running the algorithm. The inference was done for a total of $d = 4$ sub-intervals. The red dot indicates the true values of the parameters. The inference is converged very close to the true value. The Figure 4.12 displays the posterior probability distribution of r_α and r_β along with the marginal distributions of λ and μ.

5

Continuous-time Markov Chain Model for a Blinking Emitter

In the previous chapter, discrete-time Markov chains (DTMC) are used to construct a model for a blinking emitter. This chapter deals with modelling of the blinking emitters using a continuous-time Markov chain (CTMC). Compared to the discussion in the previous chapter, the assumption is now that the emitter can change its state at *any* point in time within a sampling interval. The first Section 5.1 describes the model in detail. In the last section 5.2 provides the numerical results for inference of switching rates from simulated blinking data.

5.1 Continuous-time Markov Chain model description

In the previous chapter, it was argued that the single-interval model is a good approximation to the continuous time model in the limit that the switching rates are very small. In that limit the occasional switch during a detection interval introduces negligible error and most

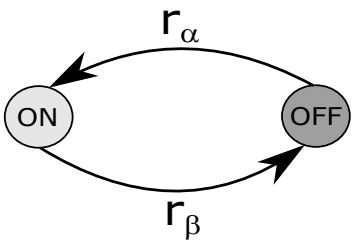


Figure 5.1: Graphical representation of a continuous time Markov process modelling the time evolution of the state. There are infinite time-steps within a given sampling interval. In this case, r_α and r_β are the switching *rates*.

intervals are either entirely *on* or entirely *off*. The emitter can switch at any time, so that the state is a continuous function of time, i.e. $s(t)$. This situation is modelled by a continuous time Markov chain (CTMC) as depicted in the Figure 5.1. At any given time, the state can still only have one of two values but can switch arbitrarily often in any given time interval. The detection events are as before, with the detectors reporting the accumulated counts from a finite time window of constant duration T.

The time spent in any state is exponentially distributed as discussed in chapter 2, and since there are only two states, the next state is deterministic. Accordingly, a natural relationship between the probabilities α_1 and β_1 in the previous model, or more generally a model that allows d switches in a sampling interval with respective probabilities α_d and β_d, and the switching rates r_α and r_β is given by Equation 2.33 and Equation 2.34 from Section 2.2.2, which are,

$$\alpha_d = 1 - e^{-r_\alpha T/d}, \tag{5.1}$$

$$\beta_d = 1 - e^{-r_\beta T/d}. \tag{5.2}$$

They are based on the probability of no switching occurring during an interval of duration T/d. Note that in the discrete model not remaining in the *off* state was equivalent to switching *on*. In this continuous model however, switching can occur at any time, and so it is possible for the emitter to switch *on* and then rapidly switch *off* again within any given time window. For large d, however, the probability of two switching events in a period of T/d becomes negligible. Thus in the fully continuous model, the probabilities for the states of the emitter at the boundaries of the measurement windows of length T can be calculated by considering d consecutive time windows of length T/d and with switching probabilities α_d and β_d, and then taking the limit as d approaches infinity.

The CTMC can be solved by the forward Kolmogorov equations discussed in Section

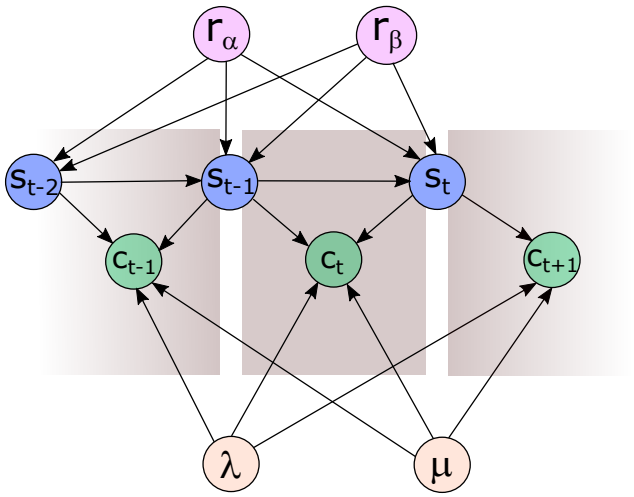


Figure 5.2: A BN representing the joint probability distribution of problem parameters. The pattern of nodes and connections inside the central grey region represents sampling interval. Here r_α and r_β indicate the *rates* of switching. This BN is used to find the variable independencies.

