Imperial College London

Department of Mathematics

Exact solutions of master equations for the analysis of gene transcription models

Justine Dattani

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy
Imperial College London, November 2015

Abstract

This thesis is motivated by two associated obstacles we face for the solution and analysis of master equation models of gene transcription. First, the master equation – a differential-difference equation that describes the time evolution of the probability distribution of a discrete Markov process – is difficult to solve and few approaches for solution are known, particularly for non-stationary systems. Second, we lack a general framework for solving master equations that promotes explicit comprehension of how extrinsic processes and variation affect the system, and physical intuition for the solutions and their properties.

We address the second obstacle by deriving the exact solution of the master equation under general time-dependent assumptions for transcription and degradation rates. With this analytical solution we obtain the general properties of a broad class of gene transcription models, within which solutions and properties of specific models may be placed and understood. Furthermore, there naturally emerges a decoupling of the discrete component of the solution, common to all transcription models of this kind, and the continuous, model-specific component that describes uncertainty of the parameters and extrinsic variation. Thus we also address the first obstacle, since to solve a model within this framework one needs only the probability density for the extrinsic component, which may be non-stationary. We detail its physical interpretations, and methods to calculate its probability density.

Specific models are then addressed. In particular we solve for classes of multistate models, where the gene cycles stochastically between discrete states. We use the insights gained from these approaches to deduce properties of several other models. Finally, we introduce a quantitative characterisation of timescales for multistate models, to delineate "fast" and "slow" switching regimes. We have thus demonstrated the power of the obtained general solution for analytically predicting gene transcription in non-stationary conditions.

Declaration of originality

I hereby declare that the work presented in this thesis is my own, unless otherwise stated and referenced.

Justine Dattani

To my parents,

on whose shoulders I am standing.

Acknowledgments

I am incredibly priviledged and grateful to have had Prof. Mauricio Barahona as my PhD supervisor; it is difficult to describe the all-permeating effects of his support, guidance, presence, and friendship on my academic development and general well-being, without degenerating into paragraphs of flowery superlatives. He is a true super-human, and I admire him more than his modesty would allow him to acknowledge.

I thank the Engineering and Physical Sciences Research Council (EPSRC) and the Department of Mathematics at Imperial College London for funding my PhD, and giving me the opportunity to attend numerous conferences and workshops. I would also like to thank Philipp, Ed, Juan, and Andrea who proofread chapters and gave me invaluable feedback, as well as all the people who worked behind the scenes to provide help and support, in particular Dr John Gibbons, Dr Tony Bellotti, Rusudan Svanidze, Anderson Santos, and Kalra Taylor.

The past 4 years would have been comparatively colourless without the friendship and support of my office mates over the years, who ensured that no workday was ever dull: Michael, Mariano, Antoine, Elias, Juan, and Ben, I'm going to miss you.

I would especially like to thank Ed for his friendship, love and support, and for keeping me happy, fed, and relatively rested during crunch time. Finally, I would not be here today without my parents Anne-Marie and Satish, who have always done everything they could so that I could do anything I wanted to. This thesis is their achievement too.

Justine Dattani
London, February 2016

Contents

List of Figures

List of Tables

Part I

Introduction to gene transcription models and master equations

Chapter 1

Introduction

As early as the 1920s geneticists predicted that genes could "play a fundamental role in determining the nature of all cell substances, cell structures and cell activities" (Muller 1922; Pontecorvo 1968), via a "giant hereditary molecule" (Koltsov (1927), as cited in Soyfer 2001). It is now known that genetic information is transferred to the biological system via the reproduction (*transcription*) of a section of DNA (a *gene*) to produce messenger RNA (mRNA) molecules; the mRNA molecules are then used as templates for the synthesis of a protein, or small number of proteins (*translation*). This process of information transfer is known as *gene expression*, and occurs in all known life forms to control the structure, functions, decision processes, adaptability, and phenotypes of the cell. Understanding gene expression and its regulation is thus an issue of fundamental interest.

Despite gene expression being a control mechanism, its coordination is mired in mystery. The process requires a complex sequence of biological steps and potentially involves a very low number of biomolecules, leading to a distinct kind of heterogeneity at the single cell level, beyond the standard randomness and variability widely present in biochemical systems. Heterogeneity is observed ubiquitously, even in clonal cell populations in a homogeneous environment (Elowitz and Leibler 2000; Elowitz *et al.* 2002; Ozbudak *et al.* 2002; Blake *et al.* 2003; Raj *et al.* 2006; Zenklusen *et al.* 2008; Zenobi 2013; Bahar Halpern *et al.* 2015), and across time in single cells (Rosenfeld *et al.* 2005; Golding *et al.* 2005; Cai *et al.* 2006; Stevense *et al.* 2010; Yunger *et al.* 2010; Suter *et al.* 2011; Harper *et al.* 2011; Larson *et al.* 2013; Corrigan and Chubb 2014; Francesconi and Lehner 2014; Singer *et al.* 2014). So much so, that gene expression is now routinely labelled as being a "fundamentally stochastic process" (Elowitz *et al.* 2002; Raser and E. K. O'Shea 2005; Kærn *et al.* 2005; Raj and Oudenaarden 2008; Raj and Oudenaarden 2009; Shahrezaei and Swain 2008; Li and Xie 2011; Junker and Oudenaarden 2014). Since expression levels have such important consequences for the proper functioning of the cell and the cell population as a whole, the challenge of understanding gene expression brings us towards the realms of philosophy: how do cells operate with such randomness at their local level, yet together they form a sentient being that can philosophise about how it can write about philosophising? In that sense, the study of gene expression forms part of the "quantum field theory" of biology[1].

[1] In fact, Erwin Schrödinger himself addressed these questions in his book *What is life?* (Schrödinger 1948).

One can argue that, almost by definition, interesting philosophical questions have as their essence innumerable facets, and are intractably nested in layers of other questions. It is 'turtles all the way down', even in a grain of salt[2]. The questions of life, even from a purely biological viewpoint, are a case in point; a real understanding of gene expression and its regulation is hindered by the complexity of each cell, each network of intracellular interactions and molecular species, and how each cell positions itself and its contributions within a wider population of cells and their own fluctuating environment, to name but a few of the turtles (Burger 1999; Tanaka *et al.* 2003; Rockman and Kruglyak 2006; Junker and Oudenaarden 2014). We are only able to observe relatively few of the factors that determine gene expression and regulation at any one time. The data we obtain can be thought of as the output from a black box, that contains all the information and processes that we have not observed (Lestas *et al.* 2010).

The rapid technological advances of this century are transforming molecular cell biology into a data-rich field. Mathematical models and analysis are now essential tools for increasing the yield of information from experiments by extracting information from the data and inferring properties hidden within the black box. However, the utility of the current mathematical models commonly used in the field is reaching two bottlenecks. First, **we need to obtain exact, analytical solutions of models that match the kind of data being produced**. Since we are now able to capture single-molecule mRNA counts that change stochastically over time, mathematical formulations and understanding must be in terms of discrete probability distributions. We need to solve models that give us full, time-dependent distributions of integer random variables, not just averages, continuous approximations or stationary snapshots. Second, in order to keep up with the extraordinary rate at which experiments are performed, **we are in particular need of mathematical models and their properties that are general enough to be applicable to a wide range of investigations and datasets, while also being adaptable to each experimental hypothesis**.

With these requirements in mind, Part II of this thesis answers the first bottleneck by presenting the exact solution and properties of a broad class of gene transcription models. The mathematical framework this provides is general enough to include several of the most well-used models as specific cases, and naturally accounts for non-stationary systems. In so doing, we respond to the second bottleneck. In Part III some of the general results are applied to the commonly-used class of multistate models. The use of a general framework, rather than solving individual models one-by-one, gives us a contextual vantage point from which we can gain physical intuition and predict several fruitful further directions. Exact expressions are also derived for a quantitative characterisation of timescales, and implications in the context of a broad class of relevant gene transcription models are discussed. Finally, further directions to explore are discussed in Part IV.

The mathematical framework that will be considered here is not restricted in its scope to modelling gene transcription; the results are directly applicable to several gene expres-

[2]See Carl Sagan's short essay *Can we know the universe? Reflections on a grain of salt* (Sagan 1979).

sion models for protein levels that use timescale arguments to omit transcription (Paulsson 2004; Hornos *et al.* 2005; Iyer-Biswas *et al.* 2009; Ge *et al.* 2015). Further afield, our framework it is analogous to $G/G/\infty$ systems in queueing theory (Brémaud 2001; Kendall 1953). We will also draw upon several concepts and classical results from other fields, including stochastic processes, engineering, and the analysis and geometry of differential equations. However, the motivation to solve models of gene transcription has driven the direction of this work, so the focus of the physical interpretations and applications will remain firmly on mRNA molecules as the objects of interest. A comprehensive account of the fields of research that form the context of this work will not be helpful here, since they are both vast and advancing at pace. The following brief review sets out the biological background and quantitative paradigms that constitute the motivational basis of this work, focusing on the aspects of stochastic gene expression and master equation models of gene transcription that will be most relevant in this thesis. For more general overviews and further reading I recommend (Kærn *et al.* 2005; Raj and Oudenaarden 2008; Li and Xie 2011; Satija and Shalek 2014; Levine *et al.* 2014) for the causes and consequences of stochastic gene expression and heterogeneity, transcriptional machinery, and experimental strategies, and (Wilkinson 2009; Coulon *et al.* 2013; Sanchez-Osorio *et al.* 2014) for some quantitative approaches for modelling stochastic gene expression.

1.1 Brief review of approaches for the stochastic modelling of gene transcription

While this thesis has been motivated so far in the context of heterogeneous gene expression, and the increasing volume and quality of experimental data, variation in gene expression is not itself a new discovery. Indeed, by 1932 Haldane was already predicting differences in the time of action of genes (Haldane 1932), although it took several decades for clear experimental evidence to emerge (Schwartz 1962; Spudich and Koshland Jr 1976). The measurements of gene products themselves were often limited to protein concentrations, which could be described in detail by traditional rate equations – ODE formulations of individual reactions and species (Smolen *et al.* 2000).

Observed fluctuations were, however, recognised as phenomena worthy of investigation. Differing attitudes for which level to view the fluctuations from led to the emergence of two main approaches for modelling the stochasticity. The Langevin approach maintained the viewpoint of chemical concentrations as the measurements of fundamental interest. Drawing upon the central limit theorem, the fluctuations were assumed to be Gaussian and theoretical results were based on applications of the fluctuation-dissipation theorem (Nitzan *et al.* 1974; Keizer 1975; Keizer 1976; Keizer 1977; Grossmann 1976). The second approach left the perspective of concentrations behind, in favour of the more elementary viewpoint based on the collision theory of chemical reactions. Believing that the state of a chemical reaction system should describe the number of molecules, not the concentration, stochasticity emerged naturally from the idea that chemical reactions take place when diffusing molecules collide with sufficient energy. The resulting differential-

difference equations for the probability distributions of the number of molecules of each chemical species of interest became known as the *chemical master equations*, or simply, *master equations* (Kramers 1940; Montroll and Shuler 1957; Bartholomay 1958; McQuarrie 1967; Gardiner and Chaturvedi 1977; Kampen 1992).

However, the use of stochastic modelling in molecular biology was slow to take hold. This was partly due to a belief that for simple reactions, the probability distribution of the number of molecules in the system exhibits only small fluctuations around the mean (McQuarrie 1967), but also because the master equations could only be solved for a handful of simple reactions (Kampen 1992; Gillespie 1977; Gardiner *et al.* 1976; Berg 1978). Various methods were developed, including van Kampen's system size expansion (Kampen 1976; Kampen 1992), generating function methods (Bartholomay 1958; McQuarrie 1967), and Gardiner's Poisson expansion (Gardiner and Chaturvedi 1977), but the methods could not always be applied systematically and were not easy to use. In practice, Gillespie's exact stochastic simulation algorithm (Gillespie 1976; Gillespie 1977) was (and still is) used instead, which does not give expressions for the distributions that can be analysed mathematically.

For studies in gene expression, the lack of reliable single-cell data was an additional major hindrance to motivating the use of master equation models. However, a shift in that direction was sparked by the pioneering work of Ko and his co-workers in the early 1990s. They obtained fluorescence intensity data for expression of the β-galactosidase gene in single cells under different induction doses (Ko *et al.* 1990). The expression data displayed clear bimodality, meaning that the distributions had clear peaks at two distinct levels of expression. Since expression levels were not distributed around the mean value, traditional ODE models would not be able to capture the characteristics of the data. The data also confirmed a conclusion reached several decades before by Novick and Wiener (Novick and Weiner 1957) that induction increased the proportion of cells with high expression levels, rather than increasing the expression in each cell linearly. However, Ko went a step further in 1991 by proposing a corresponding stochastic model where in each cell the gene switches between an active (ON) state, and an inactive (OFF) state according to a time-homogeneous Markov chain (Ko 1991). The implication was that gene expression levels would change in single cells over time, since trancription would only occur during periods when the gene is active. This picture, usually referred to as *transcriptional bursting* or *transcriptional pulsing*, is now known to be a common characteristic of gene expression, from bacteria to mammals (Blake *et al.* 2003; Raser and E. J. O'Shea 2004; Rosenfeld *et al.* 2005; Golding *et al.* 2005; Raj *et al.* 2006; Chubb *et al.* 2006; Shahrezaei and Swain 2008; Stevense *et al.* 2010; Larson *et al.* 2013). Furthermore, since the master equation for Ko's ON-OFF model (also known as the *random telegraph* model) was solved exactly under the assumption of ergodicity (Peccoud and Ycart 1995; Raj *et al.* 2006; Iyer-Biswas *et al.* 2009), it has become widely used to explain the variability observed in gene expression data (Raj *et al.* 2006; Zenklusen *et al.* 2008; Suter *et al.* 2011; Larson *et al.* 2013).

As our experimental capabilities progress, the limitations of the two-state ON-OFF

model and its solution are becoming increasingly evident. First, several studies have shown that mechanisms for regulating transcription initiation can be highly complex and combinatorial, and as such they are major sources of gene variability, including nucleosome occupancy, TATA box strength, the number of transcription factor binding sites, and their location on the promoter (Blake *et al.* 2006; Sánchez *et al.* 2008; Sánchez *et al.* 2013; Senecal *et al.* 2014; Jones *et al.* 2014; Francesconi and Lehner 2014; Lin *et al.* 2015). Some regulatory factors can induce different transcription rates, or interact with each other in various ways, so that multi-state extensions of the two-state model are appropriate (Sánchez and Kondev 2008; Coulon *et al.* 2013; Sánchez *et al.* 2013; Senecal *et al.* 2014; Corrigan and Chubb 2014). Second, since the two-state model assumes that the gene state switching events are modelled by a Markov chain with constant transition rates, waiting times in each state are exponentially distributed. However, recent time-lapse recordings of mammalian gene expression show peaked waiting time distributions in the inactive state (Suter *et al.* 2011; Harper *et al.* 2011), suggesting that at least two inactive states or non-Markovian dynamics at the promoter are necessary to capture the observed refractory periods. Third, feedback motifs are known to be a common feature of gene regulatory networks (Shen-Orr *et al.* 2002; Ji *et al.* 2013; Kueh *et al.* 2013), which are not accounted for in the two-state model. Fourth, the model possesses a stationary solution, so is unable to describe the observations of time-varying distributions that are due to, for example, circadian genes (Bieler *et al.* 2014; Lück *et al.* 2014; Suter *et al.* 2011), the cell cycle (Spellman *et al.* 1998; Sigal *et al.* 2006; Singer *et al.* 2014), external signalling (Larson *et al.* 2013; Corrigan and Chubb 2014; Olson *et al.* 2014), or coupling between cells (Feillet *et al.* 2014; Kuramoto 1975).

Unfortunately, exact solutions of the master equation for models of these observations are rare, and are usually stationary solutions obtained via a probability generating function (Hornos *et al.* 2005; Zhang *et al.* 2012; Huang *et al.* 2014). Instead, quantitative modelling and analysis of the stochasticity observed in gene expression data must still be done via the methods known in the 1970s (McQuarrie 1967; Kampen 1992; Gardiner 1985), moment-based measures of noise (Swain *et al.* 2002; Thattai and Oudenaarden 2001), spectral methods (Walczak *et al.* 2009), approximations of the master equation under various assumptions including timescale separations (Shahrezaei and Swain 2008; Thomas *et al.* 2012), numerical solutions (Munsky and Khammash 2006), or stochastic simulations of the time evolution of the cell population (Gillespie 1976; Gillespie 1977). A general framework for gene transcription models that is applicable to non-stationary systems is lacking. To fill this gap is precisely the aim of this thesis.

1.2 Outline of the thesis

The following chapter covers some preliminary material on the biological processes and data that motivate the models we consider, and a simplified derivation of the master equation itself for chemical reactions that are relevant to gene transcription and degradation

of gene products.

Part II gives results within the general framework that we consider, namely models of mRNA transcription and degradation with time-varying rates and cell-to-cell correlations. Chapter 3 lays down this general theoretical framework, and I show that the solution of the master equation in this general setting always takes the form of a Poisson mixture distribution. The random variable described by the mixing density is completely characterised by the transcription and degradation rates of the model. In effect, the general solution decouples the discrete Poisson contribution of the solution that is common to all gene transcription models, from the model-specific extrinsic component that is encapsulated entirely in the mixing density. Obtaining the solution of any gene transcription model is thus reduced to the task of calculating the model-specific mixing density; Chapter 4 gives some intuition for the physical interpretation of the extrinsic component, and the advantages conferred to us by decomposing the solution in this way. In particular, the extrinsic random variable is continuous and naturally encompasses time-dependent correlations between cells, and there are several classical methods for obtaining its probability density. In Chapter 5 the ensemble and temporal moments and measures of noise are discussed as straightforward corollaries of the full Poisson mixture solution.

Part III focuses on the implications of the general results from Part II on the multistate promoter models that are found ubiquitously in the analysis of gene expression data. Chapter 6 derives the solution of the cyclic promoter progression model in two different ways, each giving us different tools for analysis and understanding of the behaviour of the solution. Chapter 7 capitalises on the extra structural knowledge of model solutions that we obtain by determining the mixing density to infer which other models can be solved and the forms of their solutions. In Chapter 8 an exact quantitative characterisation of the timescales implied by ON-OFF model parameters is derived, and the validity of timescale separation arguments is discussed in terms of recent time-lapse experimental data from the literature.

Finally, Chapter 9 in Part IV discusses the connections of our framework to Poisson expansions of feedback models, the potential gains from extending the framework to feedback models, and proposes some initial approaches for pursuing them.

Chapter 2

Preliminaries

The purpose of this chapter is to provide only the necessary background material and and context for what follows in the rest of the thesis. We first give a basic introduction to the processes involved in gene expression and transcriptional regulation, based on Alberts *et al.* 2008, followed by an overview of the types of data that motivate our theoretical work. The following section gives a short derivation of the master equation for the single-species models that we will consider.

2.1 Biological background and context

2.1.1 Gene expression

Biology is a science of incredible diversity, based upon a set of common underlying principles. From the relatively simple, single-celled archaea to large, highly complex organisms, common to all life forms[1] is the cell as the fundamental unit of life. At the cellular level too, there can be hundreds of different cell types that form a single species, yet the underlying principles of structure and function share a number of common traits. In particular, for both prokaryotes (single-celled organisms without a nucleus, i.e. archaea and bacteria) and eukaryotes (organisms whose cells have a nucleus) the mechanisms for storing and using hereditary information that influences form and function of all organisms are surprisingly similar; the information is stored in coded form as a sequence using four letters, within the now famous deoxyribonucleic acid (DNA) molecules.

In the traditional view, the information is divided into units, called *genes*, each of which encodes a defined protein component. However, several more recent findings require the concept of a gene to be revised. For example, alternative splicing allows a single section of DNA to code for multiple proteins, and some genes code for functional products that are not used to synthesize proteins. Since there is no widely accepted consensus on the definition of a gene, for the modelling in this thesis it will be adequate to think of it as a particular locus of a DNA molecule, that is copied to make the products we are interested in.

Although DNA encodes all the necessary information to replicate an organism, the information needs to be *expressed* in order to produce the gene products that are required by the cell. Different cells need different gene products in varying amounts, at different times, and according to its environment, so the cell expresses the information from the

[1] The classification of viruses as non-cellular life forms is still a controversial issue (Forterre 2010).

relevant genes when required, in a process known as *gene expression*. There are several steps involved, but gene expression can be broadly split into two processes: transcription, followed by translation.

In *transcription*, the gene is copied, or *transcribed*, into a ribonucleic acid (RNA) molecule. The resulting RNA molecule is therefore sometimes referred to as a *transcript*. Since the gene can be used repeatedly for transcription, the transcript is a disposable copy of the information stored in the DNA. The RNA molecules that serve as messengers of this information to guide protein synthesis are known as *messenger RNA* (mRNA) molecules, and are the type that we model in this thesis. Molecular complexes called ribosomes *translate* the mRNA molecules into *proteins*, in a process known as *translation*. Depending on its stability, each mRNA molecule can be used to synthesize many protein molecules before being degraded.

We will be concerned with modelling transcription. Since translation does not expend mRNA molecules, in general we will not need to consider translation in a direct manner.

2.1.2 Transcriptional regulation

Every cell needs to be able to control the type and amount of proteins and other gene products, in order to interact with and respond to internal signals, to other cells, and to their environment in general. In a multicellular organism, each DNA molecule has the same sequence and therefore the cell type depends on which genes are expressed, and which aren't. Regulation of gene expression is thus essential for the proper and efficient functioning of any cell.

The main control point for gene regulation is at the start of the gene expression process, where the gene can be turned 'on' or 'off' (see Fig. 2.1 E,F). The strategies and mechanisms used for activating or inhibiting transcription initiation are numerous, and are generally different for prokaryotic and eukaryotic cells. To limit our scope, we will focus on the general mechanisms that motivate our work and are necessary to understand it.

For both prokaryotes and eukaryotes, transcription is initiated when an RNA polymerase molecule binds to the *promoter*, a specialised sequence on the DNA strand that is usually located just upstream of the starting point for transcription. Regulatory proteins called *transcription factors* (TFs) bind to the DNA at specific loci in order to influence the action of the RNA polymerase. In bacteria, genes are 'on' by default, so repressor proteins are their main tools for the regulation of gene expression (Jacob and Monod 1961). The repressor binds to the DNA close to the promoter site, thus blocking the RNA polymerase from binding and initiating transcription. When the repressor unbinds from its position on the DNA strand, for example due to a conformational change caused by the binding of another regulatory molecule, the gene is available for transcription once more (Jacob and Monod 1962).

Transcriptional regulation is a far more complex process in eukaryotic cells, being a combination of the structural effects of the DNA, and more sophisticated interactions of TFs. For eukaryotes, TFs usually activate transcription, and their action is said to be

'combinatorial' (Remenyi *et al.* 2004; Lin *et al.* 2015): regulation requires interactions between several proteins; RNA polymerase is often not able to bind to the promoter without this recruitment by TFs (Struhl 1999). Thus, eukaryotic genes are generally 'off' by default. The recruitment of TFs can be influenced by several factors, including the existence of a conserved gene sequence called a TATA box, silencer elements for repressor molecules, and chromatin structure. Chromatin is the complex formed by the DNA with proteins, which causes the DNA strands to be highly condensed and wrapped around proteins called histones. For TFs to bind to the DNA at the correct positions, the chromatin structure must be unwound into an 'open' domain. Thus even in the presence of several transcription factors, the genes in a closed chromatin domain remain inactive.

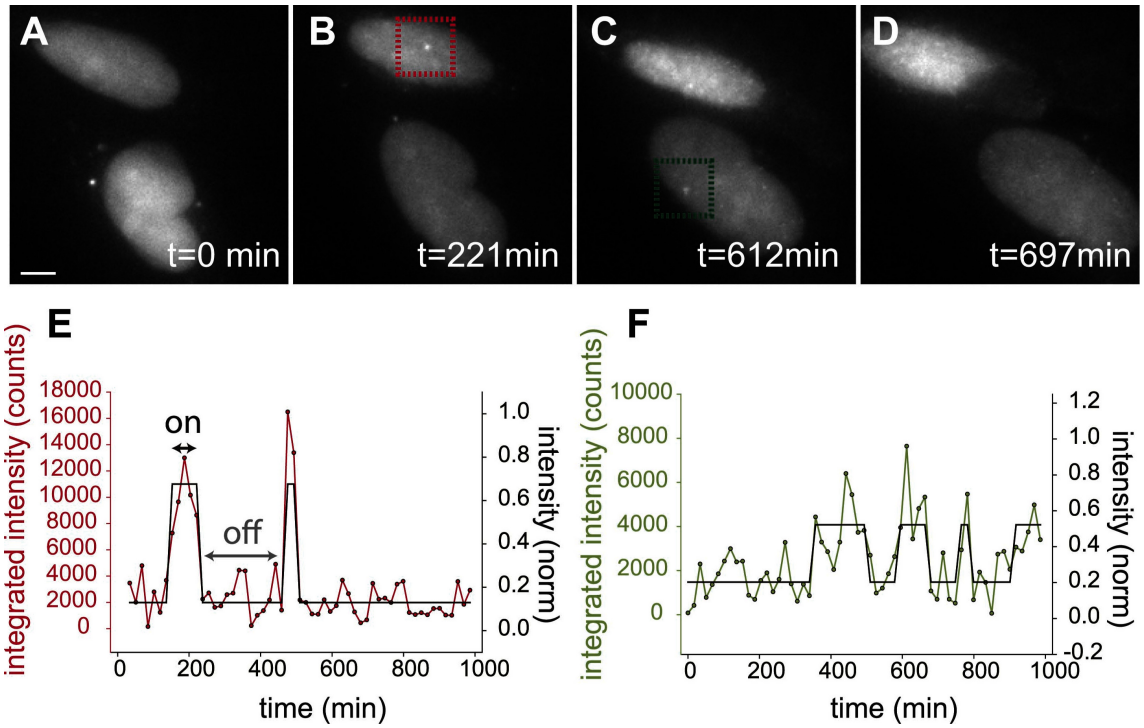

Figure 2.1: Time-lapse microscopy using the MS2-GFP detection system. *(A-D)* Fluorescence microscopy images of human U2-OS cells at different time points. Transcription sites are visible as fluorescent spots. Scale bar = 4 μm. *(E, F)* Fluorescence of single cells as a function of time with highlighted ON and OFF intervals of transcription. Reprinted from eLife, *2*, Larson *et al.*, Direct observation of frequency modulated transcription in single cells using light activation, e100750 (2013) under the Creative Commons CC0 public domain dedication.

2.1.3 Experimental capabilities and data

To help understand gene expression and its regulation, we would ideally like to observe all the relevant molecules and their interactions, in single cells over time. We are still far from this ideal scenario which, given the complexities of gene regulation, may well always remain science fiction. Nevertheless, our experimental capabilities are advancing at an astonishing rate, especially at the mRNA level. Where only a little over a decade ago

single-cell observations of mRNA molecules in living cells were unheard of, we are now able to reliably observe and quantify single molecules in single cells over time.

With this pace of progress in mind, the gene transcription models considered in this thesis are formulated mathematically in order to obtain probability distributions for the mRNA copy number in single cells over time. Since the main motivation of these models is to have the quantitative tools necessary to analyse the single-molecule, time-lapse data that are emerging, we mention here the main single-mRNA detection approaches that form the context for the examples we use in the main body of the text.

Population snapshot methods

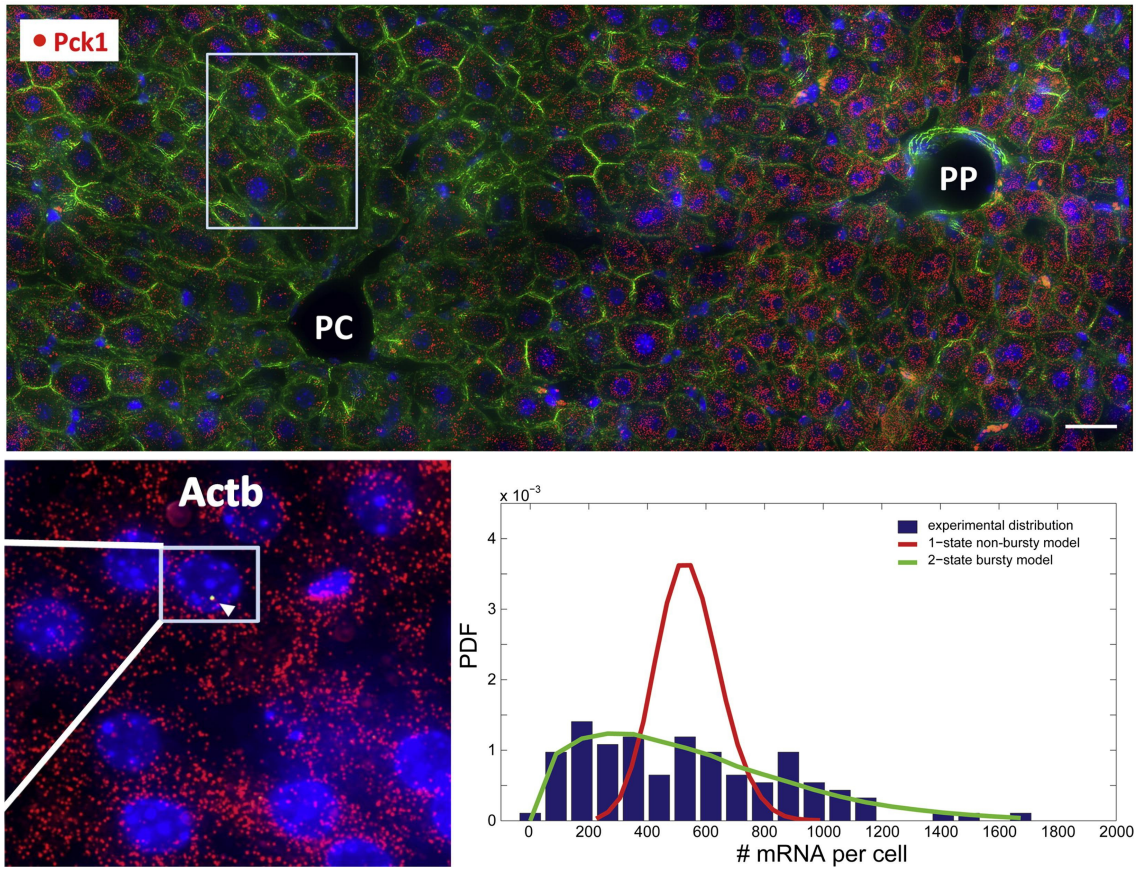

Figure 2.2: Single-molecule population snapshot measurements of mRNA molecules in mouse liver cells, obtained using smFISH. *Top:* Red dots are single mRNA molecules, blue areas are stained nuclei, and green is staining of the cell membrane. PP, periportal zone. PC, pericentral zone. Scale bar is 30 mm. *Below left:* Dual-colour labelling of transcription sites reveal active transcription sites, indicated by the arrowhead. *Below right:* Distribution of the number of mRNA molecules per cell (bars), and theoretical probability density functions (lines) from two different mathematical models. Reprinted from Molecular Cell, *58*, K. Bahar Halpern *et al.*, Bursty gene expression in the intact mammalian liver, 147-156, Copyright 2015, with permission from Elsevier.

Population snapshot approaches tell us the number (or relative number) of mRNA molecules in single cells of a population, at a single time point. The data is usually

represented in the form of bar graphs or histograms (see Fig. 2.2 for an example).

Single-molecule fluorescence in situ hybridization (smFISH) allows one to count mRNA molecules in single cells and view their localisation (Femino *et al.* 1998; Raj and Oudenaarden 2008; Little *et al.* 2013; Lyubimova *et al.* 2013). Samples are fixed and then hybridised using a set of fluorescently labelled oligonucleotides. Although this method provides only a single-time picture of the expression levels, the resolution of the molecules is unparalleled, even with dual-colour labelling (Bahar Halpern *et al.* 2015).

Single-cell quantitative reverse transcriptase polymerase chain reaction (RT-PCR), and digital single-cell RT-PCR are highly sensitive methods for counting mRNA molecules in hundreds of single cells. The method yields absolute numbers, rather than arbitrary fluorescence units, but does not give us any spatial information (Bengtsson *et al.* 2005; Warren *et al.* 2006; Ståhlberg *et al.* 2012).

Time-lapse methods

Time-lapse methods rely on engineering cells so that the transcripts of interest fluoresce, usually using fluorescent 'tagging' proteins. Previously, resolution at the mRNA level was low, so we would infer the number of molecules present in a single cell by the fluorescence intensity. However, methods have improved drastically in the past few years, so we are now able to track single molecules (Fig 2.1 A-D), as well as obtain time-lapse traces that indicate the expression levels of single cells over time (Fig 2.1 E-F, for example). As such, we are starting to overcome some of the usual limitations associated with fluorescence measurements, i.e., background fluorescence, low resolution, and inferring molecule numbers from arbitrary fluorescence units. However, the confounding effects of the fluorescent tags must still be kept in mind.

The MS2-GFP mRNA detection system is the predominant procedure for the direct observation of single mRNA molecules in living cells, over time (Bertrand *et al.* 1998; Golding and Cox 2004; Golding *et al.* 2005; Muramoto *et al.* 2012; Larson *et al.* 2013; Coulon *et al.* 2014; Corrigan and Chubb 2014). The gene is engineered to encode for a number of RNA hairpins, and an MS2-coat protein is tagged with a fluorescent protein, usually GFP, that binds with high affinity to the RNA loops (Bertrand *et al.* 1998; Golding and Cox 2004). When the MS2-GFP fusion binds to the mRNA, the molecules fluoresce and can be followed over time using time-lapse microscopy (see Fig. 2.1). Two-colour single-mRNA imaging is also emerging (Coulon *et al.* 2014). The main disadvantage for this method is that degradation of the reporter molecules is strongly inhibited, the experiments are constrained to timescales of only a few hours, and relatively few cells can be imaged (Golding *et al.* 2005).

High-throughput single-molecule assays have been developed recently, although the method has so far been constrained to *in vitro* studies (Chong *et al.* 2014). The method allows us to monitor the process of transcription itself, so that we can view transcription initiation time points, and measure how long it takes to produce each transcript. A final approach worth mentioning, pioneered by the Naef lab, is to measure the fluorescence

of short-lived proteins (Suter *et al.* 2011; Molina *et al.* 2013). The protein dynamics mimic those of the mRNA molecules because of the short half-lives. The cells can then be tracked over days instead of hours, and the data is not skewed by the slow degradation of the reporter molecules.

2.2 Master equations for gene transcription

In Chapter 1 we laid down the motivations for modelling gene expression with master equations, and developing a framework for solving them. The purpose of this section is to outline the arguments used to derive the master equation, and to acknowledge the underlying assumptions that are made, so that in later chapters we can start from this basis and immediately write down the master equation for a given model of gene transcription.

2.2.1 Derivation of the master equation

We are interested in modelling N_t, the number of mRNA molecules in a single cell over time. It is understood that N_t refers only to mRNA molecules that were transcribed at a specific gene that we are interested in. Since the number of transcripts of any particular kind in a single cell can be very low[2], we do not permit N_t to be approximated by a continuous variable. In other words, N_t takes only non-negative integer values.

As mentioned in Chapter 1, and illustrated by the experimental data in Figs 2.1 and 2.2, N_t changes stochastically in time. To accommodate our discrete, stochastic state variable N_t mathematically, we need to describe its value by $P(n,t) \coloneqq \Pr(N_t = n)$, the *probability* that N_t has the value $n \in \mathbb{N}$ at time t.

Starting from the collision theory of chemical kinetics, that chemical reactions occur when two molecules collide with sufficient energy, we can derive an equation for the time evolution of $P(n,t)$ in a well-stirred system at thermal equilibrium (Gillespie 1992a). Equations of this kind are known generally as Kolmogorov forward equations, but when the random variable is discrete, as we have here, they are more often known as *chemical master equations*, or simply *master equations*.

It is now known that the cell is a densely crowded, compartmentalised system, so that molecules can not freely diffuse within it. Nevertheless, the master equation as derived using arguments based on conventional chemical reaction rates has proved incredibly useful for characterising and explaining stochasticity in gene expression (McQuarrie 1967). With this caveat in mind, we briefly outline these arguments to derive master equations for models of transcription for a single gene. Extensions for systems with more than one chemical species follow in a similar manner; see any of the references Gardiner 1985; Kampen 1992; Gillespie 1992b for derivations with more generality.

Let $o(dt)$ denote terms such that $\lim_{dt\to 0} o(dt)/dt = 0$, and let $n \to m$ denote the event of a transition from state n to state m. We make the following assumptions:

[2] Average mRNA copy numbers usually range from less than one to a few hundred (Milo and Phillips 2015); the median in mammalian cells has been reported as being only 17 (Schwanhäusser *et al.* 2011).

i. The probability of a transition $n \to n+1$ in the infinitesimal time interval $(t, t+dt)$ is $a(n,t)dt + o(dt)$, where $a(n,t) \in \mathbb{R}_{\geq 0}$.

ii. The probability of a transition $n \to n-1$ in the infinitesimal time interval $(t, t+dt)$ is $b(n,t)dt + o(dt)$, where $b(n,t) \in \mathbb{R}_{\geq 0}$.

iii. The probability of a transition $n \to n \pm j,\ j > 1$, in the time interval $(t, t+dt)$ is $o(dt)$.

Notice that under these assumptions, the system has the *Markov property*: the transition probabilities depend only on the current state of the system.

Now $P(n, t+dt)$ is given by the sum of:

i. the probability that the system is in state n at time t, and remains there until $t+dt$,

ii. the probability that the system is in state $n-1$ at time t, and transitions to state n by $t+dt$

iii. the probability that the system is in state $n+1$ at time t, and transitions to state n by $t+dt$, and

iv. the probability that the system is in state $n \pm j,\ j > 1$ at time t, and transitions to state n by $t+dt$.

Using the multiplication law of probability, i.e. $\Pr(A \text{ and } B) = \Pr(A) \times \Pr(B \text{ given } A)$, we have

$$\begin{aligned} P(n,t+dt) &= P(n,t) \times [1 - \Pr\{n \to n \pm j,\ j \geq 1,\ \text{in } (t,t+dt)\}] \\ &\quad + \Pr\{n-1 \to n \text{ in } (t,t+dt)\} \\ &\quad + \Pr\{n+1 \to n \text{ in } (t,t+dt)\} \\ &\quad + \Pr\{n \pm j \to n,\ j \geq 1,\ \text{in } (t,t+dt)\} \\ &= P(n,t)[1 - a(n,t)dt - b(n,t)dt - o(dt)] \\ &\quad + P(n-1,t)[a(n,t)dt + o(dt)] \\ &\quad + P(n+1,t)[b(n,t)dt + o(dt)] + o(dt). \end{aligned}$$

Moving the $P(n,t)$ term from the right hand side of the equation to the left, dividing by dt, and taking the limit $dt \to 0$, we obtain the master equation

$$\frac{dP(n,t)}{dt} = -[a(n,t) + b(n,t)]P(n,t) + a(n-1,t)P(n-1,t) + b(n+1,t)P(n+1,t).$$

Suppose the state transition $n \to n+1$ can only happen via a transcription event, and the state transition $n \to n-1$ can only happen via a degradation event. Suppose also that mRNA molecules do not directly influence the processes of transcription at their gene, and degradation. Then transcription is a *zero-order* chemical reaction: we have

$a(n,t) = \alpha(t)$. On the other hand, degradation is a *first-order* chemical reaction: for each mRNA molecule, the probability that it degrades in $(t, t+dt)$ is $\beta(t)dt + o(dt)$, $\beta(t) \in \mathbb{R}_{\geq 0}$, so with n mRNA molecules present in a system, the probability of a transition $n \to n-1$ in $(t, t+dt)$ is $\beta(t)ndt + o(dt)$, i.e. $b(n,t) = \beta(t)n$. We are then left with

$$\frac{dP(n,t)}{dt} = -[\alpha(t) + \beta(t)n]P(n,t) + \alpha(t)P(n-1,t) + \beta(t)(n+1)P(n+1,t).$$

Example 2.2–1 walks through the steps of solving a simple master equation model of gene transcription, using the probability generating function method that we will use in the next chapter.

Example 2.2–1 — Gene transcription model with constant rates

The most simple gene transcription model assumes a constant transcription rate and linear degradation, i.e. $a(n,t) = \alpha$ and $b(n,t) = \beta n$, where $\alpha, \beta > 0$ are constants. The master equation for this model is therefore

$$\frac{dP(n,t)}{dt} = -[\alpha + \beta n]P(n,t) + \alpha P(n-1,t) + \beta(n+1)P(n+1,t). \tag{2.1}$$

The usual approach for obtaining $P(n,t)$ is to use the probability generating function, namely,

$$G(z,t) := \sum_{n=0}^{\infty} z^n P(n,t),$$

to transform the master equation (2.1) into a partial differential equation (Bartholomay 1958; Kampen 1992; Gardiner 1985). It is fairly straightforward to check that:

$$\begin{aligned}
\sum z^n \frac{P(n,t)}{dt} &= \frac{\partial G}{\partial t} \\
\sum n z^n P(n,t) &= z \sum \frac{\partial}{\partial z} z^n P(n,t) = z\frac{\partial G}{\partial z} \\
\sum z^n P(n-1,t) &= z \sum z^{n-1} P(n-1,t) = zG \\
\sum (n+1) z^n P(n+1,t) &= \frac{\partial G}{\partial z}.
\end{aligned}$$

so if we multiply Eq. (2.1) through by z^n and sum over all n, we can use the identities above to find that G satisfies

$$\frac{\partial G}{\partial t} = \alpha(z-1)G - \beta(z-1)\frac{\partial G}{\partial z}.$$

We can solve this equation for G using the method of characteristics with initial con-

dition $P(0,0) = 1$, $P(n,0) = 0$ for $n > 0$, to find

$$G(z,t) = e^{\frac{\alpha}{\beta}[1-e^{-\beta t}](z-1)}. \tag{2.2}$$

Recall that, by definition, $P(n,t)$ is the coefficient of z^n in $G(z,t)$ so by expanding Eq. (2.2) we obtain

$$P(n,t) = \frac{1}{n!}\left(\frac{\alpha}{\beta}\left[1-e^{-\beta t}\right]\right)^n e^{\frac{\alpha}{\beta}[1-e^{-\beta t}]},$$

i.e.

$$N_t \sim \text{Poi}\left(\frac{\alpha}{\beta}\left[1-e^{-\beta t}\right]\right).$$

Part II

Exact solution of the master equation

Part II: Exact solution of the master equation

Introduction

Fundamentally speaking, expression of a single gene can be measured in terms of the number of mRNA molecules present that were transcribed at the gene in question. Since mRNA copy number increases only via transcription events and decreases via degradation events, in a coarse-grained sense gene transcription models simply describe the processes of transcription and degradation. Specificity and complexity of any *particular* model are defined by the assumptions and levels of detail used to describe those processes for the context in which the model will be used.

In this part we work within the following simple framework: we consider the processes of transcription and degradation in a very general sense, and infer the resulting probability distribution of the mRNA copy number over time, with several corollaries. The constraints on the mathematical or statistical descriptions of transcription and degradation are weak, so the results are applicable to most types of transcription models considered in the field. In particular, time-dependence due to correlated expression between cells in the population can be naturally accounted for here, a property that is often lacking in gene transcription models as stationarity is so often assumed. The chapters in Part III will apply the general results given here to more specific descriptions of transcription and degradation.

In Chapter 3, we describe the theoretical framework and derive the main result that will underpin the rest of the thesis: that the probability mass function of the mRNA copy number is always a Poisson mixture. In Chapter 4 we give some physical intuition for the mixture distribution, place it into the context of existing approaches for solving models of gene transcription, and collect together results from several dissociated fields that can help obtain it. Chapter 5 contains corollaries of the result in Chapter 3 to do with moments and measures of noise.

Chapter 3

Derivation of the solution

This chapter consists of the foundations that will underpin the rest of this work. We describe the general, theoretical framework and model, and derive the main result that will be used throughout the thesis. Despite its central importance, the mathematics required here are fairly basic elements of probability theory and stochastic processes. However, it is the tool that has allowed us to understand what is already known about most gene transcription models, as specific cases of this result. This understanding has also guided us towards all the new results that will be discussed in the rest of the thesis, as it allows us to decouple the "Poisson" contribution that is common to all gene transcription models of the kind we consider, from the model-specific contribution.

3.1 Theoretical framework

We are interested in the expression in time of a specific gene, measured by the number of mRNA molecules present in each cell of a population that were transcribed at that gene. We assume that the mRNA copy number changes only via transcription and degradation events, both of which occur according to the principles of stochastic chemical kinetics described in Chapter 2. We will refer to models of this broad category as *transcription-degradation models*.

Our measure for gene expression, the mRNA copy number, is a stochastic process in continuous time t which we will denote by

$$\mathcal{N} \equiv \{N_t\}_{t\geq 0} \equiv \{t \mapsto \eta_\omega(t) : \omega \in \Omega_\mathcal{N}\}.$$

$\{N_t\}_{t\geq 0}$ is a set of random variables N_t, one for each time t, and $\{t \mapsto \eta_\omega(t) : \omega \in \Omega_\mathcal{N}\}$ is the set of all possible sample paths for the mRNA copy number in a single cell. We will denote the probability mass function for N_t by

$$P(n,t) \equiv \Pr(N_t = n),\ n \in \mathbb{N}.$$

If $\mathcal{N}$ is a stationary process so that the law of N_t is independent of t, we will simply write $P(n) \equiv \Pr(N_t = n)$.

In order to account for the variability in the transcription and degradation reaction rates, both in time and between cells, we allow them to be described by the stochastic

processes

$$\mathcal{M} \equiv \{M_t\}_{t\geq 0} \equiv \{t \mapsto \mu_\omega(t) : \omega \in \Omega_\mathcal{M}\} \text{ and}$$
$$\mathcal{L} \equiv \{L_t\}_{t\geq 0} \equiv \{t \mapsto \lambda_\omega(t) : \omega \in \Omega_\mathcal{L}\},$$

respectively. The notation here is the same as that for $\mathcal{N}$: M_t and L_t are random variables, $t \mapsto \mu_\omega(t)$ and $t \mapsto \lambda_\omega(t)$ are sample paths of the transcription and degradation rates, respectively, and $\Omega_\mathcal{M}$ and $\Omega_\mathcal{L}$ are the sample spaces. For continuous random variables such as M_t and L_t, we will denote the cumulative density function (cdf) and probability density function (pdf) according to usual conventions with a capital F and lower-case f, respectively. For example $F_{M_t}(m,t)$ and $f_{M_t}(m,t)$ are the cdf and pdf of M_t at time t. We will sometimes refer to the transcription and degradation rates collectively as *cellular drives*, or simply *drives*.

We do not specify any functional form for the cellular drives here, except to require that $t \mapsto \mu_\omega(t)$ and $t \mapsto \lambda_\omega(t)$ do not depend on the number of mRNA molecules already present. Thus these sample paths may be thought of as example functions of time that describe the transcription and degradation rates in response to changing cellular architecture and environmental conditions in an example cell. Note that by allowing the drives to be described by stochastic processes in this way, we have the freedom to explicitly specify how correlated we would like them to be[1]. In general, we will mention correlation between sample paths in a broad manner, rather than in precise, quantitative terms. Therefore, to avoid confusion with statistical measures of correlation and dependence, we will refer to cells or cellular drives as having a "degree of synchrony". A population will be *perfectly synchronous* if the sample paths of the drives for every cell in the population is identical, or equivalently, if M_t and L_t have zero variance. For example, if the transcription rate is a deterministic sinusoidal function $t \mapsto g(t)$ and is perfectly synchronous for all cells, we would have $f_{M_t}(m,t) = \delta(g(t) - m)\ \forall\ t \geq 0$, where δ is the Dirac delta function. If, however, cellular variability means that the transcription rate will be slightly out of phase between cells, we can impose that f_{M_t} be a more dispersed density: the wider the density, the more asynchronous the drive is. This control is discussed in Chapter 4 with several examples.

Note also that this framework generalises models that account for each step of transcription or degradation individually: the only gene expression model without a time-dependent drive is the standard first-order model with constant transcription and degradation rates μ and λ, and stationary solution $N_t \sim \text{Poi}(\mu/\lambda) \forall\ t$ (see Example 3.2–1). All other models implicitly or explicitly have drives that are not constant.

Ultimately, we would like to obtain the probability distribution of N_t for all t, within this context of time-varying and/or stochastic parameters. The following section derives an expression for $P(n,t)$ in the general case, which will be used in Chapters 4 and 5 to

[1] In fact, by defining $\mathcal{M}$ and $\mathcal{L}$ by the random variables M_t and L_t for each t, we can specify any statistical characteristics of the population that we choose.

derive properties of transcription-degradation models, and then throughout the rest of the thesis to solve and analyse several specific gene transcription models.

3.2 The exact Poisson mixture solution

We first consider a population whose cellular drives are perfectly synchronous, and use the information we gain to derive the probability distribution $P(n,t)$ for the general case.

Perfectly synchronous case: no cell-to-cell variation in cellular drives

Consider first a population of cells with perfectly synchronous transcription and degradation rate functions $\mathcal{M}_{\text{sync}} = \{\mu(t)\}_{t\geq 0}$ and $\mathcal{L}_{\text{sync}} = \{\lambda(t)\}_{t\geq 0}$. In other words, every cell in the population has transcription rate $t \mapsto \mu(t)$ and degradation rate $t \mapsto \lambda(t)$, so M_t and L_t have zero variance $\forall\, t \geq 0$. We thus have an immigration-death process that can be summarized by the following reaction diagram:

$$\emptyset \xrightarrow{\mu(t)} \text{mRNA} \xrightarrow{\lambda(t)} \emptyset. \tag{3.1}$$

Note that this is not a birth-death process because the mRNA do not reproduce; so the transcription rate $\mu(t)$ has units t^{-1}. In contrast, degradation is a first-order reaction, so the unit for the degradation rate is $t^{-1}n^{-1}$.

Denote the probability mass function for the mRNA copy number N_t for this special case with synchronous drives by

$$P_{\text{sync}}(n,t) := \Pr\left(N_t = n | \mathcal{M}_{\text{sync}}, \mathcal{L}_{\text{sync}}\right).$$

The master equation for $P_{\text{sync}}(n,t)$ is then given by:

$$\frac{d}{dt}P_{\text{sync}}(n,t) = \mu(t)P_{\text{sync}}(n-1,t) + (n+1)\lambda(t)P_{\text{sync}}(n+1,t) - \left[\mu(t) + n\lambda(t)\right]P_{\text{sync}}(n,t). \tag{3.2}$$

Using the definition of the probability generating function

$$G(z,t) := \sum_{n=0}^{\infty} z^n P_{\text{sync}}(n,t),$$

we can transform the master equation (3.2) into

$$\frac{\partial G}{\partial t} = (z-1)\mu(t)G - (z-1)\lambda(t)\frac{\partial G}{\partial z}. \tag{3.3}$$

We will solve this equation using the method of characteristics, assuming for now that there are initially n_0 mRNA molecules in each cell. Introduce the characteristic equations

$$\frac{dt}{ds} = 1, \quad \frac{dz}{ds} = (z-1)\lambda(t), \quad \frac{dG}{ds} = (z-1)\mu(t)G,$$

where s is a parametrisation of the characteristic curve, and without loss of generality we can let $t(0) = 0$. Solving this system, we have

$$\begin{aligned} t &= s, \\ z - 1 &= [z(0) - 1]\, e^{\int_0^s \lambda(\sigma)\, d\sigma}, \quad \text{and} \\ G(s) &= G(0)\, e^{[z(0)-1] \int_0^s \mu(\tau) e^{\int_0^\tau \lambda(\sigma)\, d\sigma}\, d\tau}. \end{aligned}$$

Noticing that the initial condition $P(n_0, 0) = 1$ implies that $G(0) = G(z(0), t(0)) = z(0)^{n_0}$, and eliminating the parameter s, we obtain the solution:

$$\begin{aligned} G(z, t|n_0) &= \left[(z-1)e^{-\int_0^t \lambda(\tau)\, d\tau} + 1\right]^{n_0} e^{x(t)(z-1)} \\ &=: G_{N_t^{ic}}(z, t)\, G_{N_t^s}(z, t), \end{aligned} \tag{3.4}$$

where

$$x(t) := \int_0^t \mu(\tau) e^{-\int_\tau^t \lambda(\tau')\, d\tau'}\, d\tau. \tag{3.5}$$

We will refer to x as the *extrinsic parameter*.

Notice that $G(z, t|n_0)$ (Eq. (3.4)) is the product of two probability generating functions: $G_{N_t^{ic}}(z, t)$ is the probability generating function for a binomial random variable N_t^{ic}, with n_0 trials and success probability $e^{-\int_0^t \lambda(\tau)\, d\tau}$, and $G_{N_t^s}(z, t)$ is the probability generating function for a Poisson random variable N_t^s, with parameter $x(t)$. So for this special case where all cells have perfectly synchronous drives, we deduce (Feller 1968) that

$$N_t = N_t^{ic} + N_t^s\,; \tag{3.6}$$

$$N_t^{ic} \sim \text{Bin}\left(n_0, e^{-\int_0^t \lambda(\tau) d\tau}\right), \tag{3.7}$$

$$N_t^s \sim \text{Poi}\left(x(t)\right). \tag{3.8}$$

The physical interpretation of this breakdown is that N_t^{ic} describes the mRNA transcripts that were initially present in the cell and still remain at time t, and N_t^s describes the number present in the cell that were transcribed since $t = 0$.

N_t^{ic} and N_t^s are independent, so it is easy to read off the first two moments:

$$\begin{aligned} \mathbb{E}\left(N_t | \mathcal{M}_{\text{sync}}, \mathcal{L}_{\text{sync}}, n_0\right) &= \mathbb{E}\left(N_t^{ic}\right) + \mathbb{E}\left(N_t^s\right) \\ &= n_0\, e^{-\int_0^t \lambda(\tau) d\tau} + x(t), \\ \text{Var}(N_t | \mathcal{M}_{\text{sync}}, \mathcal{L}_{\text{sync}}, n_0) &= \text{Var}(N_t^{ic}) + \text{Var}(N_t^s) \\ &= n_0 e^{-\int_0^t \lambda(\tau)\, d\tau} \left(1 - e^{-\int_0^t \lambda(\tau)\, d\tau}\right) + x(t). \end{aligned}$$

We can go further and write down the explicit expression for $P_{\text{sync}}(n, t|n_0)$, either by

computing the coefficient of z^n in $G(z,t)$, or by using Eqs (3.6)-(3.8) to write

$$\begin{aligned} P_{\text{sync}}(n,t|n_0) &= \Pr(N_t^{ic} + N_t^s = n|n_0) \\ &= \sum_{k=0}^{n} \Pr(N_t^{ic} = k|n_0) \Pr(N_t^s = n-k) \\ &= \sum_{k=0}^{n} \binom{n_0}{k} e^{-k\int_0^t \lambda(\tau)d\tau} \left(1 - e^{-\int_0^t \lambda(\tau)d\tau}\right)^{n_0-k} \frac{x(t)^{n-k}}{(n-k)!} e^{-x(t)} . \end{aligned}$$

If the initial condition n_0 is not known, or if the state of each cell at $t = 0$ is not identical, it can be more appropriate to let the initial state be described by a random variable N_0. In this case, the law of total probability gives us

$$P_{\text{sync}}(n,t) = \sum_{n_0} P_{\text{sync}}(n,t|n_0) \Pr(N_0 = n_0) . \tag{3.9}$$

Notice that the contribution from N_t^{ic} decreases exponentially for any sensible degradation functions λ, as the transcripts that were present at $t = 0$ are expected to degrade, and the population is expected to be composed only of mRNA molecules that were transcribed since $t = 0$. Hence in this case where all cells are initialised in the same state, the distribution $P_{\text{sync}}(n,t|n_0)$ displays a transient period until the contribution from N_t^{ic} is negligible. However, $N_t = N_t^{ic} + N_t^s$ does not *necessarily* display transient behaviour until the contribution from N_t^{ic} is negligible. Let us first fix ideas using a model with a stationary solution, and then discuss the general case.

Example 3.2–1 — Time-dependent contributions from N_t^{ic} and N_t^s can balance each other

Consider the gene transcription model with constant transcription and degradation rates μ and λ. Then, when there are initially n_0 mRNA transcripts, Eqs (3.6)-(3.8) give us

$$\begin{aligned} N_t^{ic} &\sim \text{Bin}\left(n_0, e^{-\lambda t}\right), \\ N_t^s &\sim \text{Poi}\left(\frac{\mu}{\lambda}\left[1 - e^{-\lambda t}\right]\right). \end{aligned}$$

As $t \to \infty$ the distribution of N_t will tend towards $\text{Poi}(\mu/\lambda)$, the stationary distribution of the population. We will show that if the system starts at stationarity, $N_t = N_t^{ic} + N_t^s$ will be independent of time $\forall\, t \geq 0$.

Suppose that the population starts at stationarity, i.e.

$$\Pr(N_0 = n_0) = \left(\frac{\mu}{\lambda}\right)^{n_0} \frac{e^{-\mu/\lambda}}{n_0!}\,.$$

We wish to find $P_{\text{sync}}(n,t)$ given this initial distribution. We still have

$$N_t^s \sim \text{Poi}\left(\frac{\mu}{\lambda}\left[1 - e^{-\lambda t}\right]\right),$$

so we just need to determine the contribution to N_t from N_t^{ic}. In this case, we have

$$\begin{aligned}
\Pr(N_t^{ic} = k) &= \sum_{n_0=k}^{\infty} \Pr(N_t^{ic} = k | N_0 = n_0) \Pr(N_0 = n_0) \\
&= \sum_{n_0=k}^{\infty} \binom{n_0}{k} \left(e^{-\lambda t}\right)^k \left(1 - e^{-\lambda t}\right)^{n_0-k} \left(\frac{\mu}{\lambda}\right)^{n_0} \frac{e^{-\mu/\lambda}}{n_0!} \\
&= \left(e^{-\lambda t}\right)^k e^{-\mu/\lambda} \sum_{r=0}^{\infty} \left(\frac{\mu}{\lambda}\right)^{r+k} \frac{\left(1 - e^{-\lambda t}\right)^r}{k!\, r!} \\
&= \left(\frac{\mu}{\lambda}\right)^k \frac{\left(e^{-\lambda t}\right)^k e^{-\mu/\lambda}}{k!} \sum_{r=0}^{\infty} \left(\frac{\mu}{\lambda}\right)^r \frac{\left(1 - e^{-\lambda t}\right)^r}{r!} \\
&= \left(\frac{\mu}{\lambda} e^{-\lambda t}\right)^k \frac{e^{-\mu/\lambda}}{k!} e^{\frac{\mu}{\lambda}(1 - e^{-\lambda t})} \\
&= \left(\frac{\mu}{\lambda} e^{-\lambda t}\right)^k \frac{e^{-\frac{\mu}{\lambda} e^{-\lambda t}}}{k!}\,.
\end{aligned}$$

In other words $N_t^{ic} \sim \text{Poi}\left(\frac{\mu}{\lambda} e^{-\lambda t}\right)$, which cancels the time dependent contribution from $N_t^s \sim \text{Poi}\left(\frac{\mu}{\lambda}\left[1 - e^{-\lambda t}\right]\right)$. Therefore $\forall\, t \geq 0$,

$$N_t \sim \text{Poi}\left(\frac{\mu}{\lambda}\right).$$

The example shows that N_t^{ic} and N_t^s will combine to reproduce a stationary distribution at all times $t > 0$, when the system starts at stationarity. Thus we do not have a transient period where the initial distribution converges to the equilibrium of the system.

The same is true if the system is not stationary: suppose we let the system start at $t = -\infty$, with any initial condition. For $t > 0$, the system will be independent of the initial condition and will be described by N_t^s. Let us denote the state of the system for $t > 0$ by the *attracting* distribution P_*. Although $\Pr(N_t^s = n) = P_*(n,t)\ \forall\, n \in \mathbb{N}, \forall\, t > 0$, we wish to distinguish P_* from P_s because we only have equality of the two distributions when the

system starts at $t = -\infty$. P_* can be thought of as an inherent property of the system, analogous to the the stable point of a dynamical system that moves in time (sometimes called a *chronotaxic system* (Suprunenko *et al.* 2013)).

Now, if we let $\Pr(N_0 = n_0) = P_*(n_0, 0)\ \forall\ n_0$ in Eq. (3.9), the contributions from N_t^{ic} and N_t^s balance each other as they did in Example 3.2–1, and we have $P_{\text{sync}}(n,t) = P_*(n,t)\ \forall\ n,\ \forall\ t > 0$. Thus we only observe an initial transient period if the initial distribution starts away from its attracting distribution at $t = 0$. In all other cases, the following mathematical formulations are equivalent: *i)* assume that the system was initialised at $t = -\infty$ and consider only N_t^s, or *ii)* use the initial distribution $P_*(n,0)\ \forall\ n$ at $t = 0$, and consider $N_t^{ic} + N_t^s$.

We wish to focus on the time dependence of $P(n,t)$ induced through non-stationarity of the parameters, and/or synchronous behaviour of the cells within the population. So, unless otherwise stated, we will assume that the system was initialised at $t = -\infty$ and that the distribution of N_t^s is the attracting distribution $P_*(n,t)\ \forall\ t > 0$. A discussion of the transient period due to initial conditions can be found in Section 5.2. We will therefore neglect the contribution from N_t^{ic}, and use $N_t \sim \text{Poi}(x(t))$ when the drives for the population are perfectly synchronous, i.e.

$$P_{\text{sync}}(n,t) = \frac{x(t)^n}{n!}e^{-x(t)}, \quad x(t) := \int_0^t \mu(\tau)e^{-\int_\tau^t \lambda(\tau')\,d\tau'}\,d\tau. \tag{3.10}$$

General case: cell-to-cell variation in cellular drives

Let us return to the general case where several sample paths are possible, implying that the transcription and degradation rates within the population have some degree of asynchrony: $\mathcal{M} \equiv \{t \mapsto \mu_\omega(t)\}_{\omega\in\Omega_\mathcal{M}}$ and $\mathcal{L} \equiv \{t \mapsto \lambda_\omega(t)\}_{\omega\in\Omega_\mathcal{L}}$, so that M_t and L_t have non-zero variance for at least some $t \geq 0$. The results above for the special case of synchronous drives show that the relevant stochastic process to consider is in fact $\mathcal{X}$, where

$$\begin{aligned}\mathcal{X} \equiv \{X_t\}_{t\geq 0} &\equiv \{t \mapsto x_\omega(t) : \omega \in \Omega_\mathcal{X}\} \\ &= \left\{t \mapsto \int_0^t \mu_{\omega'}(\tau)e^{-\int_\tau^t \lambda_{\omega''}(\tau')\,d\tau'}\,d\tau : \omega' \in \Omega_\mathcal{M},\ \omega'' \in \Omega_\mathcal{L}\right\}.\end{aligned}$$

Notice that for all sensible cellular drives (for example, positive and finite), $\mathcal{X}$ is a continuous stochastic process. The properties of $\mathcal{X}$ are discussed in more detail in the following chapter.

Now, Eq. (3.10) tells us that the conditional distribution of N_t given $X_t = \xi$ is

$$P(n,t|X_t = \xi) = \frac{\xi^n}{n!}e^{-\xi},$$

hence to find the probability distribution for N_t we can simply take the expectation over the distribution for X_t. This brings us to our central result:

Poisson mixture result

Denote the sets of all possible sample paths for the transcription and degradation rates by

$$\{t \mapsto \mu_\omega(t) : \omega \in \Omega_\mathcal{M}\} \quad \text{and} \quad \{t \mapsto \lambda_\omega(t) : \omega \in \Omega_\mathcal{L}\},$$

respectively. Let $\mathcal{X} \equiv \{X_t\}_{t\geq 0}$ be the stochastic process defined by the collection of sample paths

$$\{t \mapsto x_\omega(t) : \omega \in \Omega_\mathcal{X}\} := \left\{t \mapsto \int_0^t \mu_{\omega'}(\tau) e^{-\int_\tau^t \lambda_{\omega''}(\tau')\, d\tau'}\, d\tau : \omega' \in \Omega_\mathcal{M},\ \omega'' \in \Omega_\mathcal{L}\right\}.$$

Then the probability mass function for the random variable N_t is given by the *Poisson mixture* (or compound) distribution

$$P(n,t) = \int \frac{\xi^n}{n!}\, e^{-\xi}\, f_{X_t}(\xi,t)\, d\xi, \tag{3.11}$$

$\forall\, t \geq 0$ if the system starts at its attracting distribution at $t = 0$. Otherwise, the probability mass function converges to Eq. (3.11). f_{X_t} is the probability density function of the random variable X_t, which we will refer to as the *mixing density*.

Equation (3.11) will be used throughout this thesis, and will be referred to as the *Poisson mixture result*. Note that when the drives are perfectly synchronous and $X_t = x(t)$, the mixture distribution is a Dirac delta function, $f_{X_t}(\xi,t) = \delta(x(t) - \xi)$, so Eq. (3.11) reduces to Eq. (3.10).

3.3 Discussion

For a very broad class of gene transcription models, the Poisson mixture result explicitly indicates two separate sources of variability in gene expression, and how they combine in the distribution of N_t. First, variation due to the Poisson processes of transcription and degradation, that is common to every model of the form shown in the reaction diagram (3.1). Second, variation or uncertainty of a certain combination of the transcription and degradation rates that we defined, which is model-dependent. In this sense, Eq. (3.11) extends the concept of separable 'intrinsic' and 'extrinsic' components of gene expression noise pioneered by Swain *et al.* 2002, to the full distribution. In fact, using the laws of total expectation and total variance (Weiss 2005), we recover the additive property of the noise η^2:

$$\begin{aligned}
\mathbb{E}(N_t) &= \mathbb{E}_{X_t}\left[\mathbb{E}_{N_t|X_t}(N_t|X_t)\right] = \mathbb{E}(X_t),\\
\mathrm{Var}(N_t) &= \mathbb{E}\left[\mathrm{Var}(N_t|X_t)\right] + \mathrm{Var}\left[\mathbb{E}(N_t|X_t)\right] = \mathbb{E}(X_t) + \mathrm{Var}(X_t),
\end{aligned}$$

so

$$\eta^2 = \frac{\mathrm{Var}(N_t)}{\mathbb{E}\left(N_t\right)^2} = \frac{1}{\mathbb{E}\left(N_t\right)} + \frac{\mathrm{Var}(X_t)}{\mathbb{E}\left(X_t\right)} = \eta_{int}^2 + \eta_{ext}^2.$$

Note that the Poisson mixture result as an extension of the concepts of intrinsic and extrinsic noise emerged naturally from our initial assumptions, as did the relevant parameter combination X_t to describe the extrinsic noise explicitly. We can use the Poisson mixture result to simplify analysis and explain properties of models in the same way as intrinsic and extrinsic noise components have (Elowitz *et al.* 2002; Volfson *et al.* 2006; Taniguchi *et al.* 2010). The problem of solving and analysing master equation models of gene transcription is reduced to the problem of obtaining and studying only the variation that is model-dependent, given by the mixing distribution f_{X_t}. This ability to discard the "Poisson part" of the model until we need it confers to us significant simplifications and advantages compared to working with the full distribution $P(n,t)$, not least because X_t is a continuous random variable. In Chapter 4 we collect together some tools and methods available to us for obtaining f_{X_t} analytically or numerically, and note how the cost of stochastic simulations can be significantly reduced with the knowledge of Eq. (3.11). Similarly, obtaining moments and expressions for the noise become simple corollaries of the Poisson mixture result, as we will show in Chapter 5. Less obvious advantages become clear in Part III, where studying f_{X_t} instead of $P(n,t)$ gives us a far deeper understanding of the models and their properties.

Let us begin.

Chapter 4

On the mixture density f_{X_t}

We showed in Chapter 3 that for transcription-degradation models the distribution $P(n,t)$ for the number of mRNA molecules at time t is given by the mixture density

$$P(n,t) = \int \frac{\xi^n}{n!}\, e^{-\xi}\, f_{X_t}(\xi,t)\, d\xi.$$

The challenge of obtaining $P(n,t)$ is thus reduced to the task of obtaining the mixing density f_{X_t}, which contains all the model-specific properties of the solution.

We start this chapter by giving some intuition for how sample paths of the stochastic process $\mathcal{X}$ relate to sample paths for $\mathcal{N}$, and how they relate to differential rate equations used in ordinary differential equation (ODE) models of gene expression. We then collect together more general results from several fields that can aid us to determine f_{X_t}.

4.1 Physical interpretations and intuition

Before we look at methods for obtaining f_{X_t}, let us first try to gain some intuition via some simple examples that:

- For a particular cell with extrinsic parameter $\{x_\omega(t)\}_{t\geq 0}$, $x_\omega(t)$ is *a priori* the expected value of $\eta_\omega(t)$, the sample path for the mRNA copy number for that cell at time t;
- the more synchronous or correlated the drives are in the population, the less dispersed the distribution f_{X_t} will be (and hence the less dispersed $P(n,t)$ will be);
- given the mean value $\mathbb{E}(N_t) = x$ for a population at time t, the Poisson distribution with parameter x is the least dispersed possible distribution for $P(n,t)$.

4.1.1 Static cell-to-cell correlations

Let's consider a gene with sinusoidal transcription rate. This type of behaviour has been observed for genes linked to circadian or cell cycles (Lück and Westermark 2015; Bieler *et al.* 2014), genes responding to cAMP signalling (Corrigan and Chubb 2014), and externally-driven gene expression (Larson *et al.* 2013; Olson *et al.* 2014), for example. As we did in Chapter 3, we'll first consider the case where the transcription and degradation rates are perfectly correlated for all cells, and then see how variability in the rates affects the extrinsic parameters and the distribution $P(n,t)$.

Without extrinsic variation

Let us assume that the transcription rate is perfectly synchronous between all cells in the population: every cell has transcription rate (Lück *et al.* 2014)

$$\mu(t) \coloneqq m[\cos(\theta t) + 1]/2,$$

say. We can then easily calculate the extrinsic parameter $x : t \mapsto x(t)$ using its definition (Eq. (3.5)), which in this hypothetical scenario is deterministic. With $\lambda \equiv 1$ we have:

$$x(t) \coloneqq \int_0^t \mu(\tau) e^{-(t-\tau)}\, d\tau.$$

Since $X_t = x(t)\ \forall\, t$, the density of X_t is given by

$$f_{X_t}(\xi, t) = \delta(\xi - x(t)),$$

where δ is the Dirac delta function (Fig. 4.1, left column).