2.2.2. Representing the *off* state by the vector $[1\ 0]^T$ and the *on* state by $[0\ 1]^T$, the transition dynamics are encoded by the equation

$$\frac{dP(t)}{dt} = QP(t), \tag{5.3}$$

where Q is the generator matrix from Equation 2.18. The formal solution of Equation 5.3 is given by $P(t) = \exp(Qt)$. From this solution, we can then determine the transition probabilities $P_{ab}(t)$ from state a to state b in the interval Δt as given in Equation 2.25–2.28.

Bayesian network for the CTMC model

A BN showing the relation between the variables in a CTMC model is shown in Figure 5.2. The s_{t-1} and s_t are the states at instantaneous time at the boundaries of a sampling interval.*i.e.* $s_j = s(t_j)$, where the n^{th} detection interval goes from t_{n-1} to t_n and $t_n = nT$. The independencies that arise between the variables in the BN, are

$$c_t \perp c_m | s_{t-1}\, s_t\, \Omega_c \quad m \neq t, \tag{5.4}$$

$$s_t \perp s_m | s_{t-1}\, \Omega_c \quad m < t-1, \tag{5.5}$$

where, $\Omega_c = r_\alpha\, r_\beta\, \lambda\, \mu\, I_c$. Here I_c represents the background information for the CTMC model.

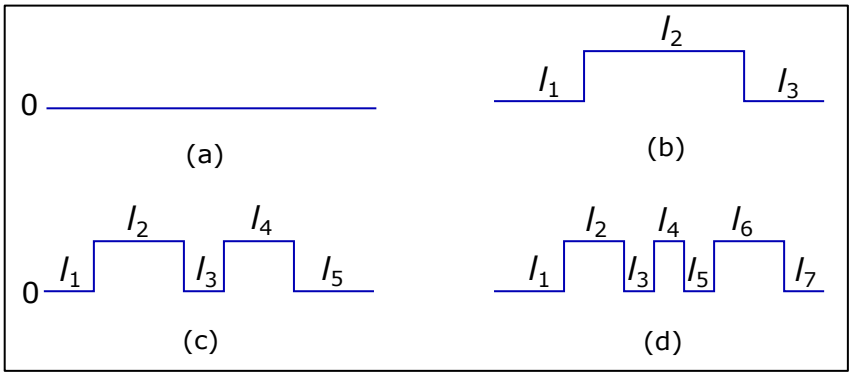


Figure 5.3: Switching events for state starts and ends in the *off* state (a) zero switch events (b) two switch events, (c) four switch events, and (d) six switch events.

5.2 Bayesian Inference on CTMC model

The basic inference for r_α and r_β using Bayes' rule is as before,

$$P(r_\alpha\, r_\beta \mid \vec{c}\, I_c) = \frac{P(r_\alpha\, r_\beta \mid I_c)\, P(\vec{c} \mid r_\alpha\, r_\beta\, I_c)}{P(\vec{c} \mid I_c)}. \tag{5.6}$$

$\vec{c}$ denotes a list of n detector counts and the parameters we want to infer are r_α and r_β, while at the same time we want to marginalise over the rates μ and λ, leading to an equivalent expression to Equation 4.8. To get traction on the problem, we add states at the *boundaries* of the detection interval s_j,

$$P(\vec{c}\,|\Omega_c) = \sum_{\vec{s}} P(\vec{c} \mid \vec{s}\,\Omega_c) P(\vec{s} \mid \Omega_c). \tag{5.7}$$

Note that in this situation the states s_j at the boundaries are not enough to specify the state during the detection interval, unlike in the DTMC single-step case where the state was constant during the whole interval. Now with the independencies given in 5.4 and 5.5, the above equation can be simplified to give

$$P(\vec{c}\,|\Omega_c) = \sum_{\vec{s}} \prod_{t=1}^{N} P(c_t\, s_t | s_{t-1}\, \Omega_c) P(s_0|\Omega_c). \tag{5.8}$$

This expression can again be converted into an efficient matrix equation as before Equation 4.18):

$$\sum_{\vec{s}} \prod_{t=1}^{N} P(c_t\, s_t | s_{t-1}\, \Omega_c) P(s_0|\Omega_c) = [1\ 1] \prod_{t=1}^{N} R_t \vec{D}_0, \tag{5.9}$$

with R_t and $\vec{D}_0$ as in Equation 4.16 and Equation 4.17, respectively.

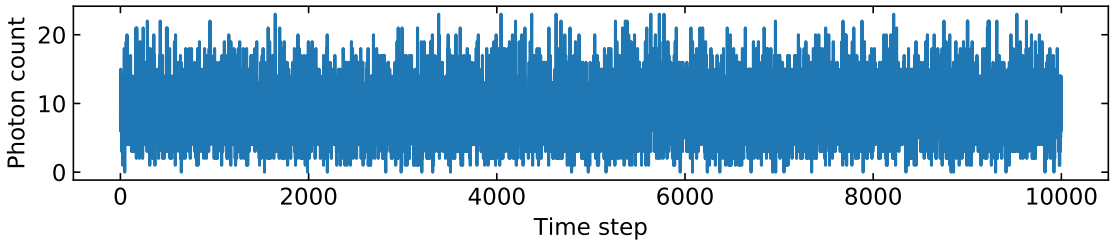


Figure 5.4: simulated data with high rate of switching, $r_\alpha = 3$ and $r_\beta = 2$.

All that remains to be done is then to calculate $P(c_t\, s_t|s_{t-1}\, \Omega_c)$. Knowledge about the fraction f of the interval for which the emitter was in the *on* state would make the evaluation of these count probabilities easier, as it would result in a weighted Poissonian. That is,

$$P(c_t\, s_t|s_{t-1}\, \Omega_c) = \int_0^1 df\, P(c_t|s_t\, s_{t-1}\, f\, \Omega_c) P(s_t\, f|s_{t-1}\, \Omega_c), \tag{5.10}$$

and

$$P(c_t|s_t\, s_{t-1}\, f\, \Omega_c) = \mathcal{P}(c_t; \mu + f\lambda) = \frac{(\mu + f\lambda)^{c_t} e^{-(\mu+f\lambda)}}{c_t!}. \tag{5.11}$$

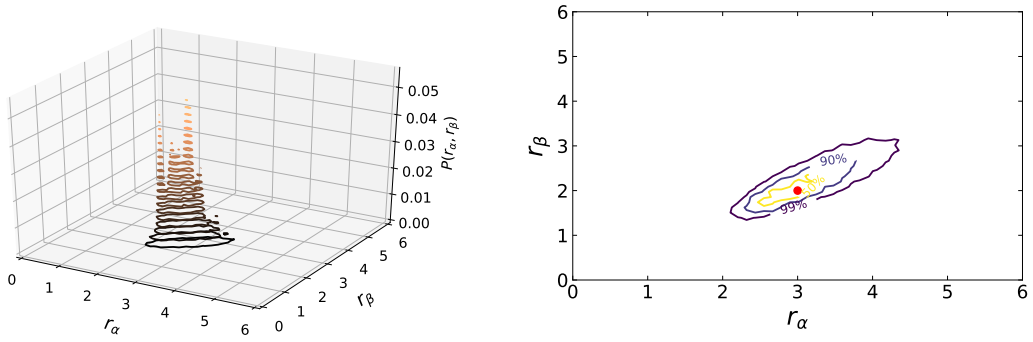
here, for simplicity, the length of the sampling interval is taken to be 1 (*i.e.* T =1)

Now, the task is to find an expression for $P(s_t\, f|s_{t-1}\, \Omega_c)$. For brevity, introduce the notation $R_{ab}(f) \equiv P(s_t = b\, f|s_{t-1} = a\, \Omega_c)$. This probability accounts for all the possible histories where the detection interval spent a fraction f in state *on* and ended in the state s_t given it started in state s_{t-1}. The different histories can be decomposed into differing numbers of switch events as shown in Figure 5.3. Consider the case where the state starts and ends in the *off* state, i.e. $R_{00}(f)$ (Figure 5.3 (a)). Clearly an odd number of switch events will not be consistent with the boundary states and hence will have probability zero. With zero switch events the entire interval is *off* and the probability is just the exponential distribution $\delta(f)e^{-r_\alpha}$, where δ is the Dirac delta function. Two switch events partitions the interval into three regions l_1, l_2, and l_3 with states *off–on–off* and probability

$$\int_0^{1-f} dl_1 e^{-r_\alpha l_1} r_\alpha e^{-r_\beta l_2} r_\beta e^{-r_\alpha l_3} = r_\alpha r_\beta e^{-r_\alpha(1-f)-r_\beta f}\,(1-f), \tag{5.12}$$

where we have used $l_2 = f$ and $l_1 + l_3 = 1 - f$.