Despite knowing the drives precisely, transcription and degradation are still stochastic processes so each cell i in the population will have a different sample path η_i for the mRNA copy number. We showed in Section 3.2 that in this special case with synchronous drives, at all times t, $N_t = \{\eta_i(t)\}$ is distributed according to a Poisson distribution with mean $x(t)$. In other words, if we took a snapshot of the population at time t and recorded the distribution of the copy numbers, we would find that

$$P(n, t) = P(n, t|\mu, \lambda) = \frac{x^n(t)}{n!} e^{-x(t)}.$$

Exemplary sample paths of $\mathcal{N}$ with the resulting distribution $P(n, t) = P(n, t|\mu, \lambda)$ are shown in Fig. 4.1, left column.

With extrinsic variation

Now suppose we extend the model to account for random cell-to-cell variation by assuming that the transcription rates within the population are not completely synchronized. We allow the phase Φ to be a random variable with some probability density that we may choose, i.e.,

$$\mu_i(t) = m[\cos(\theta t + \phi_i) + 1]/2.$$

If we now pick a cell i at random from the population at time t, we do not know precisely what the value of its extrinsic parameter $x_i(t)$ will be; it is an outcome from a random variable X_t that has non-zero variance. The density f_{X_t}, which we derive in Example 4.2–1, will depend on the density of Φ. In turn, the distribution of N_t will be given by the Poisson mixture

$$P(n, t) = \int_0^m \frac{\xi^n}{n!} e^{-\xi} f_{X_t}(\xi, t)\, d\xi,$$

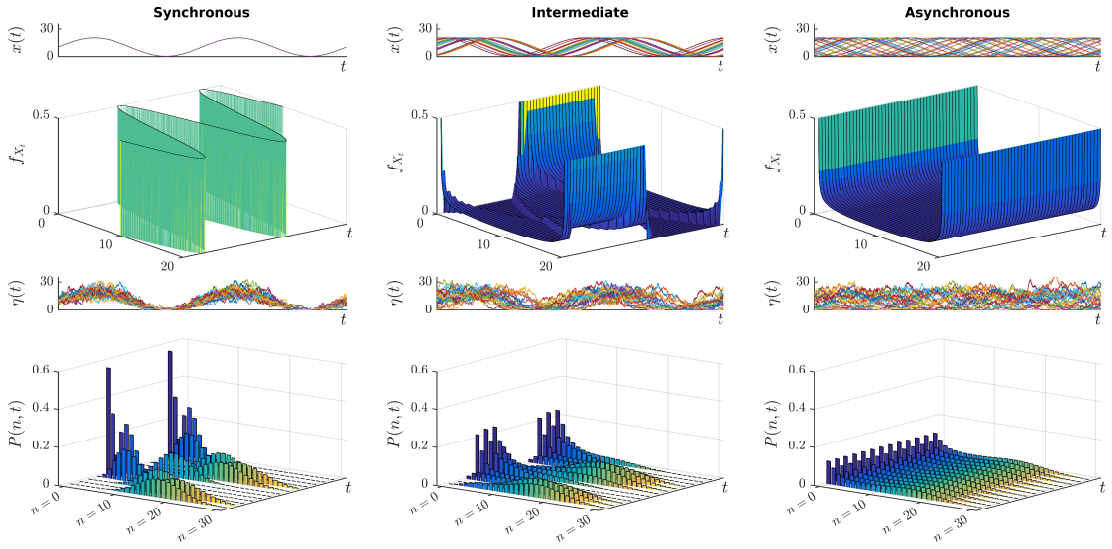

Figure 4.1: The effects of synchrony on a toy model with sinusoidal transcription rate, period T. Derivation of f_{X_t} is treated in Example 4.2–1. *Left column:* Synchronous drive, all heterogeneity is due to the Poisson processes of transcription and degradation. The graph for f_{X_t} for this synchronous case is a representation of a Dirac delta distribution with mass at $x(t)$ at time t. The snapshot distributions shown in the graph for $P(n,t)$ are Poisson with mean $x(t)$. *Middle and right columns:* The phase is uniformly distributed on an interval of range $T/2$ (middle), and T (right). Sample paths of the cellular drive $\mathcal{X}$ (top row), corresponding distribution f_{X_t} (second row), sample paths of the number of mRNA molecules $\mathcal{N}$ (third row), and the exact solution $P(n,t)$ (bottom). Throughout we have taken $m = 20$.

which will necessarily be wider than a Poisson distribution to account for our uncertainty in the extrinsic parameter. In Fig. 4.1 we show some sample paths $x_i(t)$ and $\eta_i(t)$ along with f_{X_t} and $P(n,t)$, when Φ is uniformly distributed on intervals of range $T/2$ and T, where T is the period. It is clear to see how decreasing the degree of synchrony of the cellular drives in the population increases the width of f_{X_t} and $P(n,t)$. When the drives are completely asynchronous, i.e. $\Phi \sim \text{Unif}[-T/2, T/2]$, the model shows no time dependence: we have $P(n,t) = P(n)\ \forall\, t$.

4.1.2 Dynamic cell-to-cell correlations

To give a better idea of how μ, λ, and x are related, let's consider an example with a discrete transcription rate where the degree of synchrony (cell-to-cell correlation) fluctuates over time. Suppose that for each cell i and at all times t we are able to keep track of the mRNA copy number $\eta_i(t)$ during the experiment. Assume also for the purposes of this subsection that we are able to precisely control the transcription rate over time for single cells, for example using light activation (Larson *et al.* 2013; Olson *et al.* 2014). For all times t when the light is on (resp. off), the transcription rate for all cells under the light source is $\mu(t) = m$ min^{-1} (resp. $\mu(t) = 0$ min^{-1}), say. At all times, the transcription rate is $\lambda(t) = 1$ min^{-1} per mRNA molecule.

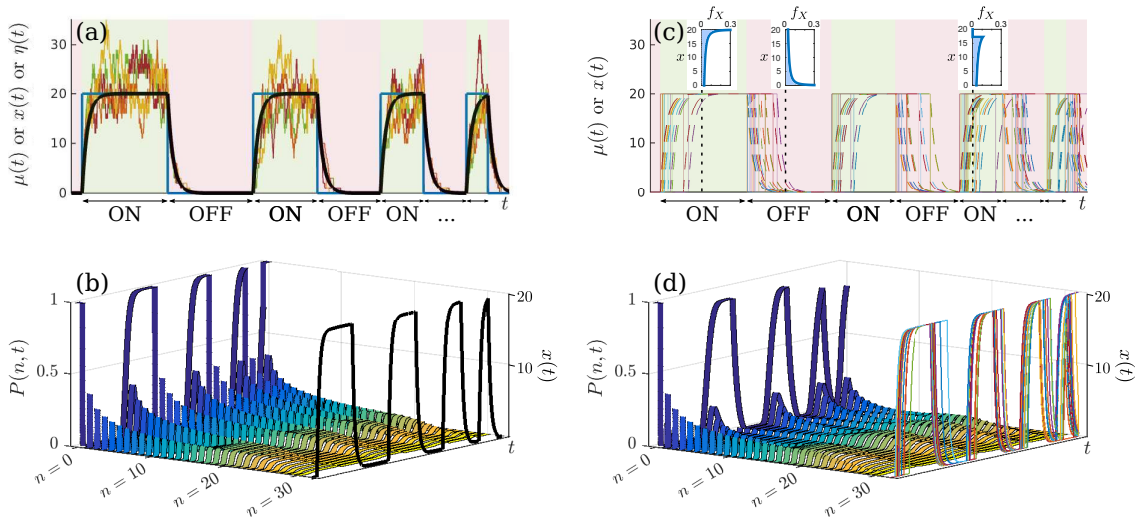

Figure 4.2: Sample paths and probability distributions for a model with temporal control of the transcription rate. *(a, b)* Immediate response of the transcription rate to the external control mechanism (light sequence) in all cells. *(a)* Time series of the light sequence, the transcription rate μ (blue) and the corresponding extrinsic parameter x (black) that apply to every cell in the population. Hence $f_{X_t}(\xi, t) = \delta(\xi - x(t))$ (representation omitted). Sample paths for the number of mRNA transcripts $\mathcal{N}$ are shown by the coloured lines. *(b)* Corresponding distribution $P(n, t|\mu, \lambda \equiv 1)$ over time, with x shown in black as a guide. A cross-section of the figure at any time t_i would show the probability mass function of a Poisson distribution with mean $x(t_i)$. *(c, d)* Response times to changes in the external control mechanism (light sequence) are exponentially distributed. *(c)* Sample paths of the transcription rate (solid lines) and the corresponding extrinsic parameter (dashed lines). Sample paths of $\mathcal{N}$ have been omitted to maintain clarity. *Inset:* The density $f_{X_{t_i}}$ for times t_i indicated by the vertical dotted lines, to give an idea of its dynamic behaviour over time. *(d)* Corresponding distribution $P(n, t)$ over time, with sample paths of $\mathcal{X}$ shown as a guide. Cross-sections are now Poisson mixture distributions. Throughout we have taken $m = 20$.

Without extrinsic variation

Suppose that we have pre-programmed the light sequence so we have a population with perfectly synchronous drives $\mu : t \mapsto \mu(t) \in \{0, m\}$ and $\lambda : t \mapsto 1$ that we know precisely. As always, each cell i in the population will have a different sample path η_i for the mRNA copy number. The trajectories $\eta_i : t \mapsto \eta_i(t)$ "follow" the extrinsic parameter $x : t \mapsto x(t)$, fluctuating around it. As before, the distribution of the η_i is a Poisson distribution with mean $x(t)$, since the mixing density is $f_{X_t}(\xi, t) = \delta(\xi - x(t))$. This will always be the case for a population with perfectly synchronous drives, but in this example it is particularly clear to see when looking at sample paths (Fig. 4.2a). Notice also that although the transcription rate μ is discrete, x is a continuous function of time.

With extrinsic variation

Now suppose that when the light is turned on or off, there is a stochastic delay before the transcription rate responds; the switch is no longer synchronised. For each cell i, the

sample path μ_i is slightly different, particularly in the period immediately following the change in the light, and it follows that the same will be the case for the sample paths x_i. Now, if we pick a cell i at random from the population at time t, we do not know precisely what the value of its extrinsic parameter $x_i(t)$ will be; it is an outcome from a random variable X_t that now has non-zero variance. The density f_{X_t} will depend on the density of the response times to the light switching, and will be time-dependent (see the insets of Fig. 4.2c). In turn, the distribution of N_t will again be given by the Poisson mixture, which will be more dispersed than a Poisson distribution. However, this time the dispersion will fluctuate according to the dispersion of f_{X_t} (see Fig. 4.2c,d). f_{X_t} will be most dynamic immediately after a change in the light sequence, since that is when the distribution of the random variable M_t for the transcription rates will be evolving. Given a long enough waiting time between changes of the light sequence, the sample paths x_i would become almost indistinguishable, with values approaching m when the light is on, and zero when the light is off. As time increases without a switch in the light sequence, f_{X_t} would tend towards a delta distribution with mass at m or zero, and hence $P(n,t)$ would tend towards a Poisson distribution with mean m or zero.

4.1.3 Relationship to ODE models of gene transcription

In chemical reaction kinetics it is common to approximate the number of molecules N_t by a continuous variable, or equivalently, to work in terms of its concentration. Given the reaction rates, we can then write down the *differential rate law*: an ordinary differential equation that describes how the concentration varies in time. Here we will show how the differential rate laws for gene transcription models of the form (3.1) are related to the mixing density f_{X_t}.

Denote the concentration of mRNA molecules in our cell population at time t by $x(t)$. Suppose we know that the transcription rate is $\mu(t)$ and the degradation rate is $\lambda(t)$. Under the assumptions laid out at the beginning of Chapter 3, transcription is a zero order reaction and degradation is a first order reaction. Hence the differential rate law is

$$\frac{dx}{dt} = \mu(t) - \lambda(t)x(t). \tag{4.1}$$

Importantly, this is precisely the equation we obtain by differentiating our integral definition of $x(t)$ (Eq. (3.5)).

Following the same reasoning as in Chapter 3 and the previous subsection, we can extend the use of the differential rate equation further to account for cell-to-cell variability or uncertainty in the reaction rates, or stochastic rates. Any sample paths $t \mapsto \mu_\omega(t)$ and $t \mapsto \lambda_\omega(t)$ satisfy the differential rate equation (4.1), in turn defining the sample path $t \mapsto x_\omega(t)$. For each t, the collection of all possible values $\{x_\omega(t) : \omega \in \Omega_\mathcal{X}\}$ defines our random variable X_t, and the collection of all such ordinary differential equations can be

represented formally by the random[1] differential equation

$$\frac{dX_t}{dt} = M_t - L_t X_t. \tag{4.2}$$

It is now clear that X_t is described by a linear first-order differential equation, where the coefficients may be random. Equations of this form describe a large variety of systems in many sciences, and as such there is a large body of classical results for determining the probability density function f_{X_t} (Kampen 1992; Soong 1973; Pawula 1967; Denisov *et al.* 2009).

In fact, several recent papers (Raj *et al.* 2006; Paszek 2007; Smiley and Proulx 2010; Lück *et al.* 2014; Liu *et al.* 2015) have reasoned that in the limit of large transcription rates, a continuous approximation for N_t is justified and subsequently used Eq. (4.2) to obtain f_{X_t}. However, in reality f_{X_t} is often qualitatively different to $P(n,t)$ (Liu *et al.* 2015). Using the Poisson mixture result, we now know that they only needed to use their "approximate solution" in the Poisson mixture expression to obtain the exact solution $P(n,t)$ for their models (see Example 4.1–1).

Example 4.1–1 — Random telegraph model, approximate solution (Raj *et al.* 2006)

The *random telegraph model* has been widely used to model gene expression since it was proposed by Ko in 1991 (Ko 1991). It is a simple two-state model that explains two key experimental observations: that gene expression is heterogeneous both across cells in a population, and over time in single cells, and that transcription occurs only during short periods of activity. In the 'active' or 'ON' state, transcription events occur according to a Poisson process with constant rate μ, and in the 'inactive' or 'OFF' state, transcription can not occur. Switching between the two states occurs randomly, with constant rates k_{on} and k_{off}. Degradation of the mRNA molecules occurs independent of the state. We will often refer to the random telegraph model or use it in examples because it is widely used in the literature.

Although the exact generating function for the solution of the random telegraph model had been published in 1995 by Peccoud and Ycart (Peccoud and Ycart 1995), the expression for $P(n)$ was not widely known until Raj's seminal paper in 2006 (Raj *et al.* 2006). The derivation of the full solution was not provided, but the authors derived an approximate solution:

> "An interesting parameter regime to investigate is that of μ being large compared to the other rates in the system. In this case, one can then

[1] We use the term *random* differential equation rather than *stochastic* differential equation because the latter are often associated with white noise, which will not usually be the case for us.

make the approximation that N is a continuous variable, $X^{\dagger}$. With this approximation, the dynamics can be described by

$$\frac{dX}{dt} = -\lambda X + \mu r(t)$$

Here, $r(t) \in \{0, 1\}$ is a random telegraph signal representing the active and inactive states of the gene."

They go on to derive the steady-state probability density function for this approximation, which is a scaled Beta distribution:

$$P(n) \approx f_X(x) = \left(\frac{\mu}{\lambda}\right)^{1-\frac{k_{\text{on}}}{\lambda}-\frac{k_{\text{off}}}{\lambda}} \frac{\Gamma\left(\frac{k_{\text{on}}}{\lambda}+\frac{k_{\text{off}}}{\lambda}\right)}{\Gamma\left(\frac{k_{\text{on}}}{\lambda}\right)\Gamma\left(\frac{k_{\text{off}}}{\lambda}\right)} x^{\frac{k_{\text{on}}}{\lambda}-1}\left(\frac{\mu}{\lambda}-x\right)^{\frac{k_{\text{off}}}{\lambda}-1} \tag{4.3}$$

where k_{on} and k_{off} are the rates of switching on and off, resp.

However, this approximation is precisely the mixture distribution required to write down the *exact* solution as a Poisson mixture. Therefore, using the Poisson mixture result (3.11) the exact solution of the random telegraph model is:

$$\begin{aligned} P(n) &= \int_0^{\mu/\lambda} \frac{\xi^n}{n!} e^{-\xi} f_X(\xi)\, d\xi \\ &= \left(\frac{\mu}{\lambda}\right)^{1-\frac{k_{\text{on}}}{\lambda}-\frac{k_{\text{off}}}{\lambda}} \frac{\Gamma\left(\frac{k_{\text{on}}}{\lambda}+\frac{k_{\text{off}}}{\lambda}\right)}{\Gamma\left(\frac{k_{\text{on}}}{\lambda}\right)\Gamma\left(\frac{k_{\text{off}}}{\lambda}\right)} \int_0^{\mu/\lambda} \frac{\xi^n}{n!} e^{-\xi} \xi^{\frac{k_{\text{on}}}{\lambda}-1}\left(\frac{\mu}{\lambda}-\xi\right)^{\frac{k_{\text{off}}}{\lambda}-1} d\xi \\ &= \frac{\Gamma\left(\frac{k_{\text{on}}}{\lambda}+n\right)}{\Gamma\left(\frac{k_{\text{on}}}{\lambda}\right)} \frac{\Gamma\left(\frac{k_{\text{on}}}{\lambda}+\frac{k_{\text{off}}}{\lambda}\right)}{\Gamma\left(\frac{k_{\text{on}}}{\lambda}+\frac{k_{\text{off}}}{\lambda}+n\right)} \frac{\left(\frac{\mu}{\lambda}\right)^n}{n!} {}_1F_1\left(\frac{k_{\text{on}}}{\lambda}+n, \frac{k_{\text{on}}}{\lambda}+\frac{k_{\text{off}}}{\lambda}+n; -\frac{\mu}{\lambda}\right), \end{aligned} \tag{4.4}$$

where ${}_1F_1$ is the confluent hypergeometric function (Abramowitz and Stegun 1964). The last line is the form of the exact solution that was stated in the supplementary material of (Raj *et al.* 2006), but was not derived. It can be deduced from the Poisson mixture form using a standard integral representation of the confluent hypergeometric function (Abramowitz and Stegun 1964).

[†]Notation changed from the original text to be consistent with the notation used in this thesis.

Note that, unlike the approximation to a continuous variable, we do not need to make any assumptions about the value of any parameters to obtain the exact solution.

4.2 Obtaining f_{X_t} using the differential equation for X_t

Obtaining the distribution of a random variable can be difficult if not impossible, and usually requires one to solve an equation for the density function explicitly. However,

since our random variable X_t satisfies the simple, first-order linear differential equation

$$\frac{dX_t}{dt} + L_t X_t = M_t, \tag{4.5}$$

we can sometimes obtain the density function f_{X_t} more directly. Here we give two examples: first, when the solution X_t of Eq. (4.5) is an invertible function of a random variable which we know the distribution of, and second, when we do not require (or can not otherwise obtain) an explicit expression for f_{X_t} and need to resort to cost-effective simulations.

4.2.1 Obtaining f_{X_t} via a change of variables

If we solve the formal differential equation for X_t directly and find that it can be written in the form

$$X_t = g(\Phi_t),$$

where Φ_t is a random variable whose probability density we know, we can calculate f_{X_t} using standard techniques for transformations of random variables. Suppose the transformation g is "nice enough" that the preimage $g^{-1}(x) = \{\phi : g(\phi) = x\}$ is countable for any x in the range of X_t. We can then obtain the cumulative distribution function F_{X_t} or the probability density function f_{X_t} from first principles. When g is injective, $\phi := g^{-1}(x)$ is unique and the procedure is fairly straightforward:

$$\begin{aligned} F_{X_t}(x,t) &= \Pr(X_t \leq x) \\ &= P(g(\Phi_t) \leq x) \\ &= P(\Phi_t \leq g^{-1}(x)) \\ &= F_{\Phi_t}(g^{-1}(x), t) \\ &= F_{\Phi_t}(\phi, t), \end{aligned}$$

and

$$\begin{aligned} f_{X_t}(x,t) &= \frac{d}{dx} F_{X_t}(x,t) \\ &= f_{\Phi_t}(g^{-1}(x), t) \left| \frac{dg^{-1}(x)}{dx} \right| \\ &= f_{\Phi_t}(\phi, t) \left| \frac{1}{g'(\phi)} \right|. \end{aligned}$$

When g is not injective, the same steps can be taken but extra care is required. The clearest way to demonstrate is with an example.

Example 4.2–1 — Sinusoidal transcription rate with distributed phase

Consider the toy model discussed at the beginning of this chapter, where the transcription rate in each cell is sinusoidal with a distributed phase Φ:

$$M_t := m[\cos(\theta t + \Phi) + 1]/2,$$

and constant degradation rate $\lambda \equiv 1$.

Solving Eq. (4.5) and ignoring the effects of initial conditions for simplicity, we obtain

$$\begin{aligned} X_t &= \frac{m}{2(1+\theta^2)}\left[1 + \theta^2 + \cos(\theta t + \Phi) + \theta \sin(\theta t + \Phi)\right] \\ &= B + A\sin(\theta t + \Phi_a), \end{aligned}$$

where $A := m/2\sqrt{1+\theta^2}$, $B := m/2$ and $\Phi_a := \Phi + \arctan(1/\theta)$. Then,

$$\begin{aligned} F_{X_t}(x,t) &:= P(X_t \le x) \\ &= P\left(\sin(\theta t + \Phi_a) \le \frac{x-B}{A}\right). \end{aligned}$$

Now, the set of intervals

$$\left\{[\phi_1, \phi_2] : \ \forall\, \psi \in [\phi_1, \phi_2]\,,\ \sin(\theta t + \psi) \le \frac{x-B}{A}\right\}$$

is countable, so we may proceed. Suppose for example that Φ_a is uniformly distributed on the interval $[-r, r]$, with $r \le \pi$. It is then easy to show that

$$F_{X_t}(x,t) = \frac{s(t)}{2r}\arcsin\left(\frac{x-B}{A}\right) + \{\text{terms independent of } x\},$$

where $s(t) \in \{0, 1, 2\}$ is the number of solutions of

$$\sin\varphi = \frac{x-B}{A}, \quad \varphi \in (\theta t - r, \theta t + r).$$

Differentiating, we finally obtain

$$\begin{aligned} f_{X_t}(x,t) &= \frac{s(t)}{2r\sqrt{A^2 - (x-B)^2}} \\ &= \frac{s(t)}{2r}\left(\left[\frac{m}{2(1+\theta^2)}\right]^2 - \left[x - \frac{m}{2}\right]^2\right)^{-\frac{1}{2}} \end{aligned}$$

(see Fig. 4.1, second row). Clearly, as r tends to zero, i.e. the phase difference between any two cells tends to zero, $f_{X_t}(x,t)$ tends to the Dirac delta function centred at

$B + A\sin(\theta\, t)$. Substituting back into the Poisson mixture solution in this limit we recover the Poisson solution for the case of a synchronous population, as expected (see Fig. 4.1, left column):

$$\begin{aligned} P(n,t) &= \frac{x(t)^n\, e^{-x(t)}}{n!} \\ &= \frac{1}{n!}\left[B + A\sin(\theta\, t)\right]^n\, e^{-[B+A\sin(\theta\, t)]} \end{aligned}$$

4.2.2 Reduction of simulation costs

An explicit expression for $P(n,t)$ is difficult to obtain for all but a handful of gene transcription models. Even with the Poisson mixture result and the tools to obtain f_{X_t} analytically that are presented in the rest of this chapter, one can not always obtain an exact, practical expression for $P(n,t)$.

Several approximation methods have been proposed to approximate $P(n,t)$ (Kampen 1992; Munsky and Khammash 2006; Thomas *et al.* 2012), but probably the most straightforward, popular approach is to use a stochastic simulation algorithm to generate sample paths of the model in question (Gillespie 1976; Gillespie 1977; Cao *et al.* 2006). The distribution of the sample paths can then be used as an approximation for the theoretical probability distribution $P(n,t)$, with the approximation approaching the theoretical one in the limit of infinite sample paths. Although this approach is simple, can be applied to complex systems, and does not require any equations to be solved, it can be extremely computationally expensive and can fail dramatically when time-dependent rates are used (Voliotis *et al.* 2015).

However, the Poisson mixture result tells us that we need only obtain f_{X_t}, and then perform the mixture. Simulating sample paths of $\mathcal{X}$ instead of $\mathcal{N}$ can drastically decrease simulation costs, as we do not need to consider the stochastic processes of transcription or degradation explicitly. We are then balancing the costs of simulating $\{\eta_\omega(t) : t \geq 0\}$ from $\{x_\omega(t) : t \geq 0\}$, versus performing some numerical integration to obtain $P(n,t)$ from f_{X_t}.

Example 4.2–2 — Stochastic simulations for the multistate "Ladder" model

Consider an extension of the random telegraph model where the gene may switch between three different gene states in a ladder structure, as shown in Fig. 4.3 (Senecal *et al.* 2014). The transcription rate is modelled by a continuous-time Markov chain with state space $\{0, \mu_1, \mu_2\}$, $\mu_1 < \mu_2$, corresponding to the states $\{\mathrm{OFF}, \mathrm{ON}_1, \mathrm{ON}_2\}$, respectively. The degradation rate λ is a constant and independent of gene state.

An exact solution for this model is not known[†], but obtaining an approximation for

f_{X_t} via stochastic simulations is extremely quick because $\mathcal{X}$ is a piecewise-deterministic Markov process (Davis 1984). In other words, the behaviour of the sample paths x_ω is governed by the random jumps between gene states at countable times t_i, but in between those times the value of x_ω evolves deterministically according to the ODE (4.1) (see Fig. 4.3 for an sample path of $\mathcal{X}$). Assuming that the cells are uncoupled, the system is ergodic so we can simulate one sample path $\{x_\omega(t)\}_{t\in[0,T]}$ and compute a time average (see Section 5.2), which will tend to the exact solution as $T \to \infty$ if simulated for example according to the Gillespie algorithm (Gillespie 1977). The stationary solution $P(n)$ can then be obtained by performing the Poisson mixture with the approximation for f_X, via numerically integration.

Without knowledge of the Poisson mixture result and the differential rate equation (4.1) for the sample paths of $\mathcal{X}$, we would need to use the Markov chain for the gene state to simulate a sample path of $\mathcal{N}$. In this case we need to obtain event times for every transcription and degradation event, as well as the gene state switching times, so in general the costs will be much higher than the method described above.

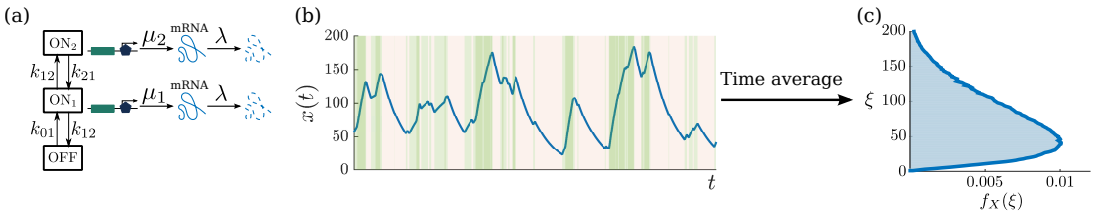

Figure 4.3: Approximating f_X via stochastic simulations. *(a)* Three-state Ladder model of gene transcription. The gene switches between an inactive state (OFF) and two active states (ON_1 and ON_2), with transition rates indicated in the figure. Transcription takes place according to a Poisson process with rate μ_1 in ON_1 and $\mu_2 > \mu_1$ in ON_2. Degradation of transcripts occurs stochastically with rate λ per molecule per unit time. *(b)* Sample path of $\mathcal{X}$, using biologically realistic parameter values (Senecal *et al.* 2014). The gene state is indicated by pink (OFF), light green (ON_1), and dark green (ON_2) shading. *(c)* The resulting stationary probability density f_X.

[†]However, in Chapter 7 we show that f_{X_t} satisfies Heun's equation (Ronveaux 1995).

The cost reduction is particularly noteworthy when we do not have an ergodic[2] system, because millions of simulations are required to obtain enough data for a reliable probability distribution of N_{t_i} at each time point t_i.

4.3 Kramers-Moyal equation for f_{X_t}

Although $\mathcal{X} \equiv \{X_t\}_{t\geq 0}$ satisfies a simple linear equation, when both $\mathcal{M} \equiv \{M_t\}_{t\geq 0}$ and $\mathcal{L} \equiv \{L_t\}_{t\geq 0}$ are stochastic processes expressions can quickly become intractable.

[2]See Section 5.2 for a definition of ergodicity and the implications this has for obtaining solutions.

See Soong 1973 for a readable introduction to the theory and results for these kinds of differential equation. Fortunately, in terms of the transcription and degradation of mRNA molecules it has been shown that most of the variability in $\mathcal{N}$ comes from the dynamics at the promoter (Coulon *et al.* 2010; Suter *et al.* 2011; Carey *et al.* 2013; Sánchez and Golding 2013). Indeed, frequently the degradation probability is approximately constant (Raj *et al.* 2006; Harper *et al.* 2011; Suter *et al.* 2011) and in this case f_{X_t} can often be determined.

Taking $\mathcal{L} \equiv \lambda = \text{constant}$, the differential equation for X_t (Eq. (4.2)) becomes an ordinary differential equation with a random inhomogeneous part, which can be considered as the output of a dynamical system governed by a simple first-order ordinary differential equation with random drive M_t. Equations of this form have been studied since 1908 when Paul Langevin investigated Brownian motion, and are still of paramount importance for applications in fields as wide as electrical engineering, neuroscience, civil engineering, communication theory, economics and finance, and biology (Langevin 1908; Zon *et al.* 2004; Fitzhugh 1983; Crandall and Zhu 1983; Rolski *et al.* 2009). As such, there is a large body of results relating to ordinary differential equations with a random drive. Importantly, the following classical results (Kramers 1940; Moyal 1949) provide us with a differential equation for f_{X_t}, and some general properties. We demonstrate a brief derivation of the differential equation for completeness.

Let f_{X_t} be the probability density function of the random variable X_t that satisfies

$$\frac{dX_t}{dt} + \lambda X_t = M_t, \tag{4.6}$$

where $\{X_t\}_{t\geq 0}$ and $\{M_t\}_{t\geq 0}$ are continuous-time processes, X_t can take a continuum of values, and λ is a constant. Denoting the conditional density for $X_{t+\Delta t} = x$ given that $X_t = x'$ by $f_{X_{t+\Delta t}|X_t}(x, t+\Delta t|x', t)$, the Markov property gives us

$$f_{X_{t+\Delta t}}(x, t+\Delta t) = \int_{-\infty}^{\infty} f_{X_{t+\Delta t}|X_t}(x, t+\Delta t \mid x', t)\, f_{X_t}(x', t)\, dx'. \tag{4.7}$$

Now consider the conditional characteristic function ϕ of the random variable $\Delta X_t \coloneqq X_{t+\Delta t} - X_t$ given that $X_t = x'$:

$$\begin{aligned}\phi(u, t+\Delta t|x', t) &= \mathbb{E}\left(e^{iu\Delta X}\middle| x', t\right) \\ &= \int_{-\infty}^{\infty} e^{iu\Delta x}\, f_{X_{t+\Delta t}|X_t}(x, t+\Delta t|x', t)\, dx,\end{aligned}$$

where $\Delta x \coloneqq x - x'$. Perform the inverse Fourier transform and expand $\phi(u, t+\Delta t|x', t)$ in a Taylor series around $u = 0$:

$$\begin{aligned}f_{X_{t+\Delta t}}(x, t+\Delta t|x', t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iu\Delta x}\, \phi(u, t+\Delta t|x', t)\, du \\ &= \sum_{n=0}^{\infty} \frac{a_n(x', t)}{2\pi n!} \int_{-\infty}^{\infty} (iu)^n\, e^{-iu\Delta x}\, du\end{aligned}$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} a_n(x', t) \frac{d^n}{dx^n} \delta(\Delta x), \tag{4.8}$$

where

$$a_n(x', t) := \frac{1}{i^n} \frac{\partial^n \phi}{\partial u^n}\bigg|_{u=0} = \mathbb{E}\left([\Delta X_t]^n | x', t\right).$$

Substituting Eq. (4.8) into Eq. (4.7) and performing the integration by parts, we obtain

$$\begin{aligned} f_{X_{t+\Delta t}}(x, t+\Delta t) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left[a_n(x,t)\, f_{X_t}(x,t)\right] \\ &= f_{X_t}(x,t) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left[a_n(x,t)\, f_{X_t}(x,t)\right]. \end{aligned}$$

Finally, putting $f_{X_t}(x,t)$ on the left hand side, dividing by Δt and taking the limit $\Delta t \to 0$ gives the *Kramers-Moyal equation*:

$$\frac{\partial f_{X_t}(x,t)}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} [\alpha_n(x,t) f_{X_t}(x,t)], \tag{4.9}$$

where

$$\alpha_n(x,t) := \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{E}\left([X_{t+\Delta t} - X_t]^n \mid X_t = x\right), \quad n = 1, 2, \ldots \tag{4.10}$$

We will refer to the α_n as the *jump moments*, although they are sometimes known as the Kramers-Moyal moments, or derivate moments. Similarly, the Kramers-Moyal equation is sometimes referred to as the kinetic equation associated with the stochastic process $\mathcal{X}^\dagger$ (Soong 1973).

4.3.1 The Fokker-Planck equation can only be an approximation

The Kramers-Moyal equation (4.9) is, *a priori*, a partial differential equation of infinite order. However, Pawula (Pawula 1967) showed that in fact it is either of infinite order, or it is of order at most two:

$$\begin{gathered} \text{If } \alpha_n < \infty \ \forall\ n \text{ and if } \alpha_n = 0 \text{ for some even } n, \text{ then} \\ \alpha_n = 0 \ \forall\ n \geq 3. \end{gathered} \tag{4.11}$$

In the finite case, Eq. (4.9) becomes the well-known Fokker-Planck equation, sometimes called the Kolmogorov forward equation:

$$\frac{\partial f_{X_t}}{\partial t} = -\frac{\partial}{\partial x}\left(\alpha_1 f_{X_t}\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\alpha_2 f_{X_t}\right).$$

Clearly, when solving a gene transcription model one would rather be in the position where f_{X_t} satisfies the Fokker-Planck equation, instead of the infinite order Kramers-

[†]Note that the Kramers-Moyal equation (4.9) is (in general) a partial differential equation of infinite order, in one dimension. It is thus different from the multidimensional Fokker-Planck equation, which is a second-order partial differential equation in several dimensions.

Moyal equation. An interesting question, then, is to ask for what kind of process $\mathcal{X}$ can we obtain the full statistical description by solving only the Fokker-Planck equation? Physically, $\alpha_n = 0$ for $n \geq 3$ implies that $\mathcal{X}$ can only change by a small amount in a small time interval: $\mathbb{E}\left([X(t+\Delta t) - X(t)]^n\right)$ approaches zero faster than Δt as $\Delta t \to 0$. To our knowledge, no one has yet been able to fully characterise all stochastic processes $\mathcal{X}$ that satisfy this condition. However, a sufficient constraint is that $\mathcal{X}$ be a continuous Markov process - for a proof see for example (Gillespie 1992b; Risken 1989).