In general the interval will be partitioned into a set of *on* states of total duration f and a set of *off* states of duration $1 - f$ so that the exponentials combine and will not depend on the durations l_j. For example with six switches there are seven regions with three switch *on*



(a) (b)

Figure 5.5: (a) Contour plot showing the posterior probability distribution for r_α and r_β inferred from the simulated data given in Figure 4.10. (b) Credible regions of the switching rates inferred from simulated data. The contours contain the 50%, 90%, and 99% credible regions. The true value is shown as red dot.

and three switch *off* events as shown in Figure 5.3 (d). The probability is,

$$r_\alpha^3 r_\beta^3 e^{-r_\alpha(1-f)-r_\beta f} \times \underbrace{\int_0^f dl_2 \int_0^{f-r_2} dl_4}_{A} \times \underbrace{\int_0^{1-f} dl_1 \int_0^{1-f-l_1} dl_3 \int_0^{1-f-l_1-l_3} dl_5}_{B} 1. \tag{5.13}$$

The integrals in A calculate the volume of a two-dimensional simplex of side f, and those in B the volume of a three dimensional simplex of side $1-f$. Since the volume of an n-dimensional simplex of side f is $f^n/n!$, Equation 5.13 evaluates to

$$\frac{f^2(1-f)^3 r_\alpha^3 r_\beta^3 e^{-r_\alpha(1-f)-r_\beta f}}{2!3!}. \tag{5.14}$$

Summing over all switch events gives

$$R_{00}(f) = \delta(f)e^{-r_\alpha} + e^{-r_\alpha(1-f)-r_\beta f} \underbrace{\sum_{k=0}^{\infty} \frac{f^{k-1}(1-f)^k}{(k-1)!k!} r_\alpha^k r_\beta^k}_{C}. \tag{5.15}$$

The summation term C can written as a Bessel function $J_1(x)$ of the first order with an additional multiplication term. The above equation becomes,

$$R_{00}(f) = \delta(f)e^{-r_\alpha} - e^{-r_\alpha(1-f)-r_\beta f} \sqrt{\frac{(f-1)r_\alpha r_\beta}{f}} \times J_1\left(2\sqrt{(f-1)f r_\alpha r_\beta}\right). \tag{5.16}$$

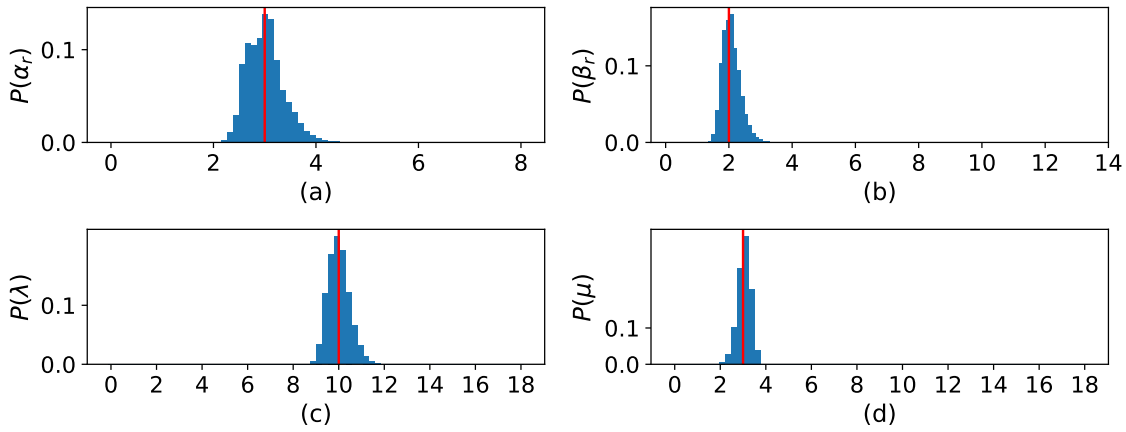


Figure 5.6: Marginal distribution: (a) and (b) shows the posterior probability distributions of r_α and r_β respectively. (c) and (d) gives the marginal distributions of the fluorescence and background rates λ and μ respectively. The red lines represents the true value used for simulation.

A similar argument yields the other possible fraction probabilities:

$$R_{01}(f) = r_\alpha J_0\left(2\sqrt{(f-1)fr_\alpha r_\beta}\right) e^{-r_\alpha(1-f)-r_\beta f}, \tag{5.17}$$

$$R_{10}(f) = r_\beta J_0\left(2\sqrt{(f-1)fr_\alpha r_\beta}\right) e^{-r_\alpha(1-f)-r_\beta f}, \tag{5.18}$$

$$R_{11}(f) = \delta(1-f)e^{-r_\beta} + e^{-r_\alpha(1-f)-r_\beta f}\sqrt{\frac{fr_\alpha r_\beta}{f-1}} \times J_1\left(2\sqrt{(f-1)fr_\alpha r_\beta}\right). \tag{5.19}$$

Finally we have

$$P(c_n\, s_n | s_{n-1}\, \Omega_c) = \int_0^1 df \mathcal{P}(c_n; \mu + f\lambda) R_{s_{n-1}s_n}(f), \tag{5.20}$$

which is an integral that needs to be done numerically.

The Figure 5.5 display the posterior probability distribution of the model parameters and credible regions calculated for the data given in Figure 5.4. Even at a high rate of $r_\alpha = 3$ and $r_\beta = 2$ the inference gave a fairly good estimation. In Figure 5.6 the marginal distributions of λ and μ along with the probability distributions of r_α and r_β are shown. The true values of λ and μ are 10 and 3 respectively.

5.2.1 Relation between single interval model and continuous-time model

The aim of this section is to see how the continuous-time model behaves in a low rate switching regime and compare it with the single interval model. For Bessel function of the

first kind,

$$J_m(z) \approx \frac{1}{\Gamma(m+1)}\left(\frac{z}{2}\right)^m \qquad z << 1 \tag{5.21}$$

where, $\Gamma(x)$ is the Gamma function. Hence at a low rate of switching, we can expand first order in r_α, r_β to obtain $R_{00}(f) \approx \delta(f)(1-r_\alpha)$, $R_{11}(f) \approx \delta(1-f)(1-r_\beta)$, $R_{01}(f) \approx r_\alpha$ and $R_{10}(f) \approx r_\beta$. Furthermore, from Equation 5.1 and Equation 5.2, $\alpha_1 \approx r_\alpha$ and $\beta_1 \approx r_\beta$. With this approximations Equation 5.20 yields,

$$P(c_n\,0|0\,\Omega_c) \approx (1-\alpha_1)\,\mathcal{P}(c_n;\mu) \tag{5.22}$$

$$P(c_n\,1|1\,\Omega_c) \approx (1-\beta_1)\,\mathcal{P}(c_n;\mu+\lambda) \tag{5.23}$$

$$P(c_n\,1|0\,\Omega_c) \approx \alpha_1 \tilde{c_n} \tag{5.24}$$

$$P(c_n\,0|1\,\Omega_c) \approx \beta_1 \tilde{c_n} \tag{5.25}$$

where,

$$\tilde{c_n} = \int_0^1 df\,\mathcal{P}(c_n;\mu+f\lambda), \tag{5.26}$$

is the probability of obtaining c_n counts averaged over all durations spent on in the interval.

The first two probabilities (Equation 5.22 and Equation 5.23) are identical to those obtained in the single interval model. The last two (Equation 5.24 and Equation 5.25) differ as in the single interval model we made the arbitrary decision to take the switch event as happening at the ends of the intervals. However, sampling intervals that contain a switch event become rare when the rates are small. So the continuous time model becomes indistinguishable from the single step model in this limit.

The single interval model loses accuracy as the rate of switching increases. Figure 5.7 shows this by examining the difference between the true switching rate (for data simulated from a continuous-switching model) and the result of a single-interval model inference. The colour map represents the distance from the true value to the maximum value from the posterior probability distribution. The black region corresponds to a small distance, and is therefore the most applicable region for the single-interval inference. As the (continuous) rates increase the single-interval model starts failing. An approximate threshold for applicability of the single-step model is when $r_\alpha r_\beta < 1$. A grid of 150×150 simulations were run to produce this colourmap, and the posterior distributions for switching rates are shown for four illustrative cases. Note that the scales are different for each contour plot. The lower left plot shows a good convergence of the inference to the true value, where the rates are small. The