Now, if $\mathcal{M}$ has stationary, independent increments, $\mathcal{X}$ is a Markov process. However, since $\mathcal{M}$ does not depend on $\mathcal{X}$ in our framework, the only example where $\mathcal{X}$ is a continuous Markov process is the Ornstein-Uhlenbeck process, where $\mathcal{M}$ is a white Gaussian process. This assumption implies that there is a non-zero probability that $\mathcal{X}$ will become negative, and hence it can not be part of a physical model of gene transcription. In other words, *no physical model of gene transcription will enjoy having a stochastic process* $\mathcal{X}$ *whose statistical properties are fully described by the Fokker-Planck equation.* Nevertheless, since the Ornstein-Uhlenbeck process is the *only* example where the Fokker-Planck equation would result in an exact solution, we include it to give an intuition as to how $\mathcal{X}$ fits in with more well-known frameworks from the physics literature.

Example 4.3–1 — When $\mathcal{X}$ is a continuous Markov process

Suppose the degradation rate $\mathcal{L} \equiv \lambda$ is a constant, and the transcription rate $\mathcal{M}$ is a constant μ with an added noise term, which is normally distributed. Then X_t satisfies the Ornstein-Uhlenbeck equation

$$dX_t = \lambda\left(\frac{\mu}{\lambda} - X_t\right) + \sigma dW_t,$$

where $\sigma > 0$ is the standard deviation of the noise and W_t denotes the Wiener process. f_{X_t} then satisfies the Fokker-Planck equation

$$\frac{\partial f_{X_t}}{\partial t} = -\lambda \frac{\partial}{\partial x}\left[\left(\frac{\mu}{\lambda} - x\right) f_{X_t}\right] + \frac{\sigma^2}{2}\frac{\partial^2 f_{X_t}}{\partial x^2},$$

with solution

$$f_{X_t}(x, t | X_0 = x_0) = \frac{\exp\left\{\frac{-[x - \frac{\mu}{\lambda} - e^{-\lambda t}(x_0 - \frac{\mu}{\lambda})]^2}{\sigma^2(1-e^{-2\lambda t})/\lambda}\right\}}{\sqrt{\pi\sigma^2(1-e^{-2\lambda t})/\lambda}}.$$

Hence

$$X_t \sim \mathrm{N}\left(\frac{\mu}{\lambda} + e^{-\lambda t}\left(x_0 - \frac{\mu}{\lambda}\right), \frac{\sigma^2}{2\lambda}\left(1 - e^{-2\lambda t}\right)\right).$$

One might hope that a higher order truncation of the Kramers-Moyal equation would yield a good approximate solution for f_{X_t}. Unfortunately, a naïve truncation is not a legitimate procedure (Risken 1989; Kampen 1992; Plyukhin 2008) as it can prescribe negative values for f_{X_t} and give a worse approximation than truncation at order two. A system-size expansion as pioneered by van Kampen (Kampen 1992) can yield better approximations (Grima *et al.* 2011).

4.4 Multistate Kramers-Moyal equation

The Kramers-Moyal equation (4.9) in theory provides a means for obtaining f_{X_t}. However, as explained in the previous section this partial differential equation will always be of infinite order for realistic models of gene transcription. In addition, the steady-state Kramers-Moyal equations will be degenerate for certain choices of transcription processes, including the important class of *multistate promoter* models (Pawula 1967), which we will define and discuss in Part III. Fortunately, by extending our state space the Kramers-Moyal equation can be extended to account for multistate models, giving a system of differential equations.

In this section we present the multistate Kramers-Moyal equations, and apply them to the random telegraph model to demonstrate how easily the known solution for this model can be rederived using these results. We defer new results to Part III, where we describe in detail how these systems of equations can be solved for more complex, as-yet unsolved examples, and how they can help us deduce which models are solvable.

Consider a multistate model where the promoter transitions between a set of J discrete states $\{s_j\}_{j=1}^{J}$, each with its own transcription rate $\mathcal{M}_j$ and degradation rate $\mathcal{L}_j$. We need to consider the expanded state space $\{X_t, S_t\}$, where $\{S_t\}_{t\geq 0}$ is a discrete random process that describes the promoter state. Our desired probability density function f_{X_t} is then given by

$$f_{X_t}(x,t) = \sum_j f_{X_t,S_t}(x,s_j,t),$$

with

$$\frac{\partial}{\partial t} f_{X_t,S_t}(x,s_j,t) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left[\alpha_n(x,s_j,t) f_{X_t,S_t}(x,s_j,t)\right] + \sum_{k=1}^{J} \beta_k(x,s_j,t) f_{X_t,S_t}(x,s_k,t), \quad (4.12)$$

where

$$\alpha_n(x,s_j,t) = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \mathbb{E}\left(\left[X_{t+\Delta t} - X_t\right]^n \mid X_t = x, S_t = s_j\right) \quad (4.13)$$

are the jump moments as before, and

$$\beta_k(x,s_j,t) = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \left[\Pr(S_{t+\Delta t} = s_j \mid X_t = x, S_t = s_k, t) - \delta_{jk}\right], \quad (4.14)$$

are the transition rates between states. δ_{jk} denotes the Kronecker δ-function (Pawula 1967; Pawula 1970). We will refer to these as the *multistate Kramers-Moyal equations*. Of course if we only have one state, there are no transition probabilities and the multistate Kramers-Moyal equations (4.12) reduce to the standard Kramers-Moyal equations (4.9) from the previous section.

If the jump moments terminate and $\{S_t\}_{t\geq 0}$ is Markovian, then so too is the joint process $\{X_t, S_t\}_{t\geq 0}$ so Equations (4.12) are of finite order. Any model with constant transition rates, as is most common in the literature, is of this type. Furthermore, unless the model uses white noise, $\alpha_{2j} = 0 \;\forall\; j = 1, \ldots, J$ so the multistate Kramers-Moyal equations reduce to:

$$\begin{aligned}\frac{\partial}{\partial t} f_{X_t,S_t}(x, s_j, t) &= -\frac{\partial}{\partial x}\left[\alpha_1(x, s_j, t) f_{X_t,S_t}(x, s_j, t)\right] + \sum_{k=1}^{J} \beta_{jk}(x, t) f_{X_t,S_t}(x, s_k, t) \\ &= -\frac{\partial}{\partial x}\left[(\mu_j(t) - \lambda_j x) f_{X_t,S_t}(x, s_j, t)\right] + \sum_{k=1}^{J} \beta_{jk}(x, t) f_{X_t,S_t}(x, s_k, t),\end{aligned}$$

for $j = 1, \ldots, J$.

The most simple, well-known multistate promoter model is the random telegraph model, which has two promoter states and constant parameters. As mentioned in Example 4.1–1, f_{X_t} for the random telegraph model was obtained as an approximation to the full solution in the supplementary material of Raj *et al.* 2006. They did this by briefly justifying and solving the multistate Kramers-Moyal equations for this particular model, seemingly without being aware of Pawula's more general results. In the following example we will show how to use the definitions of the jump moments (4.13) to write down these same equations to solve, but for the most general form of the random telegraph model. For completeness, we then solve the simultaneous equations by substitution as shown in (Raj *et al.* 2006), although a more general method that dispenses of the need for substitution by hand is presented in Chapter 6.

Example 4.4–1 — Random telegraph model with leaky transcription and state-dependent degradation rates

Consider a 2-state model where the gene state transitions stochastically between an active state s_{on} and an inactive state s_{off}. Transcription occurs at rate μ_{on} in s_{on}, and inefficiently at rate $\mu_{\text{off}} < \mu_{\text{on}}$ in s_{off}. Although degradation is usually assumed to occur independently of promoter state, it does not complicate the derivations here if we allow the degradation rates λ_{on} and λ_{off} in states s_{on} and s_{off}, resp., to be distinct.

To simplify notation we will write

$$\begin{aligned} f_{X_t}(x,t) &= f_{\text{on}}(x,t) + f_{\text{off}}(x,t), \text{ where} \\ f_{\text{on}}(x,t) &\coloneqq f_{X_t,S_t}(x, s_{\text{on}}, t), \text{ and} \\ f_{\text{off}}(x,t) &\coloneqq f_{X_t,S_t}(x, s_{\text{off}}, t). \end{aligned}$$

In order to write down the multistate Kramers-Moyal equations for f_{on} and f_{off} we need to determine the jump moments α_n for $n \in \mathbb{N}$, and the transition rates β. The latter are simply the transition rates given in the model description:

$$\begin{aligned} \beta_{\text{off,on}} &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\Pr(S_{t+\Delta t} = s_{\text{off}} \mid X_t = x, S_t = s_{\text{on}}) - \delta_{\text{off,on}} \right] \\ &= k_{\text{off}} \\ \beta_{\text{on,on}} &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\Pr(S_{t+\Delta t} = s_{\text{on}} \mid X_t = x, S_t = s_{\text{on}}) - \delta_{\text{on,on}} \right] \\ &= -k_{\text{off}} \end{aligned}$$

and similarly, $\beta_{\text{on,off}} = k_{\text{on}}$ and $\beta_{\text{off,off}} = -k_{\text{on}}$.

To determine the jump moments, first notice that in the derivation of the Kramers-Moyal equation (4.9) the jump moments α_n appear from the Taylor expansion of the conditional characteristic function, so we can write:

$$\alpha_n(x,t) = \frac{1}{i^n} \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left. \frac{\partial^n \phi}{\partial u^n} \right|_{u=0} \tag{4.15}$$

For the multistate Kramers-Moyal equations the notation is slightly different, to account for the expanded state space: we will need the jump moments $\alpha_{n,\text{on}}$ and $\alpha_{n,\text{off}}$. They will satisfy this same equation above, Eq. (4.15), but the conditional characteristic function ϕ will be conditional on the state S_t as well as on X_t.

To find this conditional characteristic function ϕ, write Eq. (4.6) in difference form:

$$\lim_{\Delta t \to 0} \frac{X_{t+\Delta t} - X_t}{\Delta t} = M_t - L_t X_t,$$

i.e.

$$\Delta X_t = M_t \Delta t - L_t X_t \Delta t + o(\Delta t),$$

where $o(\Delta t)$ denotes terms that tend to zero faster than Δt as $\Delta t \to 0$. Remember that in this random telegraph model the transcription rate and the degradation rate jump between two discrete values: $M_t \in \{\mu_{\text{on}}, \mu_{\text{off}}\}$ and $L_t \in \{\lambda_{\text{on}}, \lambda_{\text{off}}\}$. Then we have:

$$\begin{aligned} \phi(u, t+\Delta t | X_t = x, S_t = s_{\text{on}}) &= \mathbb{E}\left(e^{iu\Delta X_t} \middle| X_t = x, S_t = s_{\text{on}} \right) \\ &= \mathbb{E}\left(e^{iu[M_t \Delta t - L_t X_t \Delta t + o(\Delta t)]} \middle| X_t = x, S_t = s_{\text{on}} \right) \end{aligned}$$

$$= e^{iu[\mu_{\text{on}}\Delta t - \lambda_{\text{on}} x\Delta t + o(\Delta t)]},$$

and similarly

$$\phi(u, t+\Delta t | X_t = x, S_t = s_{\text{off}}) = e^{iu[\mu_{\text{off}}\Delta t - \lambda_{\text{off}} x\Delta t + o(\Delta t)]}.$$

Using the definition of the jump moment, Eq. (4.15), we find:

$$\begin{aligned}
\alpha_{1,\text{on}}(x,t) &= \frac{1}{i}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{\partial}{\partial u}\left\{e^{iu[\mu_{\text{on}}\Delta t - \lambda_{\text{on}} x\Delta t + o(\Delta t)]}\right\}\Big|_{u=0} \\
&= \mu_{\text{on}} - \lambda_{\text{on}} x, \text{ and similarly} \\
\alpha_{1,\text{off}}(x,t) &= \mu_{\text{off}} - \lambda_{\text{off}} x; \\
\alpha_{2,\text{on}}(x,t) &= \frac{1}{i^2}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{\partial^2}{\partial u^2}\left\{e^{iu[\mu_{\text{on}}\Delta t - \lambda_{\text{on}} x\Delta t + o(\Delta t)]}\right\}\Big|_{u=0} \\
&= 0, \text{ and similarly} \\
\alpha_{2,\text{off}}(x,t) &= 0.
\end{aligned}$$

Since the second jump moments are zero, we can immediately deduce using Eq. (4.11) that $\alpha_{n,\text{on}} = 0$ and $\alpha_{n,\text{off}} = 0 \;\forall\; n \geq 2$. Note that since the expanded state space $\{X_t, S_t\}_{t\geq 0}$ is Markovian, finite multistate Kramers-Moyal equations were expected anyway.

Putting this all together into the multistate Kramers-Moyal equations (4.12), and setting the time derivatives equal to zero so that we obtain the steady-state solution, we finally obtain

$$\begin{aligned}
\frac{d}{dx}\left[(\mu_{\text{on}} - \lambda_{\text{on}} x) f_{\text{on}}\right] &= -k_{\text{off}} f_{\text{on}} + k_{\text{on}} f_{\text{off}} \\
\frac{d}{dx}\left[(\mu_{\text{off}} - \lambda_{\text{off}} x) f_{\text{off}}\right] &= k_{\text{off}} f_{\text{on}} - k_{\text{on}} f_{\text{off}}
\end{aligned}$$

These simultaneous ordinary differential equations can be solved straightforwardly by substitution, as shown in more detail in (Raj *et al.* 2006) for example. The final solution is then

$$f_{X_t}(x,t) = C(\mu_{\text{on}} - \lambda_{\text{on}} x)^{\frac{k_{\text{off}}}{\lambda_{\text{on}}} - 1}(\lambda_{\text{off}} x - \mu_{\text{off}})^{\frac{k_{\text{on}}}{\lambda_{\text{off}}} - 1}[(\mu_{\text{on}} - \mu_{\text{off}}) - (\lambda_{\text{on}} - \lambda_{\text{off}})x] \quad \forall\, t,$$

where the constant of integration C can be determined with the integral condition

$$\int_{\min\{\mu_{\text{on}}/\lambda_{\text{on}},\, \mu_{\text{off}}/\lambda_{\text{off}}\}}^{\max\{\mu_{\text{on}}/\lambda_{\text{on}},\, \mu_{\text{off}}/\lambda_{\text{off}}\}} f_X(x)\, dx = 1.$$

When $\lambda_{\text{on}} = \lambda_{\text{off}} = \lambda$, we obtain the usual solution for the leaky random telegraph

model (Smiley and Proulx 2010):

$$f_{X_t}(x,t) = \frac{\lambda(\mu_{\text{on}} - \lambda x)^{\frac{k_{\text{off}}}{\lambda}-1}(\lambda x - \mu_{\text{off}})^{\frac{k_{\text{on}}}{\lambda}-1}}{(\mu_1 - \mu_0)^{\frac{k_{\text{on}}+k_{\text{off}}}{\lambda}-1} B\left(\frac{k_{\text{on}}}{\lambda}, \frac{k_{\text{off}}}{\lambda}\right)},$$

where B is the beta function.

4.5 Equation for f_{X_t} using Fourier transforms

The differential equation (4.2) for $\{X_t\}_{t\geq 0}$ can be thought of as a description of the time evolution of a dynamical system that interacts with a randomly fluctuating environment. As such, in the field of statistical physics this equation would be called a Langevin equation (with a non-Gaussian noise term M_t). Obtaining the probability density f_{X_t} of the solution of a Langevin equation is often a primary concern to statistical physicists, so we are able to draw on some of the results derived in that field to help us determine f_{X_t}.

For brevity, here we mention only one fairly recent result derived in Denisov *et al.* 2008; Denisov *et al.* 2009, that provides an equation and some exact solutions for f_{X_t} in certain cases. Note that their results apply to a slightly different general setting than we are constrained to by Eq. (4.5), so the expressions that we include here are only the specific cases of their original results that apply to our system. Additionally, for their results to hold we must assume that *i)* our degradation rate $\mathcal{L} \equiv \lambda$ is constant, and *ii)* M_t can be written as the time derivative of the "noise generating process" Ξ_t, i.e.:

$$\Delta\Xi_t := \Xi_{t+\Delta t} - \Xi_t = \int_t^{t+\Delta t} M_\tau \, d\tau.$$

Then, the increment $\Delta X_t := X_{t+\Delta t} - X_t$ during a time interval of Δt, $\Delta t \to 0$, can be written in the form

$$\Delta X_t = -\lambda X_t \Delta t + \Delta\Xi_t.$$

Write the probability density function of the random variable $\Delta\Xi_t$ as $f_{\Delta\Xi_t}(\xi, \Delta t)$, denote its Fourier transform by $\hat{f}_{\Delta\Xi_t}(k,t)$, and define

$$\phi_k := \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\hat{f}_{\Delta\Xi_t}(k,t) - 1\right],$$

which is a measure of the strength of the influence of M_t on the system. Denoting also by $\hat{f}_{X_t}(k,t)$ the Fourier transform of $f_{X_t}(x,t)$, we have (Denisov *et al.* 2009)

$$\hat{f}_{X_t}(k,t) = \exp\left(-\frac{1}{\lambda}\int_0^k \frac{\phi_{ze^{-\lambda t}} - \phi_z}{z}\, dz\right),$$

or equivalently

$$f_{X_t}(x,t) = \mathcal{F}^{-1}\left\{\exp\left(-\frac{1}{\lambda}\int_0^k \frac{\phi_{ze^{-\lambda t}} - \phi_z}{z}\,dz\right)\right\}, \tag{4.16}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform operator.

Example 4.5–1 — Poisson white noise: in the limit of infinitesimally short burst intervals

Pedraza and Paulsson (Pedraza and Paulsson 2008) assumed that transcriptional bursts occurred on timescales so short that they could be modelled by discrete jumps in the sample paths of $\mathcal{N}$ (see Figure 4.4). They were able to obtain expressions for the mean and variance of N_t in the stationary case, but using the Fourier transform method we can obtain a closed-form expression for $P(n,t)$ for certain waiting time and burst size distributions.

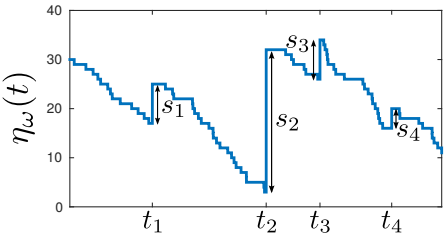

Figure 4.4: Sample path of $\mathcal{N}$ for a model with Poisson white noise: random burst times t_i occur according to a Poisson process with constant rate β, and the burst sizes s_i are drawn from a probability distribution q. Degradation of mRNA transcripts is a first-order process with constant rate.

Suppose $\mathcal{L} \equiv \lambda$ and transcription only occurs during short "burst" intervals, with exponentially distributed waiting times between bursts. Suppose further that each burst at time t increases the expectation of N_t by a random variable S, i.e. S describes the burst size. In the limit of infinitesimally short burst times, the transcription rate can be modelled by a Poisson sequence of Dirac delta pulses:

$$\mu_\omega(t) = \sum_{i=1}^{B_t} s_i\,\delta(t - t_i),$$

where

$B_t \sim \text{Poi}(\beta)$ is a Poisson process that counts the bursts up to time t,

t_i are the burst times of this Poisson process,

s_i are independent outcomes from the random variable S
with probability density q,
$\delta(\cdot)$ is the Dirac delta function.

In this case

$$\phi_k = \beta(q_k - 1),$$

where q_k is the Fourier transform of the probability density of the burst sizes, so substitution into Eq. (4.16) gives

$$f_{X_t}(x,t) = \mathcal{F}^{-1}\left\{\exp\left(-\frac{\beta}{\lambda}\int_0^k \frac{q_{ze^{-\lambda t}} - q_z}{z}\,dz\right)\right\}. \tag{4.17}$$

For certain choices of q we can evaluate Eq. (4.17) to give an explicit solution in terms of x and t only. For example, choosing the exponential distribution we have (Denisov *et al.* 2009)

$$\begin{aligned} q(s) &= \frac{r}{2}e^{-rs}, \quad r,s>0 \\ \text{so } q_k &= \frac{r^2}{r^2+k^2} \\ \text{and } f_{X_t}(x,t) &= \mathcal{F}^{-1}\left\{\exp\left(-\frac{\beta}{\lambda}\int_0^k \frac{r^2(1-e^{-2\lambda t})z}{(r^2+z^2)(r^2+z^2e^{-\lambda t})}\,dz\right)\right\} \\ &= \mathcal{F}^{-1}\left\{\left(\frac{r^2+k^2e^{-2\lambda t}}{r^2+k^2}\right)^{\frac{\beta}{2\lambda}}\right\}. \end{aligned}$$

Letting $t \to \infty$ we can invert the Fourier transform to obtain the stationary distribution:

$$f_X(x) = \sqrt{\frac{2}{\pi}}\frac{r(rx)^{(\beta-\lambda)/2\lambda}}{2^{\beta/2\lambda}\Gamma\left(\frac{\beta}{2\lambda}\right)}K_{(\beta-\lambda)/2\lambda}(rx),$$

where K is the modified Bessel function of the second kind (Abramowitz and Stegun 1964).

4.6 Discussion

As noted at the end of Chapter 3, the mixing density f_{X_t} is the essence of any transcription-degradation model of the kind we consider in this thesis. Before moving on to using the Poisson mixture result to derive solutions and properties of models, this chapter aimed to give some physical intuition for the extrinsic random variable X_t, and introduce some methods one may use to obtain its density f_{X_t}. It is far from a complete exposition of the known approaches to derive or approximate f_{X_t}, which would have taken us away from the scope of this thesis.

In particular, we could have dwelt far longer on the advantages afforded to us by the Poisson mixture result when obtaining numerical solutions. We stated in Section 4.2.2 that approximating f_{X_t} using a stochastic simulation algorithm can be much quicker than approximating $P(n,t)$, but a more detailed investigation would exemplify how significant the savings can be.

Chapter 5

Averaging in space or time

The previous two chapters were concerned with obtaining the complete statistical description $P(n,t) = \Pr(N_t = n)$ of the random process $\mathcal{N} = \{N_t\}_{t\geq 0}$. In general, this solution would give information about cell-to-cell heterogeneity of a population, as well as how that is expected to change over time. However, in certain circumstances such detailed information may not be required or relevant, and simpler characterisations based on averages are of more interest. This chapter uses what we know about the full distribution $P(n,t)$ to derive some corollaries in this direction.

In particular, a fundamental question in the study of gene expression concerns the nature and origins of the heterogeneity due to stochastic processes, or "noise", that is observed so ubiquitously in this field. Noise is characterised using different metrics according to the type of data or model being analysed, but many of the most simple, intuitive metrics use combinations of moments. These include the two that are most commonly used for gene expression data: the *coefficient of variation*, η, (or the squared coefficient of variation, η^2) (Elowitz *et al.* 2002; Swain *et al.* 2002; Paulsson 2005):

$$\eta := \frac{\text{standard deviation}}{\text{mean}}, \quad \text{or} \quad \eta^2 := \frac{\text{variance}}{\text{mean}^2},$$

and the *Fano factor* (Thattai and Oudenaarden 2001; Golding *et al.* 2005; Munsky *et al.* 2012):

$$\text{Fano factor} := \frac{\text{variance}}{\text{mean}}.$$

It is clear from the definitions that these metrics are very similar, and one is usually chosen over the other based only on how easy they are to interpret in the given context. The coefficient of variation is a dimensionless quantity, so it gives a standardised measure of dispersion. The Fano factor has been popularized more recently by van Oudenaarden (Thattai and Oudenaarden 2001; Raj and Oudenaarden 2009) specifically for the kinds of gene transcription models we consider here; under the assumption that there is no gene regulation, transcription and degradation of mRNA are zero- and first-order stochastic processes with constant rates μ and λ, say, and the stationary distribution of $\mathcal{N}$ is a Poisson distribution with parameter μ/λ. The Fano factor equals one for a Poisson distribution, so it is a useful measure of deviation from the constant-rate model of unregulated gene expression.

Although it is rarely explicitly stated, Fano factor analysis of gene transcription data or models so far have used ensemble moments in the definition (see below). However,

the Fano factor was originally defined to characterise noise in time course data (Fano 1947), using the mean and variance calculated over a time window (temporal moments, see below). We will refer to the Fano factor calculated using ensemble (resp. temporal) moments as the *ensemble* (resp. *temporal*) *Fano factor*, and discuss each separately in the corresponding sections below, as they have very different interpretations and will not give the same value.

Before we continue, let's clarify what we mean by *ensemble* and *temporal moments*: under the general framework outlined in Chapter 3, the distribution $P(n,t)$ has two dimensions across which we may take an average. First, for each fixed t we can consider the *ensemble moments*

$$\mathbb{E}\left(N_t^k\right) := \sum_n n^k \, P(n,t).$$

They correspond to averaging on a snapshot of the population at time t, and will be time-dependent if the statistical properties of the transcription and/or degradation rates are. For example, mean expression levels of circadian genes exhibit approximately periodic behaviour over 24-hour intervals (Bieler *et al.* 2014; Lück and Westermark 2015). Expressions for the ensemble moments and other corollaries will be discussed in Section 5.1.

Second, with single-cell time course data or ergodic models we can consider the *temporal moments*

$$\langle N_t^k \rangle := \frac{1}{T} \int_0^T n^k \, P(n,t) \, dt.$$

As the name suggests, temporal moments can allow us to compare the average expression of single cells over a time interval, such as a cell cycle, duration of an experiment, or after a change at the promoter. In addition, the temporal average can be used to derive an alternative solution of the master equation for ergodic models which is especially useful when we have periodic drives. All these concepts will be discussed in more detail in Section 5.2.

5.1 Ensemble moments

Ensemble moments hold an important place in gene expression analysis. Experimental methods such as northern blotting, microarrays, qPCR, and RNA-Seq are only capable of obtaining the concentration or a measure of the total amount of a substance or molecule in a cell population, equivalent to the mean value of a population snapshot. Even with single cell data, moments are useful quantitative characterisations of the data that can be used for comparisons or expression summaries. On the theoretical side, we showed in Chapter 3 that the solution $P(n,t)$ can always be written in Poisson mixture form, but if f_{X_t} can not be determined or the integration is difficult this description might not be fit for purpose. On the other hand, expressions for the moments are usually easier to obtain and/or work with, and can be used to approximate the full distribution if necessary (Lasserre 2002; Bertsimas and Popescu 2005). In addition, conclusions regarding gene expression noise are usually based on the squared coefficient of variation or ensemble Fano factor (Paulsson

2004; Ozbudak *et al.* 2002; Thattai and Oudenaarden 2001).

Our possession of the integral form of the solution (Eq. (3.11)) enables us to determine the time-dependent moments $\mathbb{E}(N_t^k)$ from $\{\mathbb{E}\left(X_t^r\right)\}_{r=1}^k$, and vice versa. We will first derive these relations, and then use them to deduce properties of the ensemble noise characteristics.

5.1.1 Derivation of the ensemble moments of N_t

In Section 3.2 we showed that N_t conditioned on $X_t = x$ is a Poisson random variable with parameter x (Eq. (3.10)):

$$P(n, t|X_t = x(t)) = \frac{x(t)^n}{n!} e^{-x(t)}.$$

Now, the k^{th} conditional moment of a Poisson random variable is

$$\mathbb{E}\left(N_t^k \middle| X_t = x\right) = \sum_{r=1}^{k} x^r \, S(k, r),$$

where

$$S(k, r) \equiv \left\{ {k \atop r} \right\} := \frac{1}{r!} \sum_{j=1}^{r} (-1)^{r-j} \binom{r}{j} j^k$$

are the Stirling numbers of the second kind (Riordan 1937). The law of total expectation then gives us

$$\begin{aligned} \mathbb{E}\left(N_t^k\right) &= \int \mathbb{E}\left(N_t^k \middle| X_t = x\right) f_{X_t}(x, t)\, dx \\ &= \sum_{r=1}^{k} S(k, r) \int x^r f_{X_t}(x, t)\, dx \\ &= \sum_{r=1}^{k} S(k, r)\, \mathbb{E}\left(X_t^r\right). \end{aligned} \tag{5.1}$$

In particular,

$$\begin{aligned} &\mathbb{E}\left(N_t\right) = \mathbb{E}\left(X_t\right), \quad \text{and} \\ &\mathbb{E}\left(N_t^2\right) = \mathbb{E}\left(X_t^2\right) + \mathbb{E}\left(X_t\right) \Rightarrow \mathrm{Var}(N_t) = \mathrm{Var}(X_t) + \mathbb{E}\left(X_t\right). \end{aligned}$$

Conversely, given the moments $\{\mathbb{E}\left(N_t^r\right)\}_{r=1}^k$ we can use the relationship (5.1) to find $\mathbb{E}(X_t^k)$: Eq. (5.1) can be written in matrix form

$$\begin{pmatrix} \mathbb{E}\left(N_t\right) \\ \mathbb{E}\left(N_t^2\right) \\ \mathbb{E}\left(N_t^3\right) \\ \vdots \\ \mathbb{E}\left(N_t^k\right) \end{pmatrix} = \begin{pmatrix} S(1,1) & 0 & 0 & \dots & 0 \\ S(2,1) & S(2,2) & 0 & \dots & 0 \\ S(3,1) & S(3,2) & S(3,3) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S(k,1) & S(k,2) & S(k,3) & \dots & S(k,k) \end{pmatrix} \begin{pmatrix} \mathbb{E}\left(X_t\right) \\ \mathbb{E}\left(X_t^2\right) \\ \mathbb{E}\left(X_t^3\right) \\ \vdots \\ \mathbb{E}\left(X_t^k\right) \end{pmatrix}.$$

The matrix above is easily invertible (Usmani 1994), but the set of equations can also be solved simply by forward substitution. For example,

$$\begin{aligned}
&\mathbb{E}\left(N_t\right) = S(1,1)\,\mathbb{E}\left(X_t\right) \\
&\Rightarrow \mathbb{E}\left(X_t\right) = \frac{\mathbb{E}\left(N_t\right)}{S(1,1)} = \mathbb{E}\left(N_t\right), \\
&\mathbb{E}(N_t^2) = S(2,1)\,\mathbb{E}\left(X_t\right) + S(2,2)\,\mathbb{E}(X_t^2) = S(2,1)\,\mathbb{E}\left(N_t\right) + S(2,2)\,\mathbb{E}(X_t^2) \\
&\Rightarrow \mathbb{E}(X_t^2) = \frac{\mathbb{E}\left(N_t^2\right) - S(2,1)\,\mathbb{E}\left(N_t\right)}{S(2,2)} = \mathbb{E}(N_t^2) - \mathbb{E}\left(N_t\right), \\
&\text{etc.}
\end{aligned}$$

See Fig. 5.1 for an illustration of the ensemble moments of X_t and N_t as they vary in time, for a model with promoter strengths coupled via a simple Kuramoto model.

> **Remark: Parameter estimation via the method of moments**
> The above shows that given the moments $\{\mathbb{E}\left(N_t^r\right)\}_{r=1}^k$, we can easily obtain $\{\mathbb{E}\left(X_t^r\right)\}_{r=1}^k$. This correspondence makes parameter estimation using the method of moments straightforward: suppose we have a model with k unknown parameters, and mRNA count data for a population at times t_i. We can estimate $\{\mathbb{E}(N_{t_i}^r)\}_{r=1}^k$ using the data, and thus find the corresponding estimates for $\{\mathbb{E}(X_{t_i}^r)\}_{r=1}^k$. These will give us k equations for the k unknown parameters to be found at each time t_i.

5.1.2 Ensemble Fano factor

Recall that the ensemble Fano factor, $\mathrm{Fano}(N_t)$, is defined by:

$$\mathrm{Fano}(N_t) := \frac{\mathrm{Var}(N_t)}{\mathbb{E}\left(N_t\right)}.$$

Since it is calculated on population snapshots, the statistics of the population may change over time whilst maintaining a constant value of $\mathrm{Fano}(N_t)$. For example, by Eq. (3.10) snapshots of a population with perfectly synchronous transcription and degradation rates has a Poisson distribution at any given time t, and thus has ensemble Fano factor equal to one. The Poisson parameter $\{x(t)\}_{t\geq 0}$ may change in time, but the degree of synchrony remains constant and therefore so does the Fano factor (Figs. 4.1 and 5.1).

In fact, $\mathrm{Fano}(N_t)$ can be thought of as a measure of synchrony in a population at time t. Using Eq. (5.1), the theoretical value for the Fano factor of the population at time t is

$$\begin{aligned}
\mathrm{Fano}(N_t) &:= \frac{\mathrm{Var}(N_t)}{\mathbb{E}\left(N_t\right)} \\
&= \frac{\mathbb{E}\left(X_t\right) + \mathrm{Var}(X_t)}{\mathbb{E}\left(X_t\right)} \\
&= 1 + \mathrm{Fano}(X_t).
\end{aligned}$$

Here we see why the ensemble Fano factor is highly suited to the analysis of transcription-degradation models and systems, and how it can be used as a measure of synchrony; all deviation from the Poisson distribution is contained in the Fano factor of the mixing density f_{X_t}. The more correlated or "synchronised" the cellular drives are, the closer the Fano factor of N_t will be to one. In particular, the Fano factor will be greater than or equal to one, with equality only when the Fano factor of X is zero, i.e. when the transcription and degradation rate functions are perfectly synchronous for all cells in the population.

Example 5.1–1 — Kuramoto model

The Kuramoto model is a set of coupled oscillators whose coupling strength can be tuned and phase coherence measured, which has found numerous applications in biological, chemical, and some physical systems (Kuramoto 1975; Strogatz 2000). We will use it to illustrate how the spatial characteristics of a population vary in time and according to the degree of synchrony of the cellular drives.

Consider a set of 100 coupled oscillators $\{\theta_i(t)\}_{i=1}^{100}$ under the Kuramoto model with coupling strength K (Kuramoto 1975; Strogatz 2000). The order parameter $r_t \in [0, 1]$ measures the phase coherence of the oscillators at time t; the closer r_t is to 1 the greater the degree of synchrony at that point in time.

Assuming transcription rate $\mu_i(t) := 8\cos\theta_i(t) + 16$ and degradation rate $\lambda \equiv 1$, we determined $x_i(t)$ and simulated $\eta_i(t)$ over a range of K (Fig. 5.1, rows 1-2). Using both sets of this *in silico* data, we calculated the ensemble mean, variance, and Fano factor for N_t at every time point t (Fig. 5.1, rows 3-6). For $K = 0.002$, the oscillators are so weakly coupled the phase coherence r_t is close to zero and the system is effectively stationary. Hence all ensemble statistics are approximately constant over time, and the Fano factor is large. On the other end of the scale, for $K = 0.4$ the oscillators are coupled strongly enough to maintain phase coherence r_t extremely close to one; the drives are almost in perfect synchrony over all times. Notice how this ensures that the Fano factor remains approximately constant around the value of one, even though the mean and variance vary in time. $K = 0.1$ takes the oscillators between the two extremes, where there is some degree of order but the oscillations occur on different timescales or with phase differences. Here, the Fano factor shows some periodic behaviour, indicating that the degree of synchrony of the population varies over time.