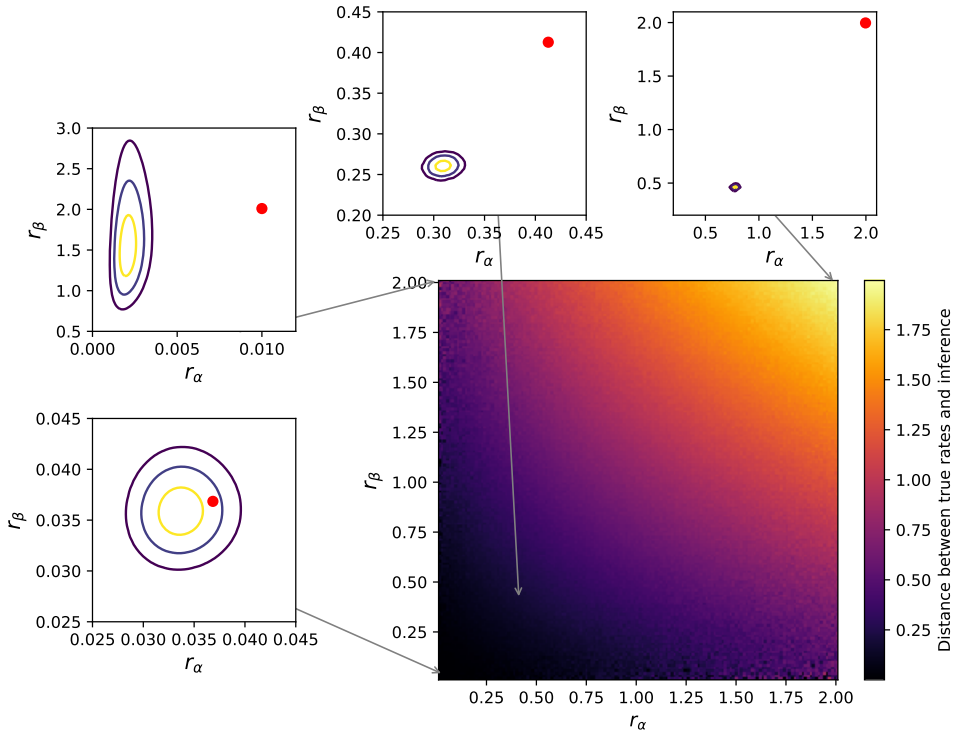


Figure 5.7: Accuracy of the single-step inference to simulated data of various switching rates. For small switching probabilities the inference gives good agreement with the known parameters, but for higher switching probabilities the inference is less accurate. An approximate threshold for applicability of the single-step model is when $r_\alpha r_\beta < 1$.

sub-figure in the top left considers a rate where the *off* rate is very high and *on* rate is low. The convergence is not good for this case, since the asymmetry of switching rates results in very few switching on events—the data remains in the *off* state for most of the time. The plot in the top centre shows the deviation of the inference from the true value as both rates increase. In the top right this is further illustrated for the extreme case where both rates are so high that there is an average of two switching events per measurement interval. The posterior distribution is tightly converged since there are a large number of switching events, but it is significantly diverged from the true value. The single-interval inference is clearly not applicable for high rates. In this region, the continuous time model performs well as shown in Figure 5.5, even though the traditional threshold technique cannot even be applied to the data in Figure 5.4.

6

Conclusion and Future Outlook

We have developed three models for extracting blinking rates from the time trace of blinking emitters. The models are based on hidden discrete and continuous-time Markov processes. The simplest and fastest model assumes the state only switches at the boundaries of the sampling intervals, and is a good choice to model a blinking time trace with very slow switching rates. The algorithm was demonstrated with simulated data where all the parameters are known (in order to establish accuracy) and also on previously published experimental data. If the true underlying mechanism can switch at any time, this single-interval model fails for reasonably high switching rates but is applicable when the switching rates are fairly low. Its key advantages are simplicity and speed.

The next model presented is again based on a discrete-time Markov process. This model divides the sampling interval into subintervals of constant state that are shorter than the sampling interval, to allow switching within a sampling interval to a limited extent. The Bayesian inference on the model is capable of inferring slow as well as fast switching rates, as long as sufficiently many subintervals are considered. The model is coded in a recursive

way, and is a bit time consuming to run for a large number of sub-intervals.

The final model developed in the thesis assumes the state can switch at any time and is based on a continuous-time Markov process. This can be considered as a limiting case of the sub-interval model above as the number of subintervals goes to infinity. The Bayesian inference on the continuous model works well for both fast and slow switching rates. Even for a time trace simulated with a rate of switching too high to see, the continuous time model is successful at inferring blinking rates. The fluorescence and background rates can also be extracted more precisely using this model.

The choice of model for a given situation depends on a number of factors. The most important factor is the time-scale of the switching relative to the sampling interval. For slow rates the single-interval model is both accurate and efficient. For faster switching rates we expect many switching events per sampling interval. Then the continuous or sub-interval models are good choices, and the efficiency of the algorithm would be the determining factor. As a starting point to generate beyond Poissonian counts the sub-interval model is more flexible as it makes less assumptions.

For an arbitrary time trace the fluorescence and background photon detection rates are not known. For all three models developed here, the Bayesian inference correctly treats these parameters as unknowns and actually yields photon emission rate inferences as well as the switching rate inferences. It should be emphasised that in each case the inference is exact according to the chosen model. For analysis of experimental data, therefore, the only assumption made in the data processing is the application of a particular model to the physical system. There are several extensions or variations from this work. For instance instead of inferring the rates we can tag each data point with the probability that the state was *on* or *off* and this create an algorithm to filter the data to extract just the *on* intervals.

Hidden Markov models are quite general and are used in a wast variety of fields. Extending the analysis presented here to other models where the *hiding* is not an accumulated Poissonian distribution would be exciting and fruitful. Similarly extending the model beyond two states would increase its generality and reach. The content of this thesis is being prepared for publication [43].

References

[1] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov. *Invited review article: Single-photon sources and detectors*. Review of scientific instruments **82**(7), 071101 (2011).

[2] J. L. O'brien, A. Furusawa, and J. Vučković. *Photonic quantum technologies*. Nature Photonics **3**(12), 687 (2009).

[3] C. Kurtsiefer, S. Mayer, P. Zarda, and H. Weinfurter. *Stable solid-state source of single photons*. Physical review letters **85**(2), 290 (2000).

[4] I. Aharonovich, S. Castelletto, D. Simpson, C. Su, A. Greentree, and S. Prawer. *Diamond-based single-photon emitters*. Reports on progress in Physics **74**(7), 076501 (2011).

[5] I. Aharonovich, A. D. Greentree, and S. Prawer. *Diamond photonics*. Nature Photonics **5**(7), 397 (2011).

[6] N. Somaschi, V. Giesz, L. De Santis, J. Loredo, M. P. Almeida, G. Hornecker, S. L. Portalupi, T. Grange, C. Antón, J. Demory, *et al. Near-optimal single-photon sources in the solid state*. Nature Photonics **10**(5), 340 (2016).

[7] X. Ding, Y. He, Z.-C. Duan, N. Gregersen, M.-C. Chen, S. Unsleber, S. Maier, C. Schneider, M. Kamp, S. Höfling, *et al. On-demand single photons with high extraction efficiency and near-unity indistinguishability from a resonantly driven quantum dot in a micropillar*. Physical review letters **116**(2), 020401 (2016).

[8] I. Aharonovich, D. Englund, and M. Toth. *Solid-state single-photon emitters*. Nature Photonics **10**(10), 631 (2016).

[9] A. M. Berhane, C. Bradac, and I. Aharonovich. *Photoinduced blinking in a solid-state quantum system*. Physical Review B **96**(4), 041203 (2017).

[10] C. Galland, Y. Ghosh, A. Steinbrück, M. Sykora, J. A. Hollingsworth, V. I. Klimov, and H. Htoon. *Two types of luminescence blinking revealed by spectroelectrochemistry of single quantum dots*. Nature **479**(7372), 203 (2011).

[11] C. Galland, Y. Ghosh, A. Steinbrück, J. A. Hollingsworth, H. Htoon, and V. I. Klimov. *Lifetime blinking in nonblinking nanocrystal quantum dots*. Nature communications **3**, 908 (2012).

[12] C. Bradac, T. Gaebel, N. Naidoo, M. Sellars, J. Twamley, L. Brown, A. Barnard, T. Plakhotnik, A. Zvyagin, and J. Rabeau. *Observation and control of blinking nitrogen-vacancy centres in discrete nanodiamonds*. Nature nanotechnology **5**(5), 345 (2010).

[13] U. Jantzen, A. B. Kurz, D. S. Rudnicki, C. Schäfermeier, K. D. Jahnke, U. L. Andersen, V. A. Davydov, V. N. Agafonov, A. Kubanek, L. J. Rogers, *et al. Nanodiamonds carrying silicon-vacancy quantum emitters with almost lifetime-limited linewidths*. New Journal of Physics **18**(7), 073036 (2016).

[14] M. Kuno, D. P. Fromm, H. F. Hamann, A. Gallagher, and D. J. Nesbitt. *Nonexponential "blinking" kinetics of single cdse quantum dots: A universal power law behavior*. The journal of chemical physics **112**(7), 3117 (2000).

[15] M. Kuno, D. Fromm, H. Hamann, A. Gallagher, and D. J. Nesbitt. *"on"/"off" fluorescence intermittency of single semiconductor quantum dots*. The Journal of chemical physics **115**(2), 1028 (2001).

[16] A. A. Cordones. *Mechanisms for Fluorescence Blinking and Charge Carrier Trapping in Single Semiconductor Nanocrystals*. Ph.D. thesis, UC Berkeley (2012).