The order parameter r_t measures the phase coherence of the oscillators, so if the ensemble Fano factor is an effective measure of the degree of synchrony in a population over time, there should be a strong correlation between r_t and $\mathrm{Fano}(N_t)$. For each value of K we plot in Fig. 5.1b the time-average of the Fano factor, $\langle \mathrm{Fano}(N_t) \rangle$, against the time-average of the order parameter, $\langle r_t \rangle$. The negative correlation is clear: the greater

the degree of synchrony, the closer the Fano factor is to 1. Similarly, there is a clear relationship between the coupling parameter K and Fano(N_t) (Fig. 5.1b inset), that mirrors the well-known relationship between r_t and K (Strogatz 2000).

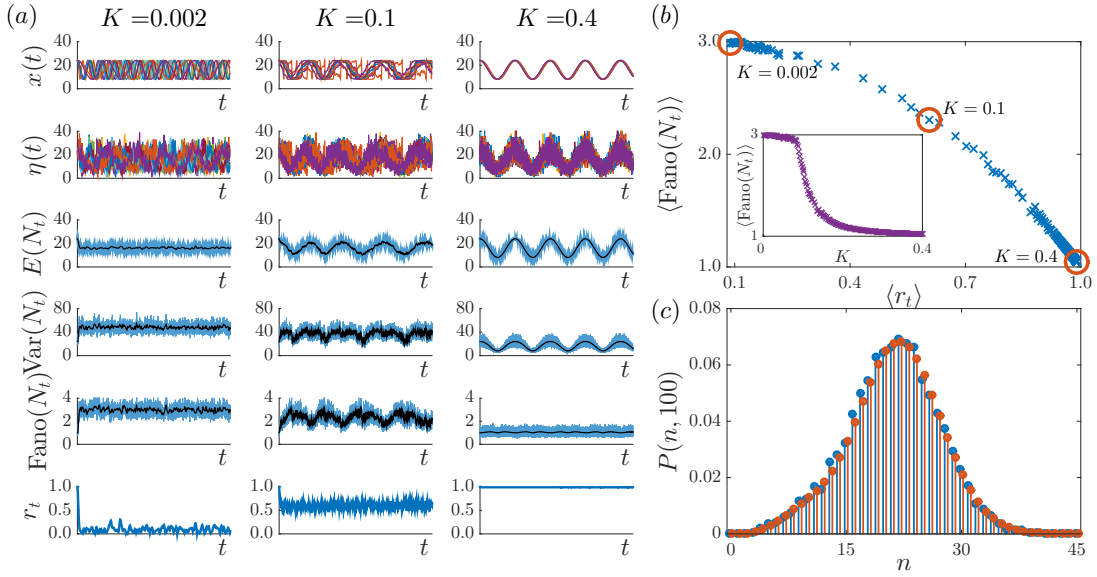

Figure 5.1: Spatial statistics and comparison of results from simulations of $\mathcal{X}$ *vs* $\mathcal{N}$ for the Kuramoto model. *(a)* Using the numerical solution of the Kuramoto model for 100 oscillators $\{\theta_i(t)\}_{i=1}^{100}$ with coupling strength $K = 0.002$ (left column), $K = 0.1$ (middle column), $K = 0.4$ (right column), the transcription rates for each oscillator are defined as $\{\mu_i(t)\}_{i=1}^{100} = 8\cos\theta_i(t) + 16$. The degradation rate is $\lambda \equiv 1$. *Rows 1-2:* Sample paths of $\mathcal{X}$ and $\mathcal{N}$, respectively. *Rows 3-5:* Mean, variance, and ensemble Fano factor calculated from the sample paths of $\mathcal{N}$ (blue lines), and the corresponding estimations calculated from the sample paths of $\mathcal{X}$ (black dashed lines). *Row 6:* The order parameter r_t that measures the phase coherence of the oscillators. *(b)* Scatter plot of the time-averaged Fano factors against the time-averaged order parameters for coupling parameter values between $K = 0.002$ and $K = 0.4$. *(c)* $P(n, t)$ at fixed time $t = 100$ for a system of 10 000 oscillators, using simulated sample paths for $\mathcal{N}$ (blue) and $\mathcal{X}$ (red).

5.1.3 Squared coefficient of variation

The coefficient of variation η, or more usually the squared coefficient of variation η^2, is a popular measure of noise for models of gene expression, particularly those considering proteins as well as mRNA transcripts.

As a dimensionless quantity η^2 is useful for comparisons across species, and in an influential paper by Swain *et al.* 2002 it was shown that it can be decomposed into a sum of its "intrinsic" and "extrinsic" components η^2_{int} and η^2_{int}, from which we can try to deduce the sources of the noise. They defined intrinsic noise as variation generated by the "inherent stochasticity of biochemical processes such as transcription and translation",

and extrinsic noise as "fluctuations in the amounts or states of other cellular components". This additive property is naturally recovered when we use the expressions for the ensemble moments (Eq. (5.1)) to write down an expression for η^2, as we did for the ensemble Fano factor. We have:

$$\begin{aligned}\eta^2 &:= \frac{\mathrm{Var}(N_t)}{\mathbb{E}\left(N_t\right)^2} \\ &= \frac{\mathbb{E}\left(X_t\right) + \mathrm{Var}(X_t)}{\mathbb{E}\left(X_t\right)^2} \\ &= \frac{1}{\mathbb{E}\left(X_t\right)} + \eta_X^2 \\ &=: \eta_{\mathrm{int}}^2 + \eta_{\mathrm{ext}}^2.\end{aligned}$$

As could be expected, the intrinsic noise for gene transcription models is simply the squared coefficient of variation for a Poisson random variable. The extrinsic noise can now be more precisely defined as the squared coefficient of variation for the random variable X_t.

5.2 Stationarity and ergodicity

The results presented so far have not placed many restrictions on the stochastic processes for transcription $\mathcal{M}$ and degradation $\mathcal{L}$, so in general the distributions f_{X_t} and $P(n,t)$ depended on time. However, many gene transcription models assume (at least implicitly) that there is no coupling or correlations between cells in the population, and that the statistics of the stochastic processes $\mathcal{M}$ and $\mathcal{L}$ do not change over time. In these cases, the system will be *ergodic* and/or *stationary*, so f_{X_t} and $P(n,t)$ will tend to equilibrium distributions f_X and $P(n)$ that are independent of time.

In this section we discuss some corollaries of the Poisson mixture result that are related to these important properties: in Section 5.2.1 we give an alternative expression for $P(n)$ under the assumption of ergodicity, which is especially useful when considering periodic drives. In Section 5.2.2 we discuss the temporal Fano factor in the context of stationarity and use a simple example to illustrate shifts between stationary and non-stationary intervals in time course *in silico* data. Before proceeding, however, let us recall the definitions of stationarity and ergodicity, the difference between them, and what they each imply in the context of gene transcription models.

A *stationary* process, or a system at *stationarity*, means that its joint probability distribution does not change when shifted in time: $\forall\, k,\ \forall\, \tau$, and $\forall\, t_1, \ldots, t_k$,

$$\Pr(N_{t_1} = n_1, \ldots, N_{t_k} = n_k) = \Pr(N_{t_1+\tau} = n_1, \ldots, N_{t_k+\tau} = n_k),$$

and similarly for the mixing density f_{X_t}. Simply speaking, and for our purposes, this means that the mRNA distribution and mixing density do not depend on time:

$$P(n, t_i) = P(n, t_j) \equiv P(n)\ \forall\, n \in \mathbb{N},\ \forall\, t_i, t_j > 0, \quad \text{and}$$

$$f_{X_{t_i}}(x, t_i) = f_{X_{t_j}}(x, t_j) \equiv f_X(x) \;\forall\, x > 0,\; \forall\, t_i, t_j > 0.$$

Since the probability distribution of a stationary process does not depend on time, the time derivatives $\partial P/\partial t$, $\partial G/\partial t$, and $\partial f_{X_t}/\partial t$ that appear in the master equation (3.2), the corresponding equation for the probability generating function (3.3), and the Kramers-Moyal equation (4.9), respectively, can be set equal to zero. This greatly simplifies the process of solving the equations, and usually the solutions themselves are much easier to work with.

However, it must be remembered that in terms of cell populations, the stationarity assumption implies that distributions from snapshot population data should look the same regardless of when they are taken. (Note that the expression for each individual cell can (and usually does) change in time, but taken as a whole the statistical properties of the population remain constant). In particular, the mean, variance, autocovariance and other functions of ensemble moments are not time-dependent. Therefore, there can not be any synchrony or coupling between cells, and $\mathcal{M}$ and $\mathcal{L}$ must be stationary.

Ergodicity is stronger than stationarity: for a random process in an ergodic population, the distribution obtained from observing a single cell for a time T, $T \to \infty$, is equivalent to the distribution obtained from a single-time snapshot of the population of C cells, $C \to \infty$. For example, if we follow a cell with sample path $\{\eta_\omega(t)\}_{t\geq 0}$ from the stochastic process $\mathcal{N}$, for any $n \in \mathbb{N}$ the proportion of time spent with $\eta_\omega(t) = n$ will be equal to $P(n)$. Since time averages do not depend on time, ergodic systems are also stationary. However, the converse is not true. Models of ergodic systems must assume that all cells are indistinguishable from each other in terms of parameter values and behaviour. The flip-side of this means that ergodic populations can not consist of subpopulations.

Example 5.2–1 — Non-ergodic, stationary systems

Consider a population consisting of two subpopulations of equal size. In the first subpopulation, the cells transcribe stochastically at constant rate μ and degrade at constant rate λ. Cells in the second subpopulation do not transcribe this particular gene, i.e. they never have any mRNA molecules of the type that we are interested in. Assume each subpopulation is at stationarity, and let N_1 and N_2 be the random variables describing the number of mRNA transcripts in a randomly chosen cell from the first and second subpopulation, respectively. N describes the number of transcripts in a cell chosen at random from the population as a whole. Then we have:

$$\Pr(N_1 = n) = \frac{\left(\frac{\mu}{\lambda}\right)^n e^{-\mu/\lambda}}{n!}, \quad n \in \mathbb{N}$$

$$\Pr(N_2 = n) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{so} \quad P(n) = \Pr(N = n) = \frac{1}{2}\Pr(N_1 = n) + \frac{1}{2}\Pr(N_2 = n)$$

$$= \begin{cases} \frac{1}{2}e^{-\mu/\lambda} + \frac{1}{2} & \text{if } n = 0, \\ \frac{1}{2}\frac{\left(\frac{\mu}{\lambda}\right)^n e^{-\mu/\lambda}}{n!} & \text{otherwise.} \end{cases}$$

This distribution is clearly stationary, but it is not ergodic: there are cells in the population (in fact, all of them) whose averaged behaviour over time would not reproduce the distribution of N shown above. On the other hand, each subpopulation taken by itself is ergodic.

Extending the idea of subpopulations, we can easily concoct a toy example that reproduces the stationary distribution of the standard random telegraph model, but is not ergodic. Suppose we have a population where each cell i has constant transcription rate m_i drawn from a random variable M that takes values on the interval $[0, \mu]$, and all cells have degradation rate λ. Then $X = M/\lambda$. Suppose the transcription rates are distributed such that X has the same scaled Beta distribution that we saw for the mixing density for the random telegraph model in Chapter 4 (Eq. (4.3)):

$$f_X(\xi) = \left(\frac{\mu}{\lambda}\right)^{1-\frac{k_{\text{on}}}{\lambda}-\frac{k_{\text{off}}}{\lambda}} \frac{\Gamma\left(\frac{k_{\text{on}}}{\lambda} + \frac{k_{\text{off}}}{\lambda}\right)}{\Gamma\left(\frac{k_{\text{on}}}{\lambda}\right)\Gamma\left(\frac{k_{\text{off}}}{\lambda}\right)} \xi^{\frac{k_{\text{on}}}{\lambda}-1} \left(\frac{\mu}{\lambda} - \xi\right)^{\frac{k_{\text{off}}}{\lambda}-1}, \text{ where } \xi \in \left(0, \frac{\mu}{\lambda}\right).$$

Then the distribution for the number of mRNA transcripts is the same as for the random telegraph model:

$$P(n) = \int_0^{\mu/\lambda} \frac{\xi^n}{n!} e^{-\xi} f_X(\xi)\, d\xi$$

$$= \frac{\Gamma\left(\frac{k_{\text{on}}}{\lambda} + n\right)}{\Gamma\left(\frac{k_{\text{on}}}{\lambda}\right)} \frac{\Gamma\left(\frac{k_{\text{on}}}{\lambda} + \frac{k_{\text{off}}}{\lambda}\right)}{\Gamma\left(\frac{k_{\text{on}}}{\lambda} + \frac{k_{\text{off}}}{\lambda} + n\right)} \frac{\left(\frac{\mu}{\lambda}\right)^n}{n!} {}_1F_1\left(\frac{k_{\text{on}}}{\lambda} + n, \frac{k_{\text{on}}}{\lambda} + \frac{k_{\text{off}}}{\lambda} + n; -\frac{\mu}{\delta}\right).$$

Again, this system is stationary but not ergodic: each cell i has its own constant transcription rate m_i, so over time its averaged behaviour would tend to a Poisson distribution with parameter m_i/λ.

The second example above highlights the fact that a distribution is not sufficient in itself to describe a population.

Many of the models that are currently used assume ergodicity and the stationary solution is used (Iyer-Biswas *et al.* 2009; Shahrezaei and Swain 2008). In particular, multistate models where the gene state is governed by a Markov chain, and the state-dependent transcription and degradation rates are constant, all have a stationary solution.

This sometimes causes confusion as to the meaning of "time-dependent solutions" in these systems. This time dependence is due only to choosing an initial distribution away from the stationary, equilibrium distribution. The time-dependent solution $P(n,t)$ thus describes the relaxation to the usual stationary solution. When enough time has passed for the initial condition to be negligible, the system can be considered to be ergodic and the stationary solution can be used. The transient part of the solution is usually only of interest if one is investigating the response of a whole population to a stimulus or step-change in the environment, for example. Timescales for the duration of this transience are derived in Chapter 8. The example below, from Iyer-Biswas *et al.* 2009 illustrates this scenario.

Example 5.2–2 — Exponential relaxation to stationarity: random telegraph model (Iyer-Biswas *et al.* 2009)

In their 2009 paper, Iyer-Biswas *et al.* 2009 considered the random telegraph model assuming that the system started out of equilibrium: that all cells start in the inactive state with zero mRNA transcripts. Not stated, but implicit in the random telegraph model because of the stochastic switching of the promoter states, is that the cells are uncoupled and so behaviour of each cell is uncorrelated with any of the others. In other words, if the system started at equilibrium it would be ergodic. Here, the authors assume that the system starts with the initial distribution $P(n,0) = \delta_{n,0}$, so we expect it to relax towards the usual stationary solution (4.4).

Denoting here the transcription rate by μ, the degradation rate by λ, and the transition rates between the active and inactive states by k_{on} and k_{off}, they derived the probability generating function $G(z,t)$ for the solution $P(n,t)$:

$$\begin{aligned} G(z,t) &= F_s(t)\,{}_1F_1\left(k_{\text{on}}, k_{\text{on}}+k_{\text{off}}; -\mu(1-z)\right) \\ &\quad + F_{ns}(t)\,{}_1F_1\left(1-k_{\text{off}}, 2-k_{\text{on}}-k_{\text{off}}; -\mu(1-z)\right), \quad \text{where} \\ F_s(t) &\coloneqq {}_1F_1\left(-k_{\text{on}}, 1-k_{\text{on}}-k_{\text{off}}; \mu e^{-t}(1-z)\right), \quad \text{and} \\ F_{ns}(t) &\coloneqq \frac{k_{\text{on}}\mu\, e^{-(k_{\text{on}}+k_{\text{off}})t}(1-z)}{(k_{\text{on}}+k_{\text{off}})(1-k_{\text{on}}-k_{\text{off}})}\,{}_1F_1\left(k_{\text{off}}, 1+k_{\text{on}}+k_{\text{off}}; \mu e^{-t}(1-z)\right), \end{aligned}$$

where ${}_1F_1$ is the confluent hypergeometric function (Abramowitz and Stegun 1964). From here, one needs to obtain $P(n,t)$ by writing G explicitly in polynomial form, or using the identity

$$P(n,t) = \frac{1}{n!}\frac{\partial G}{\partial z}\Big|_{z=0}.$$

A visualisation of the solution $P(n,t)$ is shown in Fig. 5.2, but the expression for $P(n,t)$ itself is extremely unwieldy and is omitted here. It is clear from the time dependence in G, that unless one is specifically interested in the relaxation to the stationary solution

(in this case, timescales up to about four times the lifetime of an mRNA), one should assume that the system starts at stationarity.

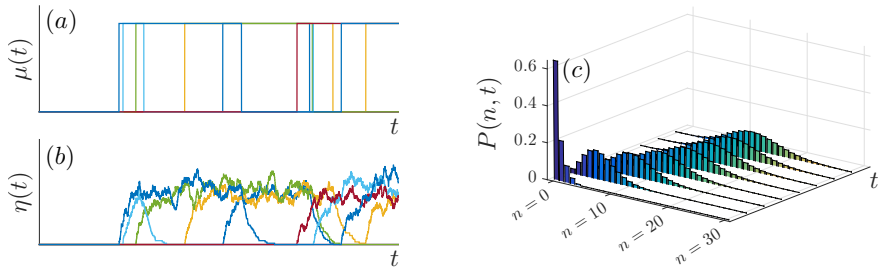

Figure 5.2: Time dependence of the random telegraph model describes only exponential convergence to the stationary distribution. *(a)* Sample paths of the transcription rate $\mathcal{M}$, *(b)* sample paths of the number of mRNA molecules $\mathcal{N}$, and *(c)* the probability distribution $P(n,t)$ for the random telegraph model with initial condition $P(0,0) = 1$ and all cells initialised in the inactive state (Iyer-Biswas *et al.* 2009). The distribution at $t = 0$ is omitted for scaling purposes.

5.2.1 Alternative solution for ergodic systems

Consider a model of an ergodic cell population: the transcription and degradation rates of each cell are sample paths from the same stationary stochastic processes $\mathcal{M}$ and $\mathcal{L}$, respectively, and the population itself starts at stationarity. Choose a cell at random, and denote its sample paths from $\mathcal{M}$ and $\mathcal{L}$ by $\{\mu_\omega(t)\}_{t\geq 0}$ and $\{\lambda_\omega(t)\}_{t\geq 0}$, respectively. As shown in Chapter 3, given these transcription and degradation rates the probability distribution for the number of mRNA transcripts $\mathcal{N}$ for this particular cell will be given by

$$P_\omega(n,t) \coloneqq P(n,t|\mu_\omega,\lambda_\omega) = \frac{x_\omega(t)^n}{n!}e^{-x_\omega(t)}, \quad \text{where}$$
$$x_\omega(t) \coloneqq \int_0^t \mu_\omega(\tau)e^{-\int_\tau^t \lambda_\omega(\tau')\,d\tau'}\,d\tau.$$

Since we have assumed ergodicity, the time average of this sequence $\{P_\omega(n,t)\}_{t\geq 0}$ is equal to the ensemble average $P(n)$ for the whole population, at any time t:

$$P(n) = \lim_{T\to\infty}\frac{1}{T}\int_0^T \frac{x_\omega(t)^n}{n!}e^{-x_\omega(t)}\,dt. \tag{5.2}$$

This expression for $P(n)$ suffers from the fact that for an exact solution in practice, it requires the sample path x_ω to be known $\forall\ t \geq 0$. Nonetheless, it can be used to our advantage when sample paths are periodic, because then the integral only needs to be over a single time period, T say:

$$P(n) = \frac{1}{T}\int_0^T \frac{x_\omega(t)^n}{n!}e^{-x_\omega(t)}\,dt. \tag{5.3}$$

The following example shows how we can use this simple solution to investigate how the choice between stochastic and periodic drives affects population distributions, for example when modelling cell cycles or oscillatory signalling. Such questions have been studied in the context of fluctuating environments, using stochastic simulations and expressions for ensemble moments (Thattai and Oudenaarden 2004; Hilfinger and Paulsson 2011).

Example 5.2–3 — Ergodic systems: stochastic *vs* deterministic drives

Oscillatory expression in single cells has been observed in several systems that could be modelled as being ergodic, due for example to dynamics linked with cell cycles or negative feedback loops. The oscillations are usually noisy and not precisely regular, but since our observations are mRNA or protein counts rather than the transcription or degradation rates themselves, it is natural to wonder whether the oscillations were caused by deterministic periodic drives (at least approximately), or stochastic drives. Additionally, deterministic periodic drives are easier to work with mathematically than stochastic drives, so we may wonder how much the output of our models might be affected if we approximated stochastic drives with deterministic ones.

As an example, consider three directly comparable gene expression models over different timescales T. The transcription rate for each model has sample paths $\{t \mapsto \mu_\omega(t) : \omega \in \Omega_{\mathcal{M}}\}$ that oscillate between the values 0 and 20 with period (or expected period) T, with either

1. sinusoidal form,
2. square wave form, or
3. random telegraph form, the stochastic analogue of the square wave.

For simplicity we take the degradation rate to be $\lambda \equiv 1$ throughout.

Sinusoidal model

Sample paths for the transcription rate of the sinusoidal model are given by

$$\mu_{\sin,\omega}(t) := 10\left[1 + \cos\left(\frac{2\pi}{T}(t+\omega)\right)\right],$$

where necessarily for ergodicity the ω are drawn from the random variable $\Omega_{\sin}$ which has a uniform distribution on the interval $[0, T)$. Since we have an ergodic model we can pick any sample path to use in Eq. (5.3), so without loss of generality take the sample path $\{\mu_{\sin,0}(t)\}_{t\geq 0}$. To obtain the corresponding sample path $\{x_{\sin,0}(t)\}_{t\geq 0}$,

solve the differential rate equation

$$\frac{dx_{\sin,0}}{dt} + x_{\sin,0} = \mu_{\sin,0} = 10\left[1 + \cos\left(\frac{2\pi t}{T}\right)\right]$$

to obtain

$$x_{\sin,0}(t) = 10\left[1 + A\cos\left(\frac{2\pi t}{T} - \alpha\right)\right],$$

where $A := T/\sqrt{T^2 + 4\pi^2}$, $\alpha := \arctan 20$, and the initial condition $x_{\sin,0}(0) = 10[1 + A\cos(-\alpha)]$ has been chosen to ensure periodicity of the solution. Then using Eq. (5.3), we can immediately write down the exact solution $P_{\sin}(n)$ for the sinusoidal model in integral form as

$$P_{\sin}(n) = \frac{10^n}{n!\,T}\int_0^T \left[1 + A\cos\left(\frac{2\pi t}{T} - \alpha\right)\right]^n e^{-10\left[1 + A\cos\left(\frac{2\pi t}{T} - \alpha\right)\right]}\,dt.$$

Squarewave model

Transcription rates for the square wave model are simply step functions that switch between the values 0 and 20 every $T/2$ time units. Without loss of generality we can take the sample path $\mu_{\text{sqw},0}$ that starts with value 20 at $t = 0$, so we have to solve:

$$\frac{dx_{\text{sqw},0}}{dt} + x_{\text{sqw},0} = \begin{cases} 20, & \text{if } t \in \left[kT,\ \left(k + \frac{1}{2}\right)T\right), k \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

We obtain:

$$x_{\text{sqw},0}(t) = \begin{cases} x^{(\text{on})}_{\text{sqw},0}(t) := 20 + [x_{\text{sqw},0}(kT) - 20]\,e^{-t}, & \text{for } t \in \left[kT, \left(k + \frac{1}{2}\right)T\right),\ k \in \mathbb{N}, \\ x^{(\text{off})}_{\text{sqw},0}(t) := x_{\text{sqw},0}\left(\left(k + \frac{1}{2}\right)T\right)e^{-t}, & \text{for } t \in \left[\left(k + \frac{1}{2}\right)T, (k+1)T\right),\ k \in \mathbb{N}, \end{cases}$$

and to ensure ergodicity we need to determine the values of $x_{\text{sqw},0}$ at the switch times kT and $(k + 1/2)T$, $k \in \mathbb{N}$ so that

$$x_{\text{sqw},0}(kT) = x_{\text{sqw},0}(lT), \quad \text{and}$$
$$x_{\text{sqw},0}\left(\left(k + \frac{1}{2}\right)T\right) = x_{\text{sqw},0}\left(\left(l + \frac{1}{2}\right)T\right)\ \forall\, k, l \in \mathbb{N}.$$

Finding the values of $x_{\text{sqw},0}$ at the switching times given any initial condition, our desired values are those in the limit of $k \to \infty$:

$$x_{\text{sqw},0}(kT) = \frac{20e^{-T/2}}{1 + e^{-T/2}}, \quad \text{and}$$
$$x_{\text{sqw},0}\left(\left(k + \frac{1}{2}\right)T\right) = \frac{20}{1 + e^{-T/2}}\ \forall\, k \in \mathbb{N}.$$

Finally,

$$P_{\text{sqw}}(n) = \frac{1}{n!\,T}\left[\int_0^{T/2}\left(x^{(\text{on})}_{\text{sqw},0}(t)\right)^n e^{-x^{(\text{on})}_{\text{sqw},0}(t)}\,dt + \int_0^{T/2}\left(x^{(\text{off})}_{\text{sqw},0}(t)\right)^n e^{-x^{(\text{off})}_{\text{sqw},0}(t)}\,dt\right].$$

Random telegraph model

Denote the solution of the random telegraph model by $P_{\text{rt}}(n,t)$. Analytical expressions for it are already known (Eq. (4.4)) and (Raj and Oudenaarden 2008): setting $\mu = 20$, $\lambda = 1$, and $k_{\text{on}} = k_{\text{off}} = 1/T$, we have:

$$\begin{aligned} P_{\text{rt}}(n,t) &= 20^{1-\frac{2}{T}}\frac{\Gamma\left(\frac{2}{T}\right)}{\Gamma\left(\frac{1}{T}\right)^2}\int_0^{20}\frac{\xi^n}{n!}\,e^{-\xi}\,\xi^{\frac{1}{T}-1}\,(20-\xi)^{\frac{1}{T}-1}\,d\xi \\ &= \frac{\Gamma\left(\frac{1}{T}+n\right)}{\Gamma\left(\frac{1}{T}\right)}\frac{\Gamma\left(\frac{2}{T}\right)}{\Gamma\left(\frac{2}{T}+n\right)}\frac{20^n}{n!}\,{}_1F_1\left(\frac{1}{T}+n,\frac{2}{T}+n;-20\right) \end{aligned}$$

It is easy to see that similarities between the three models vary considerably according to the timescale of the period T. At all timescales the distribution of the random telegraph model has greater weight at n low and n high, caused by the waiting times in either state that are longer than $T/2$. The longer waiting times allow more time for the sample paths of $\mathcal{N}$ to reach the equilibrium states at $n = 0$ and $n = 20$, although this effect is less pronounced when state switching is fast ($T = 1/5$) because the longer waiting times are not significant relative to the timescales of transcription and degradation. Indeed, the variable waiting times of the random telegraph model have a greater effect on the distribution than the wave forms do (i.e. discrete jumps or continuous sinusoids), until T is large: for all timescales shown in Fig. 5.3 apart from $T = 50$, the distribution of the square wave is more similar to the distribution of the sinusoidal model than its own stochastic analogue, the random telegraph model. When $T = 50$, the state switches for the square wave and random telegraph models are so slow that sample paths of $\mathcal{N}$ are able to reach equilibrium in both the active and inactive states, and the transient periods of convergence to either state are negligible. Hence the differences between the square wave and random telegraph models are negligible and far different to the solution of the sinusoidal model, where the expected number of mRNA transcripts in each cell is more dynamic.

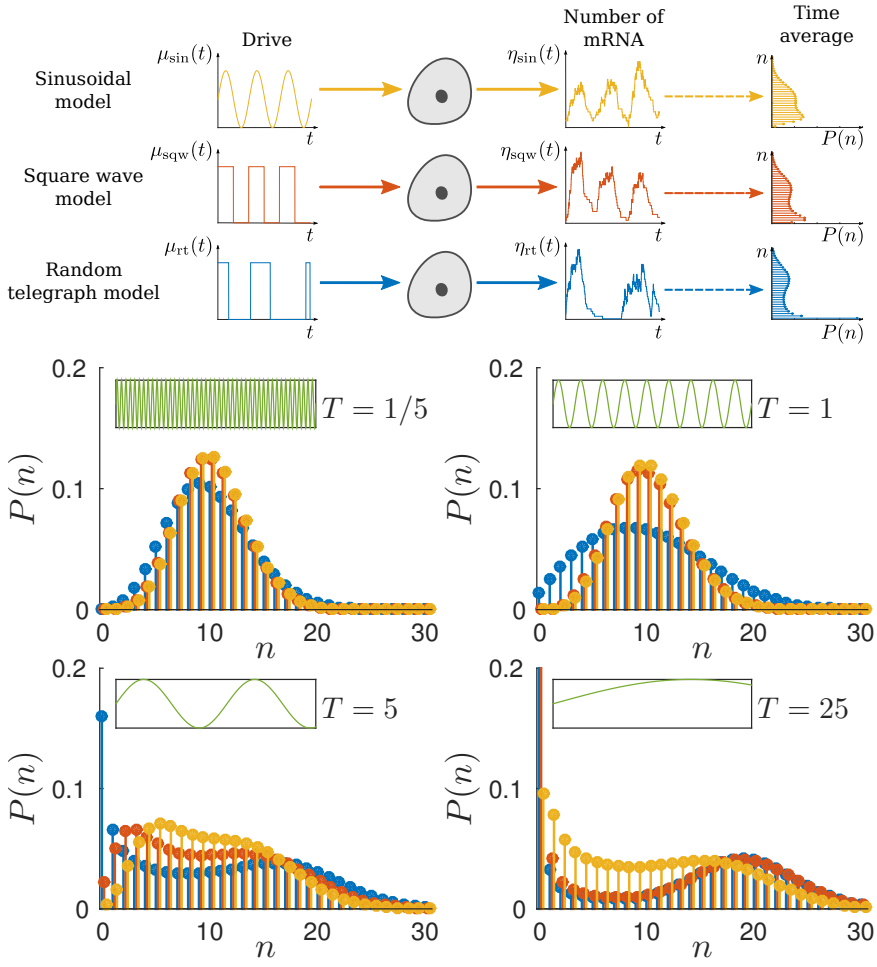

Figure 5.3: Comparison of models with periodic or stochastic transcription rate over a range of timescales. Stationary probability mass functions $P(n)$ for models with transcription rate alternating between the values 0 and 20 via *i)* a sinusoidal wave with period T (yellow), *ii)* a square wave with period T (red), or *iii)* a random telegraph process with expected waiting time $T/2$ in each state (blue). *Insets:* Example sinusoidal wave with the period indicated to illustrate the timescales for each figure relative to each other.

Furthermore, Eq. (5.2) provides another straightforward way to reduce costs when approximating $P(n)$ numerically for an ergodic system using a stochastic simulation algorithm. For models where an exact analytical solution is not known, a common approach is to simulate sample paths of the trancription and degradation rates, $\{\mu(t)\}_{t\in[0,T]}$ and $\{\lambda(t)\}_{t\in[0,T]}$, and use these to simulate the sample path of $\mathcal{N}$, $\{\eta(t)\}_{t\in[0,T]}$. The solution of the model $P(n)$ is then approximated by the proportion of time spent with $\eta(t) = n$, i.e.

$$P(n) \approx \frac{1}{T}\int_0^T \mathbb{1}_n(\eta(t))\,dt,$$

where $\mathbb{1}_n$ is the indicator function.

However, simulating $\{\eta(t)\}_{t\in[0,T]}$ can be computationally expensive, so Eq. (5.2) can reduce these costs by removing the need to simulate a sample path for $\mathcal{N}$ whatsoever. We only require a sample path $\{x(t)\}_{t\in[0,T]}$, which may even be deterministic or piecewise-deterministic, for example the Ladder model illustrated in Example 4.2–2.

5.2.2 Temporal Fano factor

The temporal Fano factor is calculated using the variance and mean of a single timeseries in some time window W. This is in fact the original definition of the Fano factor, used in signal processing to estimate statistical fluctuations of a count variable in a time window (Fano 1947). However, the random variable N_t in transcription-degradation models decreases with degradation events, and hence is not a count variable. We therefore can not use the temporal Fano factor to analyse gene expression data in the same way as it was originally intended. Nevertheless, knowing the Poisson result from Chapter 3 (Eq. (3.10)), we can use the temporal Fano factor to infer windows of stationarity and timescales when analysing single-cell time course data.

Consider a stochastic simulation $\{\eta(t)\}_{t\geq 0}$ of a cell under the leaky random telegraph model, where the transcription rate in the inactive and active promoter states are μ_0 and μ_1, respectively, and the degradation rate is λ. Given sufficient time before a state transition, there will be a time window W where $\{\eta(t)\}_{t\in W}$ can be considered to be at stationarity: its behaviour averaged over W will be described by a Poisson distribution with parameter μ_0/λ (resp. μ_1/λ) in the inactive (resp. active) state. Therefore, the temporal Fano factor for W should be approximately equal to one. After each promoter state transition there is a transient period W' during which the statistics of $\{\eta(t)\}_{t\in W'}$ are time-dependent. Over W' the averaged behaviour will therefore not be described by a Poisson distribution, and the temporal Fano factor will be greater than one.

Fig. 5.4a illustrates this dynamical behaviour of the temporal Fano factor using a fixed window length. The moving windows are indicated by the horizontal bars, and are akin to a moving average. Notice that the right-hand edge of the moving windows (indicated by the purple solid line) are a good predictor of the switching times of the gene state, as the change in the distribution causes the Fano factor to increase sharply.

We can also calculate the temporal Fano factor over growing windows, to estimate the time taken for the sample path to describe a stationary regime immediately following a switch in the gene state. We will refer to this function as the *cumulative Fano factor*. To be more specific, for a start time t_0 that we choose, the cumulative Fano factor at time T for the sample path η_ω is its temporal Fano factor calculated over the time interval $[t_0, T]$. Fig. 5.4b zooms in on the first part of the sample path $\eta_\omega(t)$ shown in Fig. 5.4a, and shows four instances of the cumulative Fano factor, starting at each of the first four gene switching event times. For example, the height of the blue line at time T indicates the temporal Fano factor for $\{\eta_\omega(t)\}_{t\in[0,T]}$, since the first switch event takes place at $t = 0$. Notice that the cumulative Fano factor increases initially as η_ω adjusts to the regime change after the switch, before returning towards the value of one as η_ω increasingly behaves like a sample

path from a stationary process. The rate at which it does this gives an indication as to the timescales involved in transitioning between the 'active' and the 'inactive' regime. The cumulative Fano factor increases again at each gene switch, thus it can also be used to estimate the switching event time where the next cumulative Fano factor function should start.