[17] C. H. Crouch, O. Sauter, X. Wu, R. Purcell, C. Querner, M. Drndic, and M. Pelton. *Facts and artifacts in the blinking statistics of semiconductor nanocrystals*. Nano letters **10**(5), 1692 (2010).

[18] P. A. Frantsuzov, S. Volkán-Kacsó, and B. Jankó. *Model of fluorescence intermittency of single colloidal semiconductor quantum dots using multiple recombination centers*. Physical review letters **103**(20), 207402 (2009).

[19] V. G. Kulkarni. *Modeling and analysis of stochastic systems* (CRC Press, 2010), 2nd ed. ed.

[20] D. R. Cox and H. D. Miller. *The theory of stochastic processes* (Chapman and Hall/CRC, 2001).

[21] P. Olofsson. *Probability, statistics, and stochastic processes* (Wiley-Interscience, 2005).

[22] M. Nirmal, B. O. Dabbousi, M. G. Bawendi, J. Macklin, J. Trautman, T. Harris, and L. E. Brus. *Fluorescence intermittency in single cadmium selenide nanocrystals*. Nature **383**(6603), 802 (1996).

[23] A. L. Efros and D. J. Nesbitt. *Origin and control of blinking in quantum dots*. Nature nanotechnology **11**(8), 661 (2016).

[24] P. Frantsuzov, M. Kuno, B. Janko, and R. A. Marcus. *Universal emission intermittency in quantum dots, nanorods and nanowires*. Nature Physics **4**(7), 519 (2008).

[25] P. A. Frantsuzov, S. VolkaÌĄn-KacsoÌĄ, and B. JankoÌĄ. *Universality of the fluorescence intermittency in nanoscale systems: experiment and theory*. Nano letters **13**(2), 402 (2013).

[26] E. Neu, M. Agio, and C. Becher. *Photophysics of single silicon vacancy centers in diamond: implications for single photon emission*. Optics express **20**(18), 19956 (2012).

[27] S. Castelletto and A. Edmonds. *680-890nm spectral range of nickel-nitrogen and nickel-silicon complex single centres in diamond*. In *Quantum Communications and Quantum Imaging X*, vol. 8518, p. 85180R (International Society for Optics and Photonics, 2012).

[28] C. Wang. *A solid-state single photon source based on color centers in diamond*. Ph.D. thesis, lmu (2007).

[29] F. Cichos, C. von Borczyskowski, and M. Orrit. *Power-law intermittency of single emitters*. Current Opinion in Colloid & Interface Science **12**(6), 272 (2007).

[30] V. Rombach-Riegraf, P. Oswald, R. Bienert, J. Petersen, M. Domingo, J. Pardo, P. Gräber, and E. Galvez. *Blinking effect and the use of quantum dots in single*

molecule spectroscopy. Biochemical and biophysical research communications **430**(1), 260 (2013).

[31] F. D. Stefani, J. P. Hoogenboom, and E. Barkai. *Beyond quantum jumps: blinking nanoscale light emitters*. Phys. Today **62**(2), 34 (2009).

[32] T. Burzykowski, J. Szubiakowski, and T. Rydén. *Analysis of photon count data from single-molecule fluorescence experiments*. Chemical Physics **288**(2-3), 291 (2003).

[33] M. Fox. *Quantum optics: an introduction*, vol. 15 (OUP Oxford, 2006).

[34] A. L. Efros and M. Rosen. *Random telegraph signal in the photoluminescence intensity of a single quantum dot*. Physical Review Letters **78**(6), 1110 (1997).

[35] E. T. Jaynes. *Probability theory: The logic of science* (Cambridge university press, 2003).

[36] P. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support* (Cambridge University Press, 2005).

[37] W. M. Bolstad. *Introduction to Bayesian statistics* (John Wiley, 2007), 2nd ed.

[38] A. Darwiche. *Modeling and reasoning with Bayesian networks* (Cambridge university press, 2009).

[39] J. Ding. *Probabilistic inferences in bayesian networks*. In *Bayesian Network* (InTech, 2010).

[40] K. B. Korb and A. E. Nicholson. *Bayesian artificial intelligence* (CRC press, 2010).

[41] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques* (MIT press, 2009).

[42] I. Ben-Gal. *Bayesian networks*. Encyclopedia of statistics in quality and reliability **1** (2008).

[43] J. Geordy, L. Rogers, C. Rogers, T. Volz, and A. Gilchrist. *Bayesian estimation of switching rates for blinking quantum emitters*. manuscript under preparation .