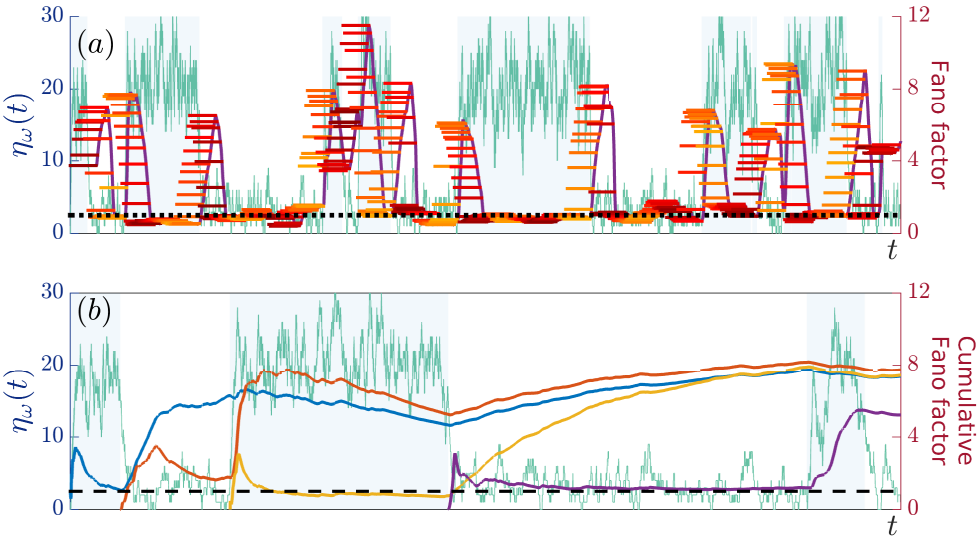

Figure 5.4: Temporal Fano factor over time for a single sample path $\eta_\omega : t \to \eta_\omega(t)$ of the leaky random telegraph model. Time periods when the gene is in the active state are indicated by background shading. *(a)* The height of each horizontal bar corresponds to the temporal Fano factor of $\{\eta(t)\}_{t\in W}$, with the time window W indicated by the length and position of the bar along the time axis. When W falls upon a section of the sample path that is at stationarity, the Fano factor is at or close to the value 1 (indicated by the black dotted line). *(b)* The Fano factor calculated over growing windows (cumulative Fano factor), starting from each of the first four gene switch event times shown in *(a)*. A cumulative Fano factor starting at time t_0 has value approximately one at time t when the sample path describes a stationary process over the domain $[t_0, t]$.

5.3 Discussion

This chapter presented some immediate corollaries of the Poisson mixture result based on ensemble and temporal moments. The ability to decouple the Poisson variation from the model-specific variation mentioned at the end of Chapter 3 manifests itself cleanly for the ensemble moments of N_t, which can easily be written in terms of moments of X_t. It is well-known in the field that (ensemble) gene expression noise can be decomposed into "intrinsic" and "extrinsic" components, although there is sometimes confusion as to the precise definition of each; often, a list of possible sources of each type of noise is given, as opposed to a precise mathematical definition (Swain *et al.* 2002; Wikipedia 2014). Using our expressions for the ensemble moments, the Fano factor and squared coefficient of variation were shown to break down naturally into the sum of a "Poisson" component common to all transcription-degradation models, and a component from the

mixing density f_{X_t} that is model-specific. Each of these noise components corresponds to the "intrinsic" and "extrinsic" noise discussed in the influential paper by Swain *et al.* 2002, clarifying precisely what they represent in the context of gene transcription models.

The use of temporal averages to analyse gene transcription models is relatively rare, apart from data analyses using autocorrelation. We have not discussed autocorrelation here, as its properties are well-defined and we would not be able to contribute any more here. Alternatively, we capitalised on the insights given by the Poisson mixture result to give a few novel examples using temporal averages. In particular, we are not aware of any investigations into the potential uses of the temporal Fano factor for the analysis of time-lapse gene transcription data beyond our short discussion in Section 5.2.2.

Part III

Multistate models

Part II: Multistate models

Introduction

In the early 1990s, Ko *et al.* observed bimodal and long-tailed gene expression in single cells (Ko *et al.* 1990). Ko attributed this heterogeneity to "the randomness corresponding to the random timing of molecular collisions and dissociations between transcription factors and a gene copy, since at any instant each copy is thought to be either 'switched on' by having a transcription complex bound to it, or 'switched off' by not having a transcription complex bound" (Ko 1991). In doing so, he sparked a paradigm shift from using deterministic models of gene expression, to models based on stochastic changes in the promoter architecture, especially for eukaryotes. As a general framework, multistate promoter models are biologically interpretable and flexible enough to model most hypotheses for transcriptional regulation. As such they are now almost ubiquitous, and are regularly used to model the effects of nucleosome occupancy and chromatin remodelling (Kim and E. K. O'Shea 2008; Tirosh and Barkai 2008), TATA box strength (Raser and E. J. O'Shea 2004; Hornung *et al.* 2012), the binding of transcription factors, and the number of their sites (Raj *et al.* 2006; Suter *et al.* 2011; Senecal *et al.* 2014), for example.

In this part we focus on stochastic models under this "multistate" framework. We use the term *multistate model* to refer to the class of models where the temporal behaviour of the transcription and/or degradation rates in single cells is governed by a continuous-time random walk between a finite number of discrete states. We refer to the subset of multistate models with a Markovian random walk by *Markov chain multistate models*. In order to avoid confusion with the term "state of the system", which is sometimes used to refer to the number of mRNA transcripts in a cell, we will refer to the discrete states here as *promoter states* or *gene states*. Each promoter state has a transcription and degradation rate associated with it, which we will take to be constants unless otherwise stated, as is usual in the literature. Thus sample paths of the stochastic processes $\mathcal{M}$ and $\mathcal{L}$ are step functions, with waiting times in each state drawn from a probability function specified by the model. As is usual for multistate models, switching events are assumed to occur independently of other cells and at random times, and therefore the models considered in this part describe ergodic systems and possess a stationary solution. As explained in Sections 3.2 and 5.2, we are not interested in the convergence to the stationary distribution from an initial condition, so we will be deriving and analysing stationary solutions $P(n)$ in this part.

In Chapter 6, we solve the multistate model with a cyclic promoter structure in two different ways, each of which conferring different insights into the behaviour and properties of the solution. One of these insights leads us to discuss in Chapter 7 how the multistate Kramers-Moyal equations for multistate models governed by a continuous-time Markov

chain (i.e., waiting times in each gene state are exponentially distributed), can help us deduce some models which are solvable, and the form that their solutions will take. We also obtain solutions for non-Markovian gene-state switching by drawing upon classical results in the electrical engineering and information theory literature, as well as relatively recent results from investigations into Dubins-Freedman processes. Finally, in Chapter 8 we introduce a quantitative characterisation of timescales for gene switching in order to delineate between "fast" and "slow" switching behaviour and classify parameter regimes.

Chapter 6

Cyclic models

Gene transcription often occurs in discrete burst periods, which are punctuated by intervals of inactivity. Quantitative data analysis for such transcription kinetics commonly uses the random telegraph model, a two-state (ON-OFF) model of gene transcription, for which the exact solution is known (Raj *et al.* 2006). However, this model implies that the waiting times in both the ON and OFF states should be exponentially distributed, whereas recent time-lapse recordings show that the silent intervals often show a refractory period (Suter *et al.* 2011; Harper *et al.* 2011; Kandhavelu *et al.* 2012). A model with several inactive states in sequence is sufficient to capture the dynamical behaviour that has been observed, so could provide a more realistic basis for data analysis (Suter *et al.* 2011; Zoller *et al.* 2015; Rybakova *et al.* 2015). We will refer to this model with one active promoter state and M inactive promoter states as the M-cycle (see Fig. 6.1).

In this chapter we will derive the exact stationary solution for the M-cycle via two different routes. First, we obtain $P(n)$ concisely in terms of standard functions and explicitly in terms of the parameters of the model, via the probability generating function. We will use it to infer noise characteristics of the model, and perform preliminary maximum likelihood estimation (MLE) and model selection using the Akaike Information Criterion (AIC) on *in silico* generated data. Second, we derive the solution of the M-cycle using the Poisson mixture result derived in Chapter 3 (Eq. (3.11)), by solving the appropriate multistate Kramers-Moyal equations (4.9) to obtain f_X. We use the solution to help us navigate parameter space when searching for different qualitative behaviours for $P(n)$.

6.1 Theoretical framework for the M-cycle model

As described briefly above, we extend the random telegraph (two-state) model to account for refractory periods in gene reactivation by adding several inactive promoter states that are visited in sequence. We will refer to the cyclic model with an active promoter state and M inactive states as the M-cycle:

Suppose we have a population of identical, uncoupled cells. In each cell, the promoter must cycle through M inactive (OFF) states $s_1, \ldots, s_M$ in sequence before returning to the active (ON) state s_{on}, where the gene is transcribed according to a Poisson process with rate μ. Regardless of promoter state, degradation of mRNA molecules occurs stochastically at rate λ per molecule per unit time. Stochastic transitions away from promoter state s_i occur at rate k_i (see Fig. 6.1).

Maintaining the same notation as in Part II, the random variables that describe the

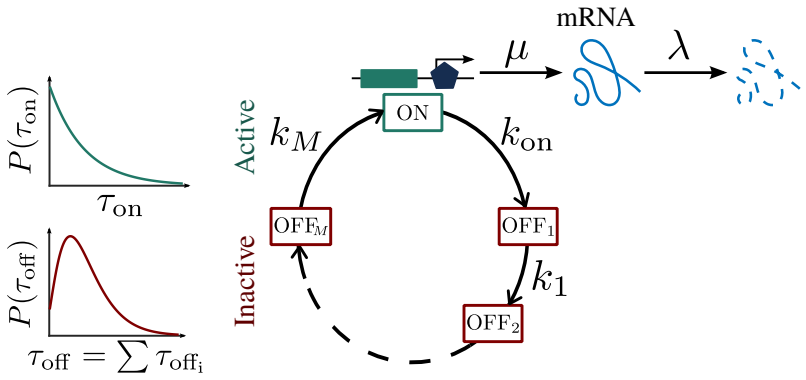

Figure 6.1: M-cycle model of gene transcription. The gene state cycles sequentially between the single active (ON) state and the M inactive (OFF) states, with transition rates k_{on} and k_i, $i = 1, \ldots, M$. When the gene is in the active state, transcription takes place as a Poisson process with constant rate μ, and mRNA molecules degrade as a first-order reaction with constant rate λ.

mRNA copy number and promoter state at time t will be denoted by N_t and S_t, respectively. We will consider the expanded state space $\{N_t, S_t\}$, and write down the master equation in terms of the joint probability mass functions

$$P_i(n,t) := \Pr(N_t = n, S_t = s_i) \quad \text{and} \quad P_{\text{on}}(n,t) := \Pr(N_t = n, S_t = s_{\text{on}}).$$

Our desired marginal distribution for the mRNA copy number, $P(n,t)$, will then be given by

$$P(n,t) = P_{\text{on}}(n,t) + \sum_{i=1}^{M} P_i(n,t).$$

To simplify notation before we continue, any indices should be considered modulo $M+1$, with $M+1 \equiv 0 \equiv \text{on}$. For example, $P_{M+1} \equiv P_0 \equiv P_{\text{on}}$. Then, the master equations for the joint probability mass functions are:

$$\frac{dP_i(n,t)}{dt} = \lambda[(n+1)P_i(n+1,t) - nP_i(n,t)] + k_{i-1}P_{i-1}(n,t) - k_iP_i(n,t), \qquad \text{for } \; i = 1, \ldots, M, \tag{6.1a}$$

$$\frac{dP_{\text{on}}(n,t)}{dt} = \lambda[(n+1)P_{\text{on}}(n+1,t) - nP_{\text{on}}(n,t)] + \mu[P_{\text{on}}(n-1,t) - P_{\text{on}}(n,t)] + k_MP_M(n,t) - k_{\text{on}}P_{\text{on}}(n,t). \tag{6.1b}$$

Notice that the stochastic processes describing the promoter state, transcription, and degradation are all stationary, and recall that the cells are assumed to be uncoupled and described by the same model and parameter values. Hence, any time dependence in the solution $P(n,t)$ would only describe the system's convergence from an initial condition towards the stationary solution $P(n)$ (see Section 5.2). Most snapshot data are assumed to be of a population at stationarity, and the expressions for $P(n,t)$ can make analysis extremely unwieldy, so for simplicity we will assume in the following that the system is

already at stationarity and will omit the variable t.

6.2 Solution for $P(n)$ via the probability generating function

In this section we will follow the usual route for solving a system of master equations of the form (6.1), by transforming them into a system of ordinary differential equations for the corresponding probability generating functions. A solution for this model has been published using this method, but the derivation is opaque in parts and the final expression is not given explicitly in terms of the parameters of the using a relevant adjugate matrix to show that $P(n)$ can be expressed concisely in terms of gamma functions and hypergeometric functions. The relevant parameter set is the model parameters themselves, and the eigenvalues of the transition matrix for the promoter cycle. As a result, properties of the solution, including expressions for the moments and noise, are transparent.

We will use this expression for $P(n)$ to show how the noise characteristics can be tuned with the timescales at the promoter, and perform maximum likelihood estimation (MLE) and model selection using the Akaike Information Criterion (AIC) on *in silico* generated data. In doing so, we exemplify how the assumption of the minimal two-state (random telegraph) model can lead to gross mis-estimation of the underlying parameters, and demonstrate that for many parameter regimes snapshot population data cannot be used to infer the underlying timescales. This quantitative data analysis is only possible due to our possession of the algebraic expression for $P(n)$ that we derive in the next subsection.

6.2.1 Derivation of the solution

Define for each state s_i, $i = 1, \ldots, M, \text{on}$, the stationary probability generating functions

$$G_i(z) \coloneqq \sum_{n=0}^{\infty} z^n P_i(n), \qquad G_{\text{on}}(z) \coloneqq \sum_{n=0}^{\infty} z^n P_{\text{on}}(n),$$

and

$$G(z) \coloneqq \sum_{n=0}^{\infty} z^n P(n) = G_{\text{on}}(z) + \sum_{i=1}^{M} G_i(z).$$

Transform the master equations (6.1) in the usual manner, by multiplying through by z^n and summing over $n \in \mathbb{N}$ to obtain, at stationarity:

$$\lambda(z-1)\frac{dG_i}{dz} = k_{i-1}G_{i-1}(z,t) - k_i G_i(z,t), \quad i = 1, \ldots, M, \tag{6.2a}$$

$$\lambda(z-1)\frac{dG_{\text{on}}}{dz} = k_M G_M(z,t) - k_{\text{on}} G_{\text{on}}(z) + \mu(z-1)G_{\text{on}}(z). \tag{6.2b}$$

Without loss of generality let $\lambda = 1$, and to simplify notation, make the change of variable

$$s \coloneqq \mu(z-1); \qquad H_i(s) \coloneqq G_i(z), \quad i = 1, \ldots, M, \text{on},$$

and write Eqs (6.2) in the form

$$\begin{pmatrix} s\frac{d}{ds}+k_1 & & & & -k_{\text{on}} \\ -k_1 & s\frac{d}{ds}+k_2 & & & \\ & -k_2 & \ddots & & \\ & & \ddots & s\frac{d}{ds}+k_M & \\ & & & -k_M & s\frac{d}{ds}+k_{\text{on}} \end{pmatrix} \begin{pmatrix} H_1 \\ H_2 \\ \vdots \\ H_M \\ H_{\text{on}} \end{pmatrix} = \begin{pmatrix} 0 & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & s \end{pmatrix} \begin{pmatrix} H_1 \\ H_2 \\ \vdots \\ H_M \\ H_{\text{on}} \end{pmatrix} \tag{6.3}$$

i.e.

$$(\theta I - K)\,\mathbf{H} = s\,\mathbf{e}_{M+1}\mathbf{e}_{M+1}^T\,\mathbf{H}, \tag{6.4}$$

where

$$\theta \coloneqq s\frac{d}{ds}, \quad \mathbf{H} \coloneqq \begin{pmatrix} H_1 \\ \vdots \\ H_M \\ H_{\text{on}} \end{pmatrix}, \quad \mathbf{e}_{M+1} \coloneqq \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

and K is the state transition matrix

$$K \coloneqq \begin{pmatrix} -k_1 & & & & k_{\text{on}} \\ k_1 & -k_2 & & & \\ & k_2 & \ddots & & \\ & & \ddots & -k_M & \\ & & & k_M & -k_{\text{on}} \end{pmatrix}$$

Before we continue, let us take a small detour to consider some properties of the promoter state transition matrix K; our desired distribution $P(n) \equiv \Pr(N_t = n)$ is a marginal of the joint distribution $\Pr(N_t = n, S_t = s_i)$, so integral conditions based on the properties of the continuous-time Markov process $\{S_t\}_{t\geq 0}$, with associated transition rate matrix K, will be required later.

Denote the probability of being in state s_i at time t by $P_{_S}(s_i, t) \coloneqq \Pr(S_t = s_i)$, and define

$$\mathbf{P}_{_S}(\mathbf{s}, t) \coloneqq \begin{pmatrix} P_{_S}(s_1, t) \\ \vdots \\ P_{_S}(s_M, t) \\ P_{_S}(s_{\text{on}}, t) \end{pmatrix}.$$

By definition of the Markov chain $\{S_t\}_{t\geq 0}$, we have

$$\frac{d\mathbf{P}_{_S}}{dt} = K\,\mathbf{P}_{_S}.$$

K clearly has a zero eigenvalue[1], and by Gershgorin's circle theorem (Horn and Johnson

[1]The columns of K sum to zero, for example.

1999) the non-zero eigenvalues of K have negative real parts. Hence a stationary solution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M, \pi_{\text{on}})^T$ always exists, that satisfies

$$K\boldsymbol{\pi} = \mathbf{0}\,; \quad \sum_{i=1}^{M+1} \pi_i = 1.$$

$\boldsymbol{\pi}$ is the eigenvector corresponding to the zero eigenvalue of K, which is normalized so that its elements sum to 1 (since $\boldsymbol{\pi}$ represents a probability mass function). It can easily be shown that $\boldsymbol{\pi}$ has elements

$$\pi_i = \frac{\prod_{j\neq i} k_j}{\sum_{r=1}^{M+1} \prod_{j\neq r} k_j}.$$

For example for $M = 2$, $\pi_{\text{on}} = \pi_3 = k_1k_2/(k_2k_{\text{on}} + k_1k_{\text{on}} + k_1k_2)$.

Returning to the derivation of $P(n)$, sum the rows of Eq. (6.3) to see that

$$\frac{dH}{ds} \equiv \frac{d}{ds}(H_1 + \cdots + H_M + H_{\text{on}}) = H_{\text{on}}(s), \tag{6.5}$$

so to determine $H(s) \equiv G(z)$ we need only determine H_{on}, and integrate once under the condition $H(0) = G(1) = \sum_n P(n) = 1$.

In order to obtain an uncoupled equation for H_{on}, we will employ a trick using an adjugate matrix and capitalise on the structure of the model. Multiply both sides of Eq. (6.4) on the left by the transposed matrix of cofactors $\text{adj}(\theta I - K)$, remembering that for any square invertible matrix A, $\text{adj}(A)\,A = \det(A)I$:

$$\begin{aligned}\det(\theta I - K)\,\mathbf{H} &= \text{adj}(\theta I - K)\, s\, \mathbf{e}_{M+1}\, \mathbf{e}_{M+1}^T\, \mathbf{H} \\ &= \begin{bmatrix} (-1)^{M+2}\mathrm{M}_{M+1,1} \\ \vdots \\ (-1)^{2M+2}\mathrm{M}_{M+1,M+1} \end{bmatrix} s\, H_{\text{on}},\end{aligned} \tag{6.6}$$

where $\mathrm{M}_{i,j}$ is the $(i,j)^{\text{th}}$ minor of $[\theta I - K]$.

Notice that the submatrix of $[\theta I - K]$ formed by deleting the last row and last column is lower triangular, so

$$(-1)^{2M+2}\mathrm{M}_{M+1,M+1} = \mathrm{M}_{M+1,M+1} = \prod_{i=1}^{M}(\theta + k_i).$$

Also, the determinant is equal to the product of the eigenvalues and hence

$$\det(\theta I - K) = \prod_{i=1}^{M+1}(\theta - \nu_i) = \theta \prod_{i=1}^{M}(\theta - \nu_i),$$

where the ν_i are the eigenvalues of the matrix K, and without loss of generality we have taken ν_{M+1} to be the zero eigenvalue. The first coordinate of this equation is now an

uncoupled equation in terms of $H_{\rm on}$ only:

$$\theta \prod_{i=1}^{M} (\theta - \nu_i)\, H_{\rm on}(s) = \prod_{i=1}^{M} (\theta + k_i)\, s\, H_{\rm on}(s)$$
$$= s \prod_{i=1}^{M} (\theta + k_i + 1)\, H_{\rm on}(s),$$

where we used the operator identity $(\theta + a)s = s(\theta + a + 1)$, for a a constant. This differential equation is satisfied by the generalized hypergeometric function ${}_M F_M$ (Erdelyi 1953), where

$${}_p F_q \left[\begin{matrix} a_1, & a_2 & \ldots, & a_p \\ b_1, & b_2 & \ldots, & b_q \end{matrix} ;\, z \right] := \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n \ldots (a_p)_n}{(b_1)_n (b_2)_n \ldots (b_q)_n} \frac{z^n}{n!},$$

where $(x)_n := x(x+1)\ldots(x+n-1) = \Gamma(x+n)/\Gamma(x)$ is Pochhammer's function[2]. Thus we can immediately write down the general solution for $H_{\rm on}$:

$$H_{\rm on}(s) = c_0\, {}_M F_M \left[\begin{matrix} 1+k_1, & \ldots, & 1+k_{_M} \\ 1-\nu_1, & \ldots, & 1-\nu_{_M} \end{matrix} ;\, s \right]$$
$$+ \sum_{i=1}^{M} c_i\, s^{\nu_i}\, {}_M F_M \left[\begin{matrix} & 1+k_1+\nu_i, & \ldots\ldots, & 1+k_{_M}+\nu_i \\ 1+\nu_i, & 1-\nu_1+\nu_i, & ..\vee.., & 1-\nu_{_M}+\nu_i \end{matrix} ;\, s \right], \qquad (6.7)$$

where $..\vee..$ denotes suppression of the term $1-\nu_i+\nu_i$, and $c_0, \ldots, c_{_M}$ are constants to be determined.

Notice that $\nu_i \in \mathbb{C}$, and we want $s^{\nu_i} \in \mathbb{R}\ \forall\ s \in \mathbb{R}$. Hence $c_i = 0$ for $i = 1, \ldots, M$. We are left only with c_0, which is given by:

$$c_0 = H_{\rm on}(0) \equiv G_{\rm on}(1) = \pi_{\rm on} = \frac{\prod_{j=1}^{M} k_j}{\sum_{r=1}^{M+1} \prod_{j\neq r} k_j}.$$

Hence Eq. (6.7) is reduced to

$$H_{\rm on}(s) = \pi_{\rm on}\, {}_M F_M \left[\begin{matrix} k_1+1, & \ldots, & k_{_M}+1 \\ -\nu_1+1, & \ldots, & -\nu_{_M}+1 \end{matrix} ;\, s \right]$$
$$=: \pi_{\rm on}\, {}_M F_M \left[\begin{matrix} (k)+1 \\ (-\nu)+1 \end{matrix} ;\, s \right],$$

where here we have introduced the notation in the latter equation, similar to that used in (Slater 1966), for compactness and clarity.

[2]The notation $(x)_n$ itself is known as *Pochhammer's symbol*. Confusingly, in combinatorics and statistics Pochhammer's function is usually denoted $x^{(n)}$ and is referred to as the *rising factorial*, whereas Pochhammer's symbol $(x)_n$ denotes the *falling factorial*. However, in the theory of special functions and in particular the hypergeometric functions that we will be using, Pochhammer's symbol is standard notation for Pochhammer's function, or the rising factorial, as defined in the text.

Now, recall from Eq. (6.5) that we can integrate $H_{\text{on}}(s)$ to obtain $H(s)$. Using standard properties of generalized hypergeometric functions (Erdelyi 1953), we get

$$H(s) = \pi_{\text{on}} \frac{\prod_{j=1}^{M}(-\nu_j)}{\prod_{j=1}^{M} k_j} \, {}_M F_M \left[\begin{matrix} (k) \\ (-\nu) \end{matrix} ; \; s \right] + c_{_H},$$

where $c_{_H}$ is a constant of integration. To see that $c_{_H} = 0$, we can use our expression for H_{on} and Eq. (6.3) to obtain $H_{_M}, \ldots, H_1$ via back substitution. Since these operations will only involve differentiation, no function H_i, $i = 1, \ldots, M$, will have any additive constant terms. Therefore, neither will $H := H_1 + \cdots + H_{_M} + H_{\text{on}}$, so we must have $c_H = 0$.

Furthermore, we need $1 = \sum_n P(n) = G(1) = H(0)^{\dagger}$, so we are left with

$$\begin{aligned} H(s) &= {}_M F_M \left[\begin{matrix} (k) \\ (-\nu) \end{matrix} ; \; s \right], \quad \text{or} \\ G(z) &= {}_M F_M \left[\begin{matrix} (k) \\ (-\nu) \end{matrix} ; \; \mu(z-1) \right]. \end{aligned} \tag{6.8}$$

Now expand $G(z)$ in a Taylor expansion around $z = 0$ so that we will be able to find $P(n)$ by reading off the coefficient of z^n:

$$\begin{aligned} G(z) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} \left. \frac{d^n G}{dz^n} \right|_{z=0} \\ &= \sum_{n=0}^{\infty} \frac{z^n}{n!} \, \mu^n \, \frac{(k_1)_n \ldots (k_M)_n}{(-\nu_1)_n \ldots (-\nu_{_M})_n} \, {}_M F_M \left[\begin{matrix} (k)+n \\ (-\nu)+n \end{matrix} ; \; -\mu \right]. \end{aligned}$$

Again using some notation introduced in (Slater 1966) for compactness:

$$\Gamma \left[\begin{matrix} (k)+n, & (-\nu) \\ (k), & (-\nu+n) \end{matrix} \right] := \frac{\Gamma(k_1+n)}{\Gamma(k_1)} \cdots \frac{\Gamma(k_{_M}+n)}{\Gamma(k_{_M})} \, \frac{\Gamma(-\nu_1)}{\Gamma(-\nu_1+n)} \cdots \frac{\Gamma(-\nu_{_M})}{\Gamma(-\nu_{_M})+n},$$

we finally have

$$P(n) = \frac{(k_1)_n \ldots (k_M)_n}{(-\nu_1)_n \ldots (-\nu_{_M})_n} \, \frac{\mu^n}{n!} \, {}_M F_M \left[\begin{matrix} (k)+n \\ (-\nu)+n \end{matrix} ; \; -\mu \right]$$

[†]Incidentally, this property provides us with an expression for the pseudo-inverse of the promoter state transition matrix K:

$$1 = H(0) = \frac{\prod_{j=1}^{M}(-\nu_j)}{\prod_{j=1}^{M} k_j} \, \pi_{\text{on}} = \frac{\prod_{j=1}^{M}(-\nu_j)}{\prod_{j=1}^{M} k_j} \, \frac{\prod_{j=1}^{M} k_j}{\sum_{r=1}^{M+1} \prod_{j \neq r} k_j},$$

i.e.

$$\sum_{r=1}^{M+1} \prod_{j \neq r} k_j = \prod_{j=1}^{M} (-\nu_j) = (-1)^M \times \text{ pseudo-inverse of } K.$$

$$= \Gamma \begin{bmatrix} (k)+n, & (-\nu) \\ (k), & (-\nu+n) \end{bmatrix} \frac{\mu^n}{n!}\, {}_M F_M \begin{bmatrix} (k)+n \\ (-\nu)+n \end{bmatrix} ; -\mu \Bigg]. \tag{6.9}$$

Example 6.2–1 — Rederivation of the solution of the random telegraph model (1-cycle)

It is easy to check that our results reproduce the known solution of the random telegraph model (Ko 1991; Peccoud and Ycart 1995; Raj *et al.* 2006; Iyer-Biswas *et al.* 2009). Substituting $M = 1$ into the above, we have

$$K = \begin{pmatrix} -k_1 & k_{\text{on}} \\ k_1 & -k_{\text{on}} \end{pmatrix}, \tag{6.10}$$

with eigenvalues $\nu_1 = -(k_{\text{on}} + k_1)$ and $\nu_2 = 0$. Eq. (6.9) then becomes

$$P(n) = \frac{\Gamma(k_1+n)}{\Gamma(k_1)} \frac{\Gamma(k_{\text{on}}+k_1)}{\Gamma(k_{\text{on}}+k_1+n)} \frac{\mu^n}{n!}\, {}_1F_1\left(k_1+n, k_{\text{on}}+k_1+n; -\mu\right), \tag{6.11}$$

as shown in (Raj *et al.* 2006; Iyer-Biswas *et al.* 2009).

Example 6.2–2 — Explicit solution for the 2-cycle

The 2-cycle has already been proposed to account for the non-exponential waiting times observed in the inactive gene state in single-molecule, single-cell time course data (Suter *et al.* 2011; Molina *et al.* 2013; Harper *et al.* 2011; Kandhavelu *et al.* 2012). However, an algebraic expression for $P(n)$ was not available so fitting and analysis was done computationally. Using Eq. (6.9) we can straightforwardly write the down the required expression.

For $M = 2$, $(\nu) = \nu_1, \nu_2$ are the non-zero eigenvalues of K, which are

$$\nu_{1,2} = \frac{1}{2}\left(-k_{\text{on}} - k_1 - k_2 \pm \sqrt{(k_{\text{on}} - k_1 - k_2)^2 - 4k_1k_2}\right), \tag{6.12}$$

so the stationary solution of the model is

$$P(n) = \frac{\Gamma(k_1+n)}{\Gamma(k_1)}\frac{\Gamma(k_2+n)}{\Gamma(k_2)}\frac{\Gamma(-\nu_1)}{\Gamma(-\nu_1+n)}\frac{\Gamma(-\nu_2)}{\Gamma(-\nu_2+n)}\frac{\mu^n}{n!}\,{}_2F_2\left[\begin{matrix}k_1+n, & k_2+n\\ -\nu_1+n, & -\nu_2+n\end{matrix};\,-\mu\right]. \tag{6.13}$$

6.2.2 Noise regulation

With the exact solution, and certain other elements of its derivation, we can obtain properties of the M-cycle and analyse how they differ from the two-state random telegraph model (the 1-cycle).

First, the moments can be straightforwardly obtained from the probability generating function (Eq. (6.8)):

$$\mathbb{E}\,(N) = \left.\frac{dG}{dz}\right|_{z=1} = \mu\,\frac{\prod_{i=1}^{M} k_i}{\prod_{i=1}^{M}(-\nu_i)}, \tag{6.14}$$

and

$$\mathrm{Var}(N) = \left.\frac{d^2G}{dz^2}\right|_{z=1} + \left.\frac{dG}{dz}\right|_{z=1} - \left(\left.\frac{dG}{dz}\right|_{z=1}\right)^2 \tag{6.15}$$

$$= \mu\,\frac{\prod_{i=1}^{M} k_i}{\prod_{i=1}^{M}(-\nu_i)}\left(\mu\,\frac{\prod_{i=1}^{M}(k_i+1)}{\prod_{i=1}^{M}(-\nu_i+1)} + 1 - \mu\,\frac{\prod_{i=1}^{M} k_i}{\prod_{i=1}^{M}(-\nu_i)}\right). \tag{6.16}$$

However, it would be helpful to write these expressions explicitly in terms of the expected waiting times in each state, denoted

$$\tau_{\mathrm{on}} = \frac{1}{k_{\mathrm{on}}}, \quad \tau_i = \frac{1}{k_i}, \quad \text{and} \quad \tau_{\mathrm{off}} = \sum_{i=1}^{M}\tau_i.$$

Thus we need to employ some manoeuvres to obtain expressions for $\prod_{i=1}^{M}(-\nu_i)$ and $\prod_{i=1}^{M}(-\nu_i+1)$. For the former, we continue from the last footnote remarking that $H(0)=1$ implies that:

$$\begin{aligned}\prod_{j=1}^{M}(-\nu_j) &= \sum_{r=1}^{M+1}\prod_{j\neq r} k_j\\ &= \sum_{r=1}^{M+1}\frac{1}{k_r}\prod_{i=1}^{M+1} k_i\\ &= \sum_{r=1}^{M+1}\tau_r\prod_{i=1}^{M+1}\frac{1}{\tau_i}\\ &= \frac{\tau_{\mathrm{on}}+\tau_{\mathrm{off}}}{\tau_{\mathrm{on}}\prod_{i=1}^{M}\tau_i}.\end{aligned} \tag{6.17}$$

In order to obtain a similar expression for $\prod_{i=1}^{M}(-\nu_i+1)$, notice that it is equal to the determinant of the matrix $[I-K]$ (since $\nu_{M+1}=0$):

$$\begin{aligned}\prod_{i=1}^{M}(-\nu_i+1) &= \prod_{i=1}^{M+1}(-\nu_i+1) \qquad (6.18)\\ &= \det(I-K)\\ &= \det\begin{pmatrix} 1+k_1 & & & & -k_{\text{on}} \\ -k_1 & 1+k_2 & & & \\ & -k_2 & \ddots & & \\ & & \ddots & 1+k_M & \\ & & & -k_M & 1+k_{\text{on}} \end{pmatrix}\\ &= (-1)^{M+2}(-k_{\text{on}})(-k_1)\dots(-k_M)+(-1)^{2M+2}(1+k_{\text{on}})(1+k_1)\dots(1+k_M)\\ &= (1+k_{\text{on}})\prod_{i=1}^{M}(1+k_i)-k_{\text{on}}\prod_{i=1}^{M}k_i\\ &= \frac{(\tau_{\text{on}}+1)\prod_{i=1}^{M}(\tau_i+1)-1}{\tau_{\text{on}}\prod_{i=1}^{M}\tau_i}. \qquad (6.19)\end{aligned}$$

Substituting these expressions in terms of waiting times (Eqs (6.17) and (6.19)) into the expressions for the mean and variance obtained above (Eqs (6.14) and (6.15)), we have

$$\begin{aligned}\mathbb{E}(N) &= \frac{\mu\prod_{i=1}^{M}k_i}{(\tau_{\text{on}}+\tau_{\text{off}})k_{\text{on}}\prod_{i=1}^{M}k_i} \qquad (6.20)\\ &= \frac{\mu\,\tau_{\text{on}}}{\tau_{\text{on}}+\tau_{\text{off}}},\end{aligned}$$

$$\text{Var}(N) = \mathbb{E}(N)\left(\frac{\mu\,\tau_{\text{on}}\prod_{i=1}^{M}(\tau_i+1)}{(\tau_{\text{on}}+1)\prod_{i=1}^{M}(\tau_i+1)-1}+1-\mathbb{E}(N)\right). \qquad (6.21)$$

Notice that the expression for $\mathbb{E}(N)$ is the same as for the random telegraph model; it is the expectation for the constitutive (Poisson) model with only one state, μ, scaled by the proportion of time the promoter is active. In particular, the expectation is independent of the individual expected waiting times τ_i in each inactive state, although the variance is not. Hence we may fix the waiting times τ_{on} and $\tau_{\text{off}}=\sum\tau_i$, thus fixing the expectation, whilst tuning the noise by varying the expected waiting time in each inactive state τ_i.

Let us investigate in more detail how noise is affected by the choice of $\tau_i, i=1,\dots,M$, and by M, the number of inactive states. We will use the ensemble Fano factor as a measure for noise, although we would arrive at the same conclusions if we chose the coefficient of variation, $\eta \coloneqq \sqrt{\text{Var(N)}}/\mathbb{E}(N)$ instead, since we have fixed $\mathbb{E}(N)=\tau_{\text{on}}/(\tau_{\text{on}}+\tau_{\text{off}})$. Since the number of inactive states will now be a variable, we will specify it using a superscript,

i.e., $\text{Fano}^{(2)}$ refers to the Fano factor for the 2-cycle. The waiting times τ_i are also variables, but it should not cause confusion if we do not specify them explicitly.

Suppose τ_{on} and τ_{off} are fixed constants, but the $\tau_i, i = 1, \ldots, M$, can vary under the constraints

$$\tau_i > 0,\ i = 1, \ldots, M, \quad \text{and} \quad \sum_{i=1}^{M} \tau_i = \tau_{\text{off}}.$$

For now, the number of inactive states M is fixed. Using Eqs (6.20) and (6.21), we have

$$\begin{aligned}\text{Fano}^{(M)} &:= \frac{\text{Var}(N)}{\mathbb{E}(N)} \\ &= \frac{\mu\,\tau_{\text{on}} \prod_{i=1}^{M}(\tau_i + 1)}{(\tau_{\text{on}} + 1)\prod_{i=1}^{M}(\tau_i + 1) - 1} + 1 - \mathbb{E}(N),\end{aligned}$$

which we can minimize over the τ_i's using for example the method of Lagrange multipliers. The minimum $\text{Fano}_{\text{min}}^{(M)}$ occurs when $\tau_i = \tau_j = \tau_{\text{off}}/M\ \forall\, i, j \in \{1, \ldots, M\}$, giving

$$\text{Fano}^{(M)} \geq \text{Fano}_{\text{min}}^{(M)} = \frac{\mu\,\tau_{\text{on}}(\tau_{\text{off}}/M + 1)^M}{(\tau_{\text{on}} + 1)(\tau_{\text{off}}/M + 1)^M - 1} + 1 - \mathbb{E}(N).$$

On the other hand, the maximum possible Fano factor occurs at the other extreme of τ_i-space, where without loss of generality $\tau_1 \to \tau_{\text{off}}$, and $\tau_i \to 0$, $i = 2, \ldots, M$. Notice that in the limit $\tau_i = 0$, $i = 2, \ldots, M$, the M-cycle is reduced to the 1-cycle, i.e. the usual two-state random telegraph model, so we have:

$$\begin{aligned}\text{Fano}^{(M)} \leq \text{Fano}_{\text{max}}^{(M)} &< \text{Fano}^{(1)} \\ &= \frac{\mu\,\tau_{\text{on}}(\tau_{\text{off}} + 1)}{(\tau_{\text{on}} + 1)(\tau_{\text{off}} + 1) - 1} + 1 - \mathbb{E}(N) \\ &= \frac{\mu\,\tau_{\text{on}}\tau_{\text{off}}^2}{(\tau_{\text{on}}\tau_{\text{off}} + \tau_{\text{on}} + \tau_{\text{off}})(\tau_{\text{on}} + \tau_{\text{off}})} + 1.\end{aligned}$$

Now to see how the number of states M affects the noise characteristics of the model, we now allow M to be a variable. Consider first the expression above for $\text{Fano}_{\text{min}}^{(M)}$, the minimum Fano factor given M. The function

$$g(M) := \frac{\mu\tau_{\text{on}}(\tau_{\text{off}}/M + 1)^M}{(\tau_{\text{on}} + 1)(\tau_{\text{off}}/M + 1)^M - 1}, M \in \mathbb{N}$$

is a monotonically decreasing function of M, hence (since $\mathbb{E}(N)$ is fixed) the minimum possible Fano factor for an M-cycle decreases as the number of inactive states M increases:

$$\text{Fano}_{\text{min}}^{(L)} < \text{Fano}_{\text{min}}^{(M)}\ \forall\, L > M.$$

Since any M-cycle reduces to the 1-cycle in the limit $\tau_1 = \tau_{\text{off}},\ \tau_i = 0,\ i = 2, \ldots, M$, we have

$$\sup\left(\text{Fano}_{\text{max}}^{(M)}\right) = \text{Fano}^{(1)}.$$

To sum up, for an M-cycle model with expected waiting times in the inactive states τ_i, $i = 1, \ldots, M$, the Fano factor $\text{Fano}^{(M)}$ satisfies

$$\text{Fano}^{(L)}_{\min} < \text{Fano}^{(M)}_{\min} \leq \text{Fano}^{(M)} < \text{Fano}^{(1)}_{\min} \equiv \text{Fano}^{(1)}_{\max} \equiv \text{Fano}^{(1)} \;\forall\; L > M > 1.$$

In other words, assuming the simple two-state random telegraph model of gene transcription necessarily implies assuming the maximum level of noise for the given expected waiting times. It is well-known that negative feedback motifs can reduce gene expression noise (Austin *et al.* 2006; Nevozhay *et al.* 2009; Shimoga *et al.* 2013), but here we have shown that we may achieve similar effects via the introduction of additional gene states, and/or tuning the transition rates towards the parameter regime $\tau_i = \tau_{\text{off}}/M,\ i = 1, \ldots, M$.

6.3 Estimating parameter values and model selection

In 2006 Raj *et al.* published their seminal paper where for the first time the exact expression for $P(n)$ for the 1-cycle (random telegraph model) was known, and could be used to fit to their fluorescence in situ hybridization (FISH) data (Raj *et al.* 2006). More recently, several different 1-cycle parameter regimes have been shown to fit snapshot data for yeast equally well (Zenklusen *et al.* 2008; Senecal *et al.* 2014), and several groups have shown that waiting time distributions in the inactive state are not exponentially distributed (which the 1-cycle can not account for) (Suter *et al.* 2011; Zoller *et al.* 2015; Harper *et al.* 2011). Nevertheless, in the absence of an exact, practical[3] solution, it remains common practice to estimate transcription rates and timescales at the promoter by fitting snapshot data to the solution of the 1-cycle, or by using expected waiting times τ_{on} and τ_{off} to estimate the 1-cycle transition rates $k_{\text{on}} = 1/\tau_{\text{off}}$ and $k_{\text{off}} = 1/\tau_{\text{on}}$ (Suter *et al.* 2011; Senecal *et al.* 2014).

Now that we have the exact solution explicitly in terms of measurable parameters, we can use it to estimate M-cycle parameter values for snapshot data, and perform quantitative model selection. As an example, we generated in-silico snapshot data for 5-cycle models designed to mimic the results published in (Zoller *et al.* 2015), where M-cycles are fitted to time-lapse data. We then used maximum likelihood estimation (MLE) to estimate the parameter values under the assumptions of several M-cycle models, and the Akaike Information Criterion (AIC) to perform model selection. We use these results:

- to exemplify how the assumption of the minimal 1-cycle (random telegraph) model can lead to gross mis-estimation of the underlying parameters; and
- to demonstrate that for many parameter regimes, snapshot population data cannot be used to infer the underlying timescales. In those cases single-cell time series analysis is necessary to validate the results.

[3] Although an exact solution for the M-cycle was published in 2012 (Zhang *et al.* 2012), it was not given explicitly in terms of the model parameters $\mu, k_1, \ldots, k_M, k_{\text{on}}$ as we were able to do here.

However, a thorough parameter fitting exercise would require more detailed investigations into which measures should be used to measure disparity between two probability distributions.

6.3.1 Generating *in-silico* data

To be able to critically analyse the output of the MLE, we generated several sets of *in-silico* data using Gillespie's stochastic simulation algorithm (SSA) (Gillespie 1977). The expression for $P(n)$ (Eq. (6.9)) was derived under assumptions that imply an ergodic system, so to obtain "snapshot" data it is sufficient to simulate a single time trace $\{\eta(t)\}_{0\leq t\leq T}$ for the number of mRNA transcripts in a single cell[4]. We conservatively discarded the first 1000 minutes after the sample path first reached its expected mean value to avoid effects from the initial condition. Then, what we refer to as the *empirical* probability distribution P_{emp} was given by

$$P_{\text{emp}}(n) = \frac{1}{T}\int_0^T \mathbb{1}_n(\eta(t))dt,$$

where $\mathbb{1}_n$ is the indicator function:

$$\mathbb{1}_n(\eta(t)) = \begin{cases} 1 & \text{if } \eta(t) = n, \\ 0 & \text{otherwise.} \end{cases}$$

We sampled 200 data points from the empirical distribution to use in the maximum likelihood estimation procedure.

To work only within biologically relevant parameter regimes, we used the models and parameter values for the NcKap1 and Hmga2 genes, which were recently estimated from time-lapse recordings of mammalian cells (Zoller *et al.* 2015).

6.3.2 Maximum Likelihood estimation and model selection

Despite the expression for $P(n)$ (6.9) being a concise function of relatively few parameter combinations, the gamma functions and generalized hypergeometric function lead to unwieldy likelihood and log-likelihood functions that are impractical to maximise analytically. As such, we used the 'Squeeze-and-breathe' numerical optimization algorithm to maximise the log-likelihood function, since it has been shown to be robust to situations where parameter values may span several orders of magnitude, and can find parameter regimes that lie outside of the range of the initial prior (Beguerisse-Díaz *et al.* 2012). The method iteratively refines a sequence of parameter distributions using local optimisation and partial resampling of the historical prior, followed by ranking and culling of the local optima. The iteration continues until the parameter distributions and the negative log-likelihood values converge.

The results of the parameter estimation process for the NcKap1 and Hmga2 genes

[4] As per the discussions in Sections 4.2.2 and Example 5.1–1, an estimate for $P(n)$ could have been obtained quicker using sample paths of $\mathcal{X}$ instead of $\mathcal{N}$. However, the simulation process here was not prohibitively slow because only one sample path was required.

NcKap1: Fitting to the 5-cycle			NcKap1: Fitting to the 1-cycle		
Parameter	True	Estimated	Parameter	Effective	Estimated
μ/λ	344.83	233.01	μ/λ	344.83	65.36
k_{on}/λ	57.47	43.94	k_{on}/λ	57.47	14.35
k_1/λ	3.05	3.06	k_{off}/λ, k_1/λ	1.61	2.32
k_2/λ	8.62	11.20			
k_3/λ	12.54	15.62			
k_4/λ	16.42	20.45			
k_5/λ	27.59	34.28			
$\ln \mathcal{L}$	-621.93	-621.41	$\ln \mathcal{L}$	-621.46	-620.86
AIC	1257.85	1256.82	AIC	1256.91	1247.73

(a) Results of the maximum likelihood estimation procedure for the NcKap1 *in silico* data

Hmga2: Fitting to the 5-cycle			Hmga2: Fitting to the 1-cycle		
Parameter	True	Estimated	Parameter	Effective	Estimated
μ/λ	51.39	37.39	μ/λ	51.39	31.84
k_{on}/λ	42.09	32.00	k_{on}/λ	42.09	41.42
k_1/λ	8.17	6.75	$k_1/\lambda, k_{\text{off}}/\lambda$	2.52	4.07
k_2/λ	9.99	11.50			
k_3/λ	12.29	14.87			
k_4/λ	16.53	21.19			
k_5/λ	30.19	33.96			
$\ln \mathcal{L}$	-412.51	-412.43	$\ln \mathcal{L}$	-414.67	-412.40
AIC	839.02	838.86	AIC	837.35	830.80

(b) Results of the maximum likelihood estimation procedure for the Hmga2 *in silico* data

Table 6.1: Results of the maximum likelihood estimation procedure for the *in silico* data, which were generated using 5-cycle models. The parameter values used for the simulations are shown in the columns entitled "True" on the left hand side. $\ln \mathcal{L}$ and AIC are the log-likelihood values. *Left:* The estimated parameter values obtained when fitting to the 5-cycle, and for comparison, the true values used to generate the data. *Right:* The estimated parameter values obtained when fitting to the 1-cycle. To enable comparison with the true values, $k_{\text{off}}/\lambda = 1/\lambda\tau_{\text{off}}$ is given in the column with the true parameter values, where τ_{off} is the total expected waiting time in the inactive states for the true model.

are shown in Table 6.1, and Figs 6.2 and 6.3. All parameters are reported in ratio to the degradation rate. k_{off} is calculated using the true parameter values according to $k_{\text{off}}^{-1} = \sum_{i=1}^{5} k_i^{-1}$, purely to enable comparisons for the estimated parameter values for the 1-cycle; it was not used in the estimation process. In other words, the "effective" parameter values quoted in the tables for the 1-cycle are the parameters one obtains when converting the 5-cycle into a 1-cycle by conserving expected waiting times in the active and inactive states. We refer to this "converted" 1-cycle as the *1-cycle equivalent* in the

figures showing probability mass functions (pmfs) and cumulative mass functions (cmfs) for the 1-cycles (bottom rows), again solely for the purposes of comparison.

Using the log-likelihood values, we can use the AIC to select the preferred model for the data we used. The AIC is a measure for statistical model selection based on information theory (Akaike 1974), defined by

$$\text{AIC} := -2\ln\mathcal{L} + 2(\text{number of independently adjusted parameters within the model}).$$

The AIC values are shown in Tables 6.1; the preferred model is the one with lowest AIC value.

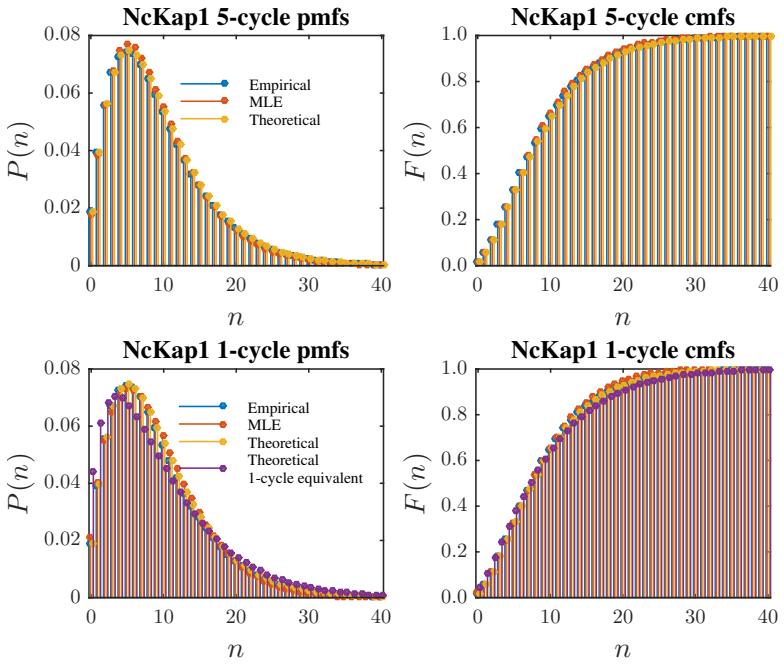

Figure 6.2: The probability mass functions $P(n)$ and cumulative mass functions $F(n)$ for the NcKap1 empirical data (blue) and the theoretical mass functions, calculated using: the parameter values from the maximum likelihood estimation (red); true parameter values (yellow); and the parameter values for the 1-cycle with the same expected waiting times in each gene state as the true model (purple). The legends apply to both the pmf and cmf graphs.

There are several observations to be made from these maximum likelihood estimates and corresponding AIC values. First, the 1-cycle maximum likelihood parameter estimates do better in terms of the log-likelihood $\ln\mathcal{L}$ than both the true values, and the estimates obtained when fitting the 5-cycle (the true model). Second, since the log-likelihood values for the fitted models are very similar, selection using the AIC amounts to choosing the model with the fewest parameters, in this case the 1-cycle. However, the 1-cycle parameter values are grossly mis-estimated. Although the 5-cycle estimates are incorrect, they are significantly closer to the true values than the 1-cycle estimates are, in particular for μ/λ.

Third, as can be clearly seen from the figures, we recover the observations from Zen-

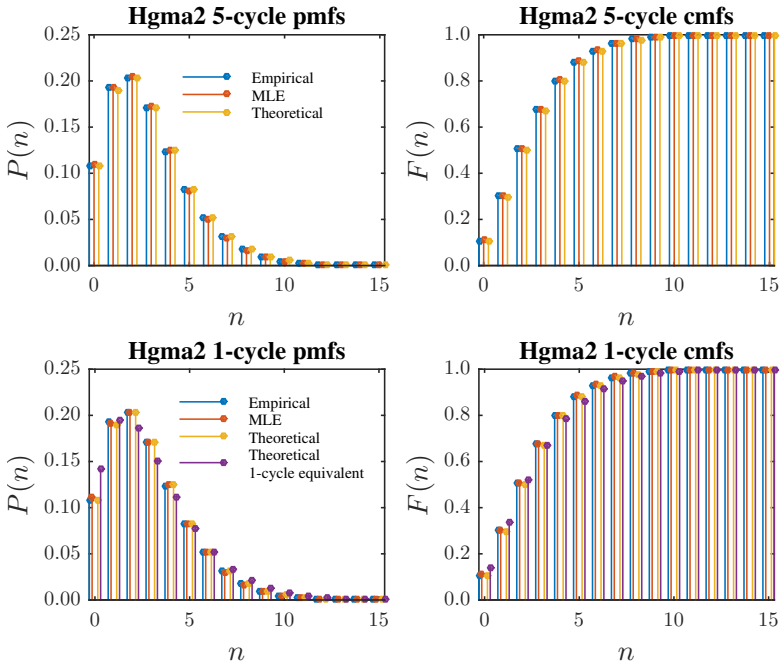

Figure 6.3: The probability mass functions $P(n)$ and cumulative mass functions $F(n)$ for the Hmga2 empirical data (blue) and the theoretical mass functions, calculated using: the parameter values from the maximum likelihood estimation (red); true parameter values (yellow); and the parameter values for the 1-cycle with the same expected waiting times in each gene state as the true model (purple). The legends apply to both the pmf and cmf graphs.

klusen *et al.* 2008 that these models are unidentifiable – several different models can fit population snapshot data equally well. The estimated parameter values outperform the true parameter values when maximising the log-likelihood function, suggesting that the log likelihood function is extremely flat for these models, and perhaps not an effective measure for these models. Precision tolerances must be kept low when maximising the log-likelihood function numerically, and one typically needs to restrain the parameter space if possible (Murphy and Vaart 2000).

Taken together, these results issue a stark warning to those inferring timescales of activities at the promoter from snapshot data (Raj *et al.* 2006; Taniguchi *et al.* 2010; Senecal *et al.* 2014; Bahar Halpern *et al.* 2015), even with more complex models than the simple 1-cycle. In particular, good fits to the empirical distribution are not necessarily indicative of the best parameter estimates; the 1-cycle equivalent effective parameters produce the worst fits to the data (Figs 6.2 and 6.3), but the timescales implied by them are closer to reality than those implied by the MLE for the 1-cycle.

There are several functions apart from the likelihood or log-likelihood that can be used to measure the disparity between two discrete probability distributions, for example the Kullback-Leibler divergence or the related Jensen-Shannon divergence, the total variation distance, or the Wasserstein (or earth mover's) distance, to name just a few. Each divergence or metric has a bias towards certain kinds of differences between two distributions,

for example the Kullback-Leibler divergence emphasizes tails more strongly than the total variation distance. A thorough parameter-fitting exercise would need to check the results obtained from several appropriate divergences or metrics, possibly defining a new measure if a particular property of the distribution needed to be given particular importance. Validation using single-cell time series analysis is in general necessary to confirm the reliability of the results (Zoller *et al.* 2015; Suter *et al.* 2011).

6.4 Solution for f_X

Recall from Chapter 3 that the probability mass function $P(n)$ for the mRNA copy number can be written in Poisson mixture form

$$P(n) = \int \frac{\xi^n e^{-\xi}}{n!} f_X(\xi)\, d\xi,$$

so that one route to obtaining $P(n)$ is simply to search for the mixing density f_X. Since we already have an explicit expression for $P(n)$ from the previous section, we will not need to use the Poisson mixture property to obtain a solution of the M-cycle. However, the value of f_X extends far beyond simply being a stepping stone towards an expression for $P(n)$. Since f_X is the central, model-specific core of any transcription-degradation model, we gain a deeper understanding of the mathematical structure and properties of the model than we do when only looking at $P(n)$. After deriving f_X using the multistate Kramers-Moyal equations, in Section 6.4.2 we will use it to help navigate parameter space to search for qualitatively different solution behaviours. We defer discussion of the mathematical structures that reveal themselves when deriving f_X until Chapter 7, as the scope extends beyond M-cycle models.

6.4.1 Derivation of the solution

The M-cycle is a multistate promoter model that falls within the framework described in Section 4.4 on the multistate Kramers-Moyal equations: we have a Markov jump process S between $M+1$ discrete states $\{s_1, \ldots, s_M, s_{\text{on}}\}$, and the bivariate process $\{X, S\}$ is Markovian. Therefore we can straightforwardly write down the multistate Kramers-Moyal equations in the same way as we did for the solution of the leaky random telegraph model (Example 4.4–1).

Define for each state $\{s_1, \ldots, s_M, s_{\text{on}}\}$ the joint probability density function $f_i(x) \coloneqq f_{X,S}(x, s_i)$, and recall the convention introduced in Section 6.1 that indices should be considered modulo $M+1$, with $M+1 \equiv 0 \equiv \text{on}$. The law of total probability gives us

$$f_X(x) = \sum_{i=1}^{M+1} f_i(x),$$

and the multistate Kramers-Moyal equations are (Pawula 1967; Pawula 1970):

$$\frac{d}{dx}\left[-\lambda x f_i(x)\right] = k_{i-1} f_{X,S}(x, s_{i-1}) - k_i f_i(x), \quad i = 1, \ldots, M, \quad \text{and} \tag{6.22a}$$

$$\frac{d}{dx}\left[(\mu - \lambda x) f_{\text{on}}(x)\right] = k_M f_M(x) - k_{\text{on}} f_{\text{on}}(x). \tag{6.22b}$$

To simplify notation, set $\lambda = 1$ and use the change of variables $z = \lambda x/\mu$. We will keep the same notation for the joint probability density functions, but specify the variable for the marginal, i.e. we will now work to obtain $f_Z(z) = f_1(z) + \cdots + f_M(z) + f_{\text{on}}(z)z \in (0, 1)$. Eqs (6.22) become

$$\frac{d}{dz}\left[A(z)\mathbf{f}(z)\right] = K\mathbf{f}(z), \tag{6.23}$$

where $\mathbf{f} = (f_1, \ldots, f_M, f_{\text{on}})^T$, $A(z) = (\mathbf{e}_{M+1}\mathbf{e}_{M+1}^T - zI)$, $\mathbf{e}_{M+1} = (0, \ldots, 0, 1)^T$, and K is the state transition matrix

$$K = \begin{pmatrix} -k_1 & & & & k_{\text{on}} \\ k_1 & -k_2 & & & \\ & k_2 & \ddots & & \\ & & \ddots & -k_M & \\ & & & k_M & -k_{\text{on}} \end{pmatrix}$$

As we saw in Section 6.2.1, K has a zero eigenvalue and by Gershgorin's circle theorem the non-zero eigenvalues of K have negative real parts, so a stationary solution exists. The probabilities $\pi_i = \int_0^1 f_i(z)\, dz$ of being in state s_i, $i = 1, \ldots, s_M, s_{\text{on}}$, evolve to an equilibrium state given by the eigenvector $\boldsymbol{\pi}$ that corresponds to the zero eigenvalue, where

$$\pi_i = \int_0^1 f_i(z)\, dz = \frac{\prod_{j \neq i} k_j}{\sum_{r=1}^{M+1} \prod_{j \neq r} k_j}. \tag{6.24}$$

Now, integrating (6.23) we have

$$\left[A(z)\mathbf{f}(z)\right]_0^1 = K \int_0^1 \mathbf{f}(z)\, dz = K\boldsymbol{\pi} = \mathbf{0},$$

hence we get the boundary conditions

$$f_i(1) = 0, \quad i = 1, \ldots, M, \quad \text{and} \tag{6.25a}$$

$$f_{\text{on}}(0) = 0. \tag{6.25b}$$

Also, for $\mathbf{1}^T := (1, \ldots, 1)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\mathbf{1}^T A(z)\mathbf{f}\right) = \mathbf{1}^T K\mathbf{f} = 0$$

(equivalent to summing the rows), so that

$$\mathbf{1}^T A(z)\mathbf{f} = -zf_1(z) - \cdots - zf_M(z) + (1-z)f_{\text{on}}(z) = C. \tag{6.26}$$

Equation (6.26) is true for all $z \in (0,1)$, so substituting in $z = 0$ or $z = 1$ we can show that the constant of integration C is equal to 0. Hence

$$f_Z = \sum_{i=1}^{M+1} f_i = \frac{1}{z} f_{\text{on}}, \tag{6.27}$$

so we need only solve the Kramers-Moyal equations (6.23) for f_{on}, the marginal probability density corresponding to the active state. In order to obtain an uncoupled equation for f_{on}, we will employ the same trick we used in Section 6.2.1, using an adjugate matrix and capitalise on the structure of Eq. (6.23).

Define $\theta := z\frac{d}{dz}$ and write the Kramers-Moyal equations (6.23) in the form

$$(\theta I + I + K)\,\mathbf{f} = \mathbf{e}_{M+1}\mathbf{e}_{M+1}^T \frac{1}{z}\,\theta\,\mathbf{f}.$$

Multiply both sides on the left by the transposed matrix of cofactors $\text{adj}(\theta I + I + K)$, remembering that for any square invertible matrix A, $\text{adj}(A)\,A = \det(A)I$:

$$\begin{aligned}\det(\theta I + I + K)\,\mathbf{f} &= \text{adj}(\theta I + I + K)\,\mathbf{e}_{M+1}\mathbf{e}_{M+1}^T \frac{1}{z}\,\theta\mathbf{f} \\ &= \begin{bmatrix} (-1)^{M+2}\mathrm{M}_{M+1,1} \\ \vdots \\ (-1)^{2M+2}\mathrm{M}_{M+1,M+1} \end{bmatrix} \frac{1}{z}\,\theta\, f_{\text{on}},\end{aligned}$$

where $\mathrm{M}_{i,j}$ is the $(i,j)^{\text{th}}$ minor of $[\theta I + I + K]$.

Notice that the submatrix of $[\theta I + I + K]$ formed by deleting the M^{th} row and the M^{th} column is lower triangular, so

$$\mathrm{M}_{M+1,M+1} = \prod_{i=1}^{M}(\theta + 1 - k_i).$$

Also, the determinant is equal to the product of the eigenvalues and hence

$$\det(\theta I + I + K) = \prod_{i=1}^{M+1}(\theta + 1 + \nu_i),$$

where the ν_i are the eigenvalues of the matrix K. The M^{th} coordinate of this equation is now an uncoupled equation in terms of f_{on} only:

$$\prod_{i=1}^{M+1}(\theta + 1 + \nu_i)\, f_{\text{on}}(z) = \prod_{i=1}^{M}(\theta + 1 - k_i)\,\frac{1}{z}\,\theta\, f_{\text{on}}(z)$$

$$= \frac{1}{z} \prod_{i=1}^{M} (\theta - k_i)\, \theta\, f_{\mathrm{on}}(z), \tag{6.28}$$

where we used the operator identity $(\theta + a)\frac{1}{z} = \frac{1}{z}(\theta + a - 1)$. This is a Fuchsian[5] equation of order $M + 1$ with singularities at 0, 1, and $+\infty$, so we can write down the solution of this equation directly in terms of generalised hypergeometric functions (Slater 1966). When the k_i are distinct modulo 1, the general solution is given by

$$\begin{aligned} f_{\mathrm{on}}(z) &= c_0\, z^0\, {}_{M+1}F_M \left[\begin{matrix} 1+\nu_1, & \ldots\ldots, & 1+\nu_M, & 1 \\ 1-k_1, & ..\vee.., & 1-k_M & \end{matrix}\; ;\; z\right] \\ &\quad + \sum_{i=1}^{M} c_i\, z^{k_i}\, {}_{M+1}F_M \left[\begin{matrix} 1+\nu_1+k_i, & \ldots\ldots, & 1+\nu_M+k_i, & 1+k_i \\ 1-k_1+k_i, & ..\vee.., & 1-k_M+k_i, & 1+k_i \end{matrix}\; ;\; z\right] \qquad (6.29a) \\ &\equiv c_0\, {}_{M+1}F_M \left[\begin{matrix} 1+\nu_1, & \ldots\ldots, & 1+\nu_M, & 1 \\ 1-k_1, & ..\vee.., & 1-k_M & \end{matrix}\; ;\; z\right] \\ &\quad + \sum_{i=1}^{M} c_i\, z^{k_i}\, {}_{M}F_{M-1} \left[\begin{matrix} 1+\nu_1+k_i, & \ldots\ldots, & 1+\nu_M+k_i \\ 1-k_1+k_i, & ..\vee.., & 1-k_M+k_i \end{matrix}\; ;\; z\right], \qquad (6.29b) \end{aligned}$$

where without loss of generality we have taken ν_{M+1} to be the zero eigenvalue of K, and $c_0, \ldots, c_M$ are constants. Here $..\vee..$ denotes suppression of the term $1 - k_i + k_i$, and ${}_{M+1}F_M$ and ${}_{M}F_{M-1}$ are generalised hypergeometric functions (Erdelyi 1953).

At this point, we will adopt the contracted notation introduced in (Slater 1966), as it will greatly improve clarity later on. For generalized hypergeometric functions we will write

$${}_pF_q \left[\begin{matrix} a_1, & \ldots, & a_p \\ b_1, & \ldots, & b_q \end{matrix}\; ;\; z\right] =: {}_pF_q \left[\begin{matrix} (a) \\ (b) \end{matrix}\; ;\; z\right],$$

for a product of several Gamma functions we will write

$$\begin{aligned} \frac{\Gamma(a_1)\Gamma(a_2)\ldots\Gamma(a_p)}{\Gamma(b_1)\Gamma(b_2)\ldots\Gamma(b_q)} &=: \Gamma \left[\begin{matrix} a_1, a_2, \ldots, a_p \\ b_1, b_2, \ldots, b_q \end{matrix}\right] \\ &=: \Gamma \left[\begin{matrix} (a) \\ (b) \end{matrix}\right], \end{aligned}$$

and a dash will denote the omission of a zero factor in a sequence, for example

$$(a)' - a_i := a_1 - a_i, \ldots, a_{i-1} - a_i, a_{i+1} - a_i, \ldots, a_p - a_i.$$

The general solution (6.29b) can then be written

$$f_{\mathrm{on}}(z) = c_0\, {}_{M}F_{M-1} \left[\begin{matrix} 1+(\nu) \\ 1-(k)' \end{matrix}\; ;\; z\right] + \sum_{i=1}^{M} c_i\, z^{k_i}\, {}_{M}F_{M-1} \left[\begin{matrix} 1+(\nu)+k_i \\ 1-(k)'+k_i \end{matrix}\; ;\; z\right], \tag{6.30}$$

[5] Fuchsian equations will be defined and discussed in Chapter 7.

where (ν) denotes the sequence $\nu_1, \ldots, \nu_M$ (we omit $\nu_{M+1} = 0$), and (k) denotes the sequence $k_1, \ldots, k_M$.

Now, recall that $f_Z(z) = f_{\text{on}}(z)/z$, and the boundary condition (6.25b) implies that $c_0 = 0$, so finally we have the general solution

$$f_Z(z) = \sum_{i=1}^{M} c_i \, z^{k_i - 1} \, {}_M F_{M-1} \left[\begin{matrix} 1 + (\nu) + k_i \\ 1 - (k)' + k_i \end{matrix} \; ; \; z \right]. \tag{6.31}$$

The constants c_i can be determined using the integral constraints $\int_0^1 f_i(z)\, dz = \pi_i$ and $\int_0^1 f_Z(z)\, dz = 1$, and the identity (Rainville 1960)

$$\int_0^1 z^{k-1} \, {}_p F_{p-1} \left[\begin{matrix} (a) \\ (b) \end{matrix} \; ; \; z \right] dz = \frac{1}{k} \, {}_{p+1} F_p \left[\begin{matrix} (a), k \\ (b), k+1 \end{matrix} \; ; \; 1 \right].$$

However, using the solution for $P(n)$ derived in Section 6.2.1 and the Poisson mixture result, we can do better. First, use our general solution for f_Z (6.31) to write down an expression for $P(n)$ using the Poisson mixture result:

$$\begin{aligned} P(n) &= \int_0^1 \frac{(\mu z)^n}{n!} \, e^{-\mu z} \sum_{i=1}^{M} c_i \, z^{k_i - 1} \, {}_M F_{M-1} \left[\begin{matrix} 1 + (\nu) + k_i \\ 1 - (k)' + k_i \end{matrix} \; ; \; z \right] dz \\ &= \frac{\mu^n}{n!} \sum_{i=1}^{M} c_i \int_0^1 e^{-\mu z} \, z^{k_i - 1 + n} \, {}_M F_{M-1} \left[\begin{matrix} 1 + (\nu) + k_i \\ 1 - (k)' + k_i \end{matrix} \; ; \; z \right] dz. \end{aligned} \tag{6.32}$$

Now, identity 4.8.3.13 from (Slater 1966) is (with letters and symbols as found in the reference):

$$\begin{aligned} &\Gamma \left[\begin{matrix} (a) + s \\ (c) + s \end{matrix} \right] \, {}_{A+E} F_{A+F} \left[\begin{matrix} (a) + s, & (e) \\ (c) + s, & (f) \end{matrix} \; ; \; y \right] \\ &\quad = \sum_{\mu=1}^{A} \Gamma \left[\begin{matrix} (a) - a_\mu \\ (c) - a_\mu \end{matrix} \right] \int_0^1 x^{s + a_\mu - 1} \, {}_A F_{A-1} \left[\begin{matrix} 1 + a_\mu - (c) \\ 1 + a_\mu - (a)' \end{matrix} \; ; \; x \right] \, {}_E F_F \left[\begin{matrix} (e) \\ (f) \end{matrix} \; ; \; xy \right] dx, \end{aligned}$$

for $E < F$ or for $E = F$ and $|y| < 1$. Use this identity with $E = F = 0$, $(e) = (f) = (0)$, $(a) = (k)$, $(c) = (-\nu)$, and $y = -\mu$ to write our explicit solution for $P(n)$ (6.9) in the form

$$\begin{aligned} P(n) &= \Gamma \left[\begin{matrix} (k) + n, (-\nu) \\ (k), (-\nu) + n \end{matrix} \right] \, \frac{\mu^n}{n!} \, {}_M F_M \left[\begin{matrix} (k) + n \\ (-\nu) + n \end{matrix} \; ; \; -\mu \right] \\ &= \frac{\mu^n}{n!} \sum_{i=1}^{M} \Gamma \left[\begin{matrix} (-\nu) \\ (k) \end{matrix} \right] \Gamma \left[\begin{matrix} (k)' - k_i \\ (-\nu) - k_i \end{matrix} \right] \int_0^1 e^{-\mu z} \, z^{k_i - 1 + n} \, {}_M F_{M-1} \left[\begin{matrix} 1 - (-\nu) + k_i \\ 1 - (k)' + k_i \end{matrix} \; ; \; z \right] dz. \end{aligned} \tag{6.33}$$

Finally, compare Eqs (6.32) and (6.33) to find

$$c_i = \Gamma \begin{bmatrix} (-\nu) \\ (k) \end{bmatrix} \Gamma \begin{bmatrix} (k)' - k_i \\ (-\nu) - k_i \end{bmatrix}.$$

Putting everything together, we finally obtain the solution

$$f_Z(z) = \Gamma \begin{bmatrix} (-\nu) \\ (k) \end{bmatrix} \sum_{i=1}^{M} \Gamma \begin{bmatrix} (k)' - k_i \\ (-\nu) - k_i \end{bmatrix} z^{k_i - 1} \, {}_M F_{M-1} \begin{bmatrix} 1 + (\nu) + k_i \\ 1 - (k)' + k_i \end{bmatrix} ; \; z \Bigg], \tag{6.34}$$

for $z \in (0, 1)$. We can easily obtain f_X by changing variables back to $X = \mu Z / \lambda$:

$$f_X(x) = \frac{\lambda}{\mu} f_Z \left(\frac{\lambda}{\mu} z \right) \tag{6.35}$$

$$= \frac{\lambda}{\mu} \Gamma \begin{bmatrix} (-\nu) \\ (k) \end{bmatrix} \sum_{i=1}^{M} \Gamma \begin{bmatrix} (k)' - k_i \\ (-\nu) - k_i \end{bmatrix} \left(\frac{\lambda}{\mu} \right)^{k_i - 1} x^{k_i - 1} \, {}_M F_{M-1} \begin{bmatrix} 1 + (\nu) + k_i \\ 1 - (k)' + k_i \end{bmatrix} ; \; \frac{\lambda}{\mu} x \Bigg], \tag{6.36}$$

for $x \in (0, \mu/\lambda)$.

Example 6.4–1 — Rederivation of the solution of the random telegraph model (1-cycle)

As for the explicit solution for $P(n)$ (6.9), we can easily check that our results reproduce the known solution of the random telegraph model (Ko 1991; Peccoud and Ycart 1995; Raj *et al.* 2006; Iyer-Biswas *et al.* 2009). For $M = 1$, $\lambda = 1$ we have eigenvalues $\nu_1 = -(k_{\text{on}} + k_1)$ and $\nu_2 = 0$. Eq. (6.34) then becomes

$$\begin{aligned} f_Z(z) &= \frac{\Gamma(k_{\text{on}} + k_1)}{\Gamma(k_1)} \frac{1}{\Gamma(k_{\text{on}} + k_1 - k_1)} z^{k_1 - 1} \, {}_1 F_0 \left[1 - k_{\text{on}} - k_1 + k_1 \; ; \; z \right] \\ &= \frac{\Gamma(k_{\text{on}} + k_1)}{\Gamma(k_1)\Gamma(k_{\text{on}})} z^{k_1 - 1} (1 - z)^{k_{\text{on}} - 1}. \end{aligned}$$

In other words, $Z \sim \text{Beta}(k_1, k_{\text{on}})$, so the final solution is a Poisson-Beta mixture:

$$\begin{aligned} P(n) &= \int_0^1 \frac{(\mu z)^n}{n!} e^{-\mu z} \frac{\Gamma(k_1 + k_2)}{\Gamma(k_1)\Gamma(k_2)} z^{k_1 - 1} (1 - z)^{k_2 - 1} \, dz \\ &= \frac{\Gamma(k_1 + n)}{\Gamma(k_1)} \frac{\Gamma(k_{\text{on}} + k_1)}{\Gamma(k_{\text{on}} + k_1 + n)} \frac{\mu^n}{\Gamma(n + 1)} \, {}_1 F_1 \left(k_1 + n, k_{\text{on}} + k_1 + n; -\mu \right), \end{aligned}$$

as shown in (Raj *et al.* 2006; Iyer-Biswas *et al.* 2009).

Example 6.4–2 — Mixture distribution for the 2-cycle

For $M = 2$, $(\nu) = \nu_1, \nu_2$ are the non-zero eigenvalues of K, which are

$$\nu_{1,2} = \frac{1}{2}\left(-k_{\text{on}} - k_1 - k_2 \pm \sqrt{(k_{\text{on}} - k_1 - k_2)^2 - 4k_1k_2}\right),$$

so the stationary solution of the mixing density is

$$\begin{aligned} f_Z(z) &= \frac{\Gamma(-\nu_1)\Gamma(-\nu_2)}{\Gamma(k_1)\Gamma(k_2)}\left(\frac{\Gamma(k_2-k_1)\,z^{k_1-1}}{\Gamma(-\nu_1-k_1)\Gamma(-\nu_2-k_1)}\,{}_2F_1\left[\begin{matrix}1+\nu_1+k_1, 1+\nu_2+k_1\\ 1-k_2+k_1\end{matrix}\,;\,z\right]\right. \\ &\quad\left. + \frac{\Gamma(k_1-k_2)\,z^{k_2-1}}{\Gamma(-\nu_1-k_2)\Gamma(-\nu_2-k_2)}\,{}_2F_1\left[\begin{matrix}1+\nu_1+k_2, 1+\nu_2+k_2\\ 1-k_1+k_2\end{matrix}\,;\,z\right]\right) \\ &= \frac{\Gamma(-\nu_1)\Gamma(-\nu_2)}{\Gamma(k_1)\Gamma(k_2)}\left(\frac{\Gamma(k_2-k_1)\,z^{k_1-1}(1-z)^{k_{\text{on}}-1}}{\Gamma(-\nu_1-k_1)\Gamma(-\nu_2-k_1)}\,{}_2F_1\left[\begin{matrix}-\nu_1-k_2, -\nu_2-k_2\\ 1-k_2+k_1\end{matrix}\,;\,z\right]\right. \\ &\quad\left. + \frac{\Gamma(k_1-k_2)\,z^{k_2-1}(1-z)^{k_{\text{on}}-1}}{\Gamma(-\nu_1-k_2)\Gamma(-\nu_2-k_2)}\,{}_2F_1\left[\begin{matrix}-\nu_1-k_1, -\nu_2-k_1\\ 1-k_1+k_2\end{matrix}\,;\,z\right]\right). \end{aligned}$$

6.4.2 Using the mixing density to analyse parameter spaces

We would like to be able to use the exact expression (6.35) for the mixing density to demarcate the parameter space into regions that produce qualitatively different solution behaviour for $P(n)$, for example to find bimodality. The usefulness of this technique has already been shown for a certain type of feedback model (Iyer-Biswas and Jayaprakash 2014).

For example, Zhang *et al.* 2012 searched for multimodality in the solution $P(n)$ of the 2-cycle via large-scale sampling of the system parameters; with our possession of f_Z we can deduce that multimodality does not exist[6] for any M-cycle: Recall from Eq. (6.28) that f_{on} is Fuchsian with only 2 finite singularities, and therefore so are f_Z and f_X. Since $P(n) = \mathbb{E}(X^n e^{-X}/n!)$, $P(n)$ can be bimodal, but not multimodal (see Chapter 7 for a more detailed discussion about what we can learn from the Fuchsian properties).

To see if we can deduce any more, for simplicity we consider the 2-cycle here. The same approach can be taken for $M > 2$, using the Euler-type identity for generalized hypergeometric functions given in (Miller and Paris 2011). Using the Euler identity to factorize $(1 - z)$ from the expression for f_Z (6.34), and writing the constant coefficients of

[6] We do not count the trivial cases where modes are adjacent to each other.

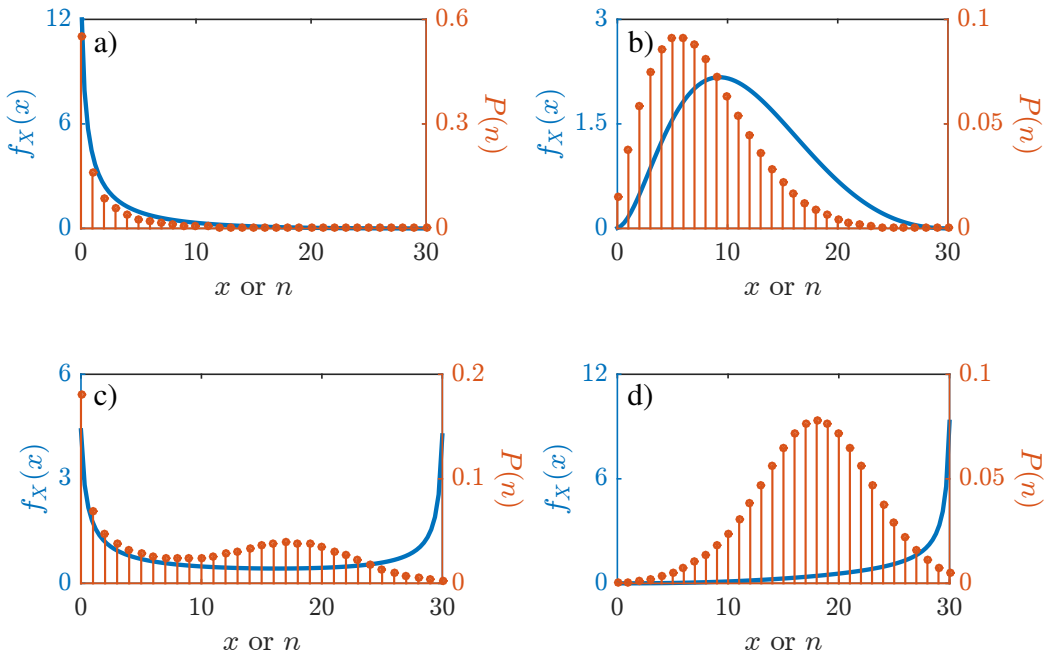

Figure 6.4: f_X (blue) and corresponding $P(n)$ (red) for four different qualitative behaviours of f_X. *(a)* f_X monotonically decreasing leads to $P(n)$ monotonically decreasing. *(b)* f_X concave downwards (peaked away from zero) leads to $P(n)$ concave downwards with peak more towards zero. *(c)* f_X concave upwards (bimodal) can lead to $P(n)$ bimodal. *(d)* f_X monotonically increasing leads to $P(n)$ concave downwards with peak more towards the upper end of the support of X_t.

each term as c_1 and c_2 to improve readability, we have

$$\begin{aligned} f_Z(z) &= (1-z)^{k_{\text{on}}-1}\left(c_1\, z^{k_1-1}\,{}_2F_1\left[\begin{matrix}-\nu_1-k_2,-\nu_2-k_2\\ 1-k_2+k_1\end{matrix};\, z\right]\right. \\ &\qquad\left. +c_2\, z^{k_2-1}\,{}_2F_1\left[\begin{matrix}-\nu_1-k_1,-\nu_2-k_1\\ 1-k_1+k_2\end{matrix};\, z\right]\right) \\ &= z^{k_1-1}(1-z)^{k_{\text{on}}-1}\left(c_1\,{}_2F_1\left[\begin{matrix}-\nu_1-k_2,-\nu_2-k_2\\ 1-k_2+k_1\end{matrix};\, z\right]\right. \\ &\qquad\left. +c_2\, z^{k_2-k_1}\,{}_2F_1\left[\begin{matrix}-\nu_1-k_1,-\nu_2-k_1\\ 1-k_1+k_2\end{matrix};\, z\right]\right) \\ &= z^{k_1-1}(1-z)^{k_{\text{on}}-1}\left(\sum_{i=0}^{\infty}\left[a_i z^i + b_i z^{i+k_2-k_1}\right]\right), \end{aligned}$$

where without loss of generality we assumed that $k_1 < k_2$, and the a_i and b_i are constants. Due to the generalised hypergeometric functions, we have not been able to factorize the expression any further. Nonetheless, f_Z and f_X enable us to search for qualitatively different behaviours of $P(n)$ using "informed" trial and improvement. Fig. 6.4 shows

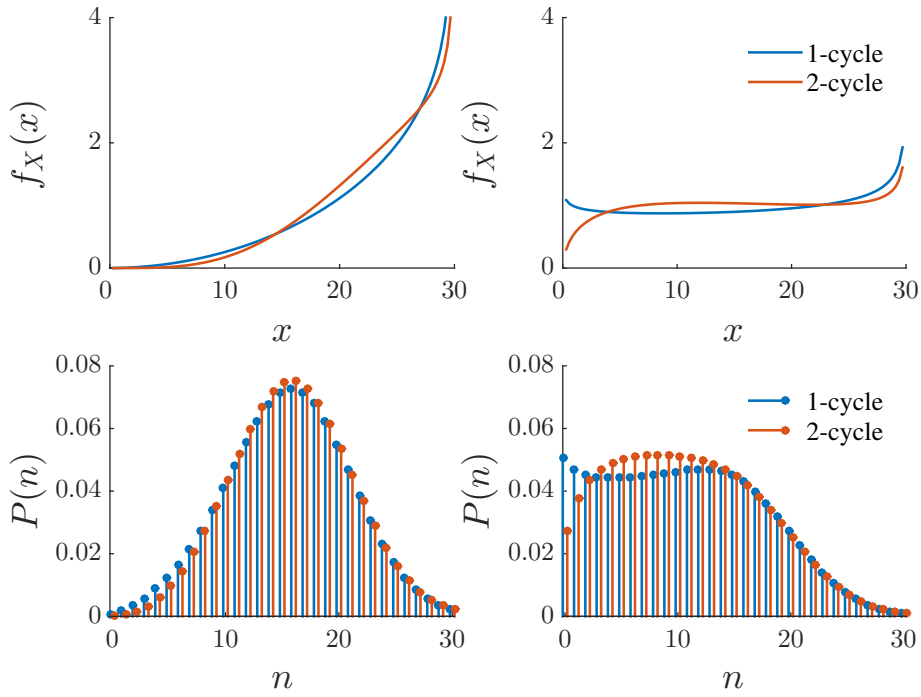

Figure 6.5: Use of the mixing density f_X *(top row)* to identify parameter regimes where the probability distributions $P(n)$ *(bottom row)* for the 1-cycle (blue) and the 2-cycle (red) are similar (*left*) and qualitatively different (*right*). Legends apply to both graphs in its row

f_X and $P(n)$ for four sets of parameters for the 2-cycle. For $P(n)$ to show bimodality for example, we require f_X to be bimodal (which corresponds to slow switching at the promoter, see Chapter 5). f_X monotonically decreasing implies that $P(n)$ will be too. f_X unimodal or monotonically increasing leads to a unimodal $P(n)$, with non-zero mode.

The same idea can be used for testing whether the addition of inactive states has any qualitative effect on the resulting distributions $P(n)$. Since all model-specific behaviour is contained in the mixing density, it suffices to find parameter regimes where f_X is qualitatively different for each model. Fig. 6.5 illustrates this concept for comparisons of the 1-cycle and 2-cycle, where the expected waiting times τ_{on} and τ_{off} were kept equal for the two models. If the mixing density is similar for the two models, we can immediately deduce that $P(n)$ will also be similar. Conversely, when searching for parameter regimes where $P(n)$ are qualitatively different, we need to find regimes where f_X displays different behaviour for each model.

6.5 Final remarks

We have shown how to calculate the exact solution of the M-cycle in two ways, each of which is useful for different purposes. Using an adjugate matrix, we determined that the relevant parameter combinations are the transition rate constants for the gene state switching process, and the non-zero eigenvalues of the transition matrix K. The appearance of the eigenvalues is significant; further properties of the solution or implied

timescales may be forthcoming via spectral analysis procedures, as has been achieved in several related fields to show how perturbations die away, to characterise convergence to stationarity, or to describe how noise propagates in a system (Walczak *et al.* 2009; Kampen 1992). Moreover, the eigenvalues for the M-cycle always appear in products of the form $\prod_{i=1}^{M}(-\nu_i + r),\ r \in \mathbb{N}$, which we can calculate as we did for $\prod_{i=1}^{M}(-\nu_i + 1)$ (see Eq. (6.18)) to obtain

$$\prod_{i=1}^{M}(-\nu_i + r) = (r + k_{\text{on}}) \prod_{i=1}^{M}(r + k_i) - k_{\text{on}} \prod_{i=1}^{M} k_i \tag{6.37}$$

$$= \frac{(r\tau_{\text{on}} + 1)\prod_{i=1}^{M}(r\tau_i + 1) - 1}{\tau_{\text{on}} \prod_{i=1}^{M} \tau_i}. \tag{6.38}$$

Hence $P(n)$ can be written explicitly in terms of only the expected waiting times τ_i in each state, and the transcription rate μ.

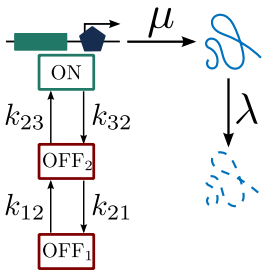

Figure 6.6: OFF-OFF-ON model of gene transcription. The gene transitions between the active state (ON) and the inactive states (OFF_1 and OFF_2) with the rates shown. Transcription only takes place in the ON state, with rate constant μ. Degradation occurs as a first-order reaction with rate constant λ, independently of gene state.

Note also that we can use the same method of solution for any multistate model with one active state and M inactive states; the solution will still be in terms of gamma functions and a ${}_M F_M$ generalised hypergeometric function with eigenvalues in the denominator, but for the parameters in the numerator we need to obtain the roots of an M^{th}-order polynomial (the polynomial is given by the $(M+1, M+1)^{\text{st}}$ minor of $[\theta I - K]$). For example, for the OFF-OFF-ON ladder model with transition rate matrix

$$K = \begin{pmatrix} -k_{12} & k_{21} & 0 \\ k_{12} & -(k_{12} + k_{23}) & k_{32} \\ 0 & k_{23} & -k_{32} \end{pmatrix}, \tag{6.39}$$

(see Fig. 6.6), we can follow the method in Section 6.2.1 step-by-step until we reach Eq. (6.6), where the $(M+1)^{\text{st}}$ coordinate is

$$\det(\theta I - K)\, H_{\text{on}}(s) = \mathrm{M}_{M+1,M+1}\, s\, H_{\text{on}}(s) \tag{6.40}$$

and

$$\mathrm{M}_{M+1,M+1} = \det \begin{pmatrix} \theta + k_{12} & -k_{21} \\ -k_{12} & \theta + (k_{23} + k_{21}) \end{pmatrix} \tag{6.41}$$

$$= (\theta + \kappa_+)(\theta + \kappa_-); \tag{6.42}$$

$$\kappa_\pm = \frac{1}{2}\left[k_{12} + k_{21} + k_{23} \pm \sqrt{(k_{12} + k_{21} + k_{23})^2 - 4k_{12}k_{23}}\right]. \tag{6.43}$$

We can then straightforwardly continue as for the M-cycle, replacing k_1 and k_2 by κ_+ and κ_- where appropriate, to obtain the solution

$$P(n) = \frac{\Gamma(\kappa_+ + n)}{\Gamma(\kappa_+)} \frac{\Gamma(\kappa_- + n)}{\Gamma(\kappa_-)} \frac{\Gamma(-\nu_1)}{\Gamma(-\nu_1 + n)} \frac{\Gamma(-\nu_2)}{\Gamma(-\nu_2 + n)} \frac{\mu^n}{n!} \, {}_2F_2 \left[\begin{matrix} \kappa_+ + n, & \kappa_- + n \\ -\nu_1 + n, & -\nu_2 + n \end{matrix} ; \, -\mu \right].$$

For these examples with only one active state, we needed only obtain $H_{\mathrm{on}}(s)$, since the transformed probability generating function $H(s)$ satisfied $dH/ds = H_{\mathrm{on}}(s)$. For models with more than one active state, we can still use the same general approach but we would find that dH/ds is a linear combination of each function $H_i(s)$ corresponding to an active state.

Similar comments can be made for finding the mixing densities of other multistate promoter models, but we can make further deductions because, as alluded to in the text, we obtain *Fuchsian* equations. These properties are discussed in Chapter 7.

Chapter 7

Which models are solvable?

The Poisson mixture result shows that extrinsic fluctuations or cell-to-cell variation in parameter values can be completely described by the single explanatory variable X_t. Further, in Chapter 4 we showed that X_t satisfies a simple first-order random differential equation that has countless applications in numerous fields, and has been studied for over a century (Soong 1973; Langevin 1908; Risken 1989). In this chapter, we take advantage of both of these insights to find shortcuts towards solutions and properties of multistate models that have not yet been solved.

In Section 7.1 we show that all Markov chain multistate models have certain regularity properties that enable several deductions about the form that solutions should take, and why. In Section 7.2, we draw from literature in other fields where f_{X_t} has already been calculated for non-Markovian ON-OFF models, so we can write down the exact Poisson mixture form for $P(n)$ directly.

7.1 Insights for Markov chain multistate models using Fuchsian theory

In Section 6.4 we showed how to use the multistate Kramers-Moyal equations to obtain $P(n)$ and the mixing density f_X for the M-cycle model. The same approaches can be used for any multistate model where the random walk between promoter states is independent of the state of the system, n: we can write down the set of L ODEs for the probability generating functions for each state, or write down the L multistate Kramers-Moyal equations. In theory, the systems of equations can be solved to obtain $P(n)$ and f_X.

Arguably more important than a solution for a particular model, however, is a means of predicting which models are solvable and the form the solutions might take. The derivation of f_X via the multistate Kramers-Moyal equations allows us to do just that for Markov chain multistate models, where the random walk between the promoter states is a Markov chain.

Essentially, for Markov chain multistate models the system of multistate Kramers-Moyal equations is *Fuchsian*, meaning that it has numerous important properties that allow us to make immediate inferences about the form of the equivalent L^{th}-order differential equation, and in some cases the form of the solution. A full treatment of these questions requires a fairly advanced level of analysis and geometry of differential equations, so after a brief interlude to introduce some context and required definitions, we will only state some relevant classical results and briefly describe the implications and possible routes for further work.

7.1.1 Interlude: A very brief introduction to Fuchsian systems

We will be considering a system of L first-order ODEs for $\mathbf{f} \coloneqq (f_1, \ldots, f_L)^T$. It is well known that an L^{th}-order ODE can be rewritten as a system of L first-order ODEs, but the converse is also true: any coupled system of first-order ODEs is equivalent to a system which comes from an ODE (Beukers 2009).

Now, consider the differential equation

$$\frac{d^L y}{dx^L} + q_1(x)\frac{d^{L-1}y}{dx^{L-1}} + \cdots + q_{L-1}(x)\frac{dy}{dx} + q_L(x)y = 0,$$

where the $q_i(x)$ are rational functions. To study this ODE it is useful to know whether it has any singularities, and how the solution will grow close to those singular points. In particular, close to a *regular* singularity at point p, $|y(x)|$ will grow at a rate of at most $|x-p|^{-L}$. Hence around regular singularities any growth is moderate, and ensures that several important properties will hold. For example, y will only have a finite number of Laurent terms.

An ODE or a system of first-order ODEs where all points are either regular or a regular singularity is called *Fuchsian*, and for the reasons mentioned above they have been of great interest to mathematicians, physicists, and engineers for centuries. In fact, Hilbert's twenty-first problem concerned Fuchsian systems, and such greats as Poincaré, Riemann, Birkhoff, and Grothendieck contributed far-reaching, general results to the field. We are thus extremely fortunate to find that our multistate Kramers-Moyal equations are Fuchsian for Markov chain multistate models (as will be shown in the next section), as it means that we can immediately apply some of these results. Here we mention the most relevant conclusions.

First, when studying a differential equation or an analytic function it is useful to find and classify the singularities. In general, detecting regular singularities can be difficult but Fuchsian systems have a particularly simple form: if our $L \times L$ system of ODEs for $\mathbf{y} \coloneqq (y_1, \ldots, y_L)^T$ is given by

$$\frac{d\mathbf{y}}{dx} = \mathcal{A}(x)\mathbf{y}, \quad \text{with} \quad \mathcal{A}(x) = \sum_{i=1}^{r} \frac{A_i}{x - p_i}, \tag{7.1}$$

where the A_i are constant matrices, the system is Fuchsian with finite regular singularities $\{p_i\}_{i=1}^r$. The point $x = \infty$ is regular if and only if $\sum_i A_i = 0$. Hence it is relatively easy to determine whether we have a Fuchsian system, and usually we simultaneously obtain the regular singular points. We will see this at play in the next section where we use this property to show that we have a Fuchsian system ourselves.

It has been proven (under certain assumptions) that a solution for L^{th}-order equations with r singularities exists (Plemelj 1964), and in certain cases one can construct an ansatz for the analytic solution near the singularities according to prescribed forms (Ilyashenko and Yakovenko 2008). Further, Fuchsian equations with few singularities can be completely characterised, or characterised to some extent:

One singularity: Any Fuchsian equation with only one singularity can be transformed into the equation

$$\frac{d^L y}{dx^L} = 0,$$

which has a regular singularity at $x = \infty$ (Beukers 2009). The solutions therefore take the form

$$y = c_1 x^{L-1} + \cdots + c_{L-1} x + c_L,$$

where $c_1, \ldots, c_L$ are constants.

Two singularities: Any Fuchsian equation with two singularities can be transformed into an Euler equation:

$$x^L \frac{d^L y}{dx^L} + a_1 x^{L-1} \frac{d^{L-1} y}{dx^{L-1}} + \cdots + a_{L-1} x \frac{dy}{dx} + a_L y = 0,$$

which has regular singularities at $x = 0$ and $x = \infty$ (Beukers 2009). The change of variable $x = e^u$ reduces this equation to a linear differential equation with constant coefficients.

Three singularities: Any Fuchsian equation with three singularities can be transformed into the equation

$$x^{L-1}(1-x)\frac{d^L y}{dx^L} + x^{L-2}(a_1 x - b_1)\frac{d^{L-1} y}{dx^{L-1}} + \cdots + a_L y = 0, \tag{7.2}$$

with regular singularities at $x = 0,\ x = 1$, and $x = \infty$ (Slater 1966). Assuming that $b_1, \ldots, b_{L-1}$ are distinct modulo 1, Eq. (7.2) has L linearly independent solutions for $|x| < 1$, given by the generalised hypergeometric functions

$${}_L F_{L-1}\left[\begin{matrix} a_1, & \ldots, & a_L \\ b_1, & \ldots, & b_{L-1} \end{matrix};\ x\right], \text{ and}$$

$$x^{1-b_i}\,{}_L F_{L-1}\left[\begin{matrix} 1+a_1-b_i, & \ldots\ldots, & 1+a_{L-1}-b_i, & 1+a_L-b_i \\ 1+b_1-b_i, & ..\vee.., & 1+b_{L-1}-b_i, & 2-b_i \end{matrix};\ x\right],\ i = 1, \ldots, L-1,$$

where the $\vee$ indicates that the term $1 + b_i - b_i$ is omitted from the sequence in the denominator. This was the case we saw for the M-cycle in Chapter 6.

Order two, four singularities: Second-order Fuchsian equations with four singular points are usually called Heun's equation. Its solutions are listed in Maier's recent article (Maier 2007). See Ronveaux's excellent book (Ronveaux 1995) for a comprehensive discussion of Heun's equation, its solutions and its properties.

Order one, r singularities: Any first-order Fuchsian equation with singularities $\{p_1, \ldots, p_r\}$ can be written in the form

$$\frac{dy}{dx} + \left(\frac{a_1}{x - p_1} + \cdots + \frac{a_1}{x - p_r}\right) y = 0,$$

where the $a_1, \ldots, a_r$ are scalar constants, and hence y has the form

$$y = \frac{C}{(x-p_1)^{a_1} \ldots (x-p_r)^{a_r}},$$

where C is a constant of integration.

7.1.2 Multistate Kramers-Moyal equations are Fuchsian

Now we know some basics of Fuchsian systems, let us return to multistate models of gene transcription and use this knowledge to our advantage. To avoid confusion with the M-cycle model from the previous chapter, in this section we specify the general Markov chain multistate model we will be considering here, and show that we obtain a Fuchsian system. The following sections will use this information to draw conclusions regarding specific gene transcription models.

We consider a multistate model with L distinct promoter states $\{s_1, \ldots, s_L\}$, where in state s_i the transcription and degradation rates are constants μ_i and λ_i, respectively. Promoter state transitions are modelled by a Markov chain, with k_{ij} the transition rate from state s_j to s_i. We assume that the cells are uncoupled, so the system has a stationary solution. For simplicity, we assume the system is already at stationarity.

Ultimately, we are interested in the probability distribution $P(n)$ for the number n of mRNA transcripts, but we showed in Section 3.2 that $P(n)$ can be written as a mixture distribution

$$P(n) = \int \frac{\xi^n}{n!} e^{-\xi} f_X(\xi)\, d\xi.$$

Recall from Section 4.4 that to find f_X, we can expand the state space to $\{X, S\}$ and consider the joint probability density functions $f_i(x) := f_{X,S}(x, s_i)$ for $i = 1, \ldots, L$. Then we have $f_X = \sum_i f_i$, and the f_i satisfy the multistate Kramers-Moyal equations

$$\frac{d}{dx}\left[(\mu_i - \lambda_i x) f_i\right] = \sum_{j=1}^{L} k_{ij} f_j - \left(\sum_{j=1}^{L} k_{ji}\right) f_i.$$

In matrix form, we have

$$\frac{d}{dx}\left[\operatorname{diag}\{\mu_i - \lambda_i x\}\mathbf{f}\right] = K\mathbf{f}, \tag{7.3}$$

where

$$\mathbf{f} := (f_1, \ldots, f_L)^T,$$
$$\operatorname{diag}\{\mu_i - \lambda_i x\} := \operatorname{diag}(\mu_1 - \lambda_1 x, \ldots, \mu_L - \lambda_L x),$$
$$\text{and } K_{i,j} := \begin{cases} -\sum_{r \neq j} k_{rj} & \text{if } i = j, \\ k_{ij} & \text{otherwise.} \end{cases}$$

Using the product rule on the left hand side of Equation (7.3), we can rewrite it as:

$$\begin{aligned}\frac{d\mathbf{f}}{dx} &= [\text{diag}\{\mu_i - \lambda_i x\}]^{-1}(K + \text{diag}\{\lambda_i\})\mathbf{f} \\ &= \sum_{i=1}^{L} \frac{A_i}{x - \frac{\mu_i}{\lambda_i}}\,\mathbf{f},\end{aligned}$$

where the A_i, $i = 1, \ldots, L$, are $L \times L$ matrices formed of the i^{th} row of $-(K + \text{diag}\{\lambda_i\})/\lambda_i$, and zeros everywhere else, i.e.

$$(A_i)_{n,m} = \begin{cases} -\left(\frac{K_{i,m}}{\lambda_i} + \delta_{i,m}\right) & \text{if } n = i, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, by definition (see Eq. (7.1)), we have a Fuchsian system with finite singularities $\{p_i := \mu_i/\lambda_i\}_{i=1}^{L}$, and $p_\infty = \infty$. Notice that for a Markov chain multistate model we have at least three singularities[1], including p_∞. Furthermore, by employing a Möbius transformation we can map any three singularities to a second set of three distinct points. Since it is not usual to have different degradation rates for different gene states, we will assume that distinct singularities simply correspond to distinct transcription rates, and without loss of generality we will always map the smallest and largest finite singularities to zero and one, respectively. In doing so, we transform the extrinsic random variable X to the random variable Z, which takes values from the interval $(0, 1)$.

It is worth pointing out here that the Fuchsian equations that correspond to the Fuchsian system are usually in terms of one of the components of $\mathbf{f}$. For our multistate models, conservation of probability constrains the rows of the transition matrix K to sum to zero. We then have one degree of freedom less than the dimension of the system, so that for an L- state model we obtain Fuchsian equations of order $L - 1$. This effect was seen in the derivation of the mixing density for the cyclic model (Section 6.4), but for ease of reading it will be demonstrated for the two state model in Section 7.1.3.

Note that we have not imposed a Fuchsian system on our model; the general multistate model considered in this chapter is almost ubiquitous in studies involving gene transcription models for a single gene. In studying them from the context of the Poisson mixture result in Chapter 3, we were able to see this extremely fortunate[2] underlying structure.

Armed with this information, we can now look back at the properties of Fuchsian systems to help us in two directions. First, to understand the results we already have for multistate gene transcription models in this wider context, and second, to guide us towards potentially profitable directions to explore. As mentioned before, a thorough exploration

[1] For simplicity we are ignoring any cases where $\mu_i/\lambda_i = \mu_j/\lambda_j,\ \mu_i \neq \mu_j,\ \lambda_i \neq \lambda_j,\ i \neq j$, because they are not likely to be relevant for gene expression models.

[2] It is "fortunate" that the multistate Kramers-Moyal equations for the mixture distribution in these models is Fuchsian, as it means that we can draw upon the tremendous amount of work that has been conducted over centuries to understand the properties of Fuchsian systems. On the other hand, in a sense it is not surprising that the system is Fuchsian because as a model representing variation in transcription rates due to promoter state switching, we would expect it to behave moderately near singularities.

of the possibilities afforded to us with the Fuchsian tools would require a strong grasp of areas of mathematics that are beyond the scope of this thesis. Nevertheless, with only the basic properties mentioned in Section 7.1.1 we will use the following sections to:

- clarify why the solution of the two-state (random telegraph) model is completely characterised and takes the form that it does;
- say why the mixing density for any Markov chain L-state model with only two distinct transcription rates should be a linear combination of ${}^{k}{}_{L-1}F_{L-2}$ functions; and
- justify that the mixing density for any three-state model with three distinct transcription rates should give us Heun's equation, the solution of which is known for certain cases.

7.1.3 Two-state model as a Fuchsian system

Recall that a Markov chain multistate model has a minimum of three singularities, including $p_\infty = \infty$. Therefore for the two state model we have two finite singularities, which can be mapped using a Möbius transformation to zero and one. We have no extra degrees of freedom, so the model is completely characterised.

As mentioned in Section 7.1.1, *a priori* the solution of a Fuchsian equation deriving from a second order Fuchsian system with three singularities should be a linear combination of ${}_2F_1$ functions. However, because the 2×2 transition matrix K only has rank 1, the ${}_2F_1$ functions reduce to ${}_1F_0$. Let's illustrate this property to see how the reduction happens.

Assume we have already applied a Möbius transformation, so the singularity for the inactive state is at zero and the singularity for the active state is at one. As we did when deriving the mixing density for the cyclic model in Section 6.4, denote the transformed extrinsic random variable by Z and the joint probability density functions by f_{on} and f_{off}. $f_Z = f_{\text{on}} + f_{\text{off}}$ thus has support on the interval $(0, 1)$. k_{on} and k_{off} are the transition rates to the active and inactive states respectively, as shown in the multistate Kramers-Moyal equations for this system:

$$\frac{d}{dz}\left[-z f_{\text{off}}\right] = -k_{\text{on}} f_{\text{off}} + k_{\text{off}} f_{\text{on}} \tag{7.4a}$$

$$\frac{d}{dz}\left[(1-z) f_{\text{on}}\right] = \; k_{\text{on}} f_{\text{off}} - k_{\text{off}} f_{\text{on}}. \tag{7.4b}$$

Direct substitution method

With only two equations, this system can easily be solved by direct substitution. Sum Eqs (7.4) to obtain

$$\frac{d}{dz}[-z f_{\text{off}}(z) + (1-z) f_{\text{on}}(z)] = 0,$$

i.e.

$$-z f_{\text{off}}(z) + (1-z) f_{\text{on}}(z) \;=\; c_1,$$

where c_1 is a constant of integration. Using the boundary condition $f_{\text{on}}(0) = 0$ (see Eq. (6.25b) for a reminder), we find that $c_1 = 0$, so we can write

$$f_{\text{off}} = \frac{1-z}{z} f_{\text{on}} \tag{7.5}$$

and substitute into Eq. 7.4b. Rearranging, we find

$$z(1-z)\frac{df_{\text{on}}}{dz} = \left[k_{\text{on}} + (1 - k_{\text{on}} - k_{\text{off}})z\right] f_{\text{on}},$$

with general solution

$$\begin{aligned} f_{\text{on}}(z) &= c_2 z^{k_{\text{on}}} (1-z)^{k_{\text{off}}-1} \\ &= c_2 z^{k_{\text{on}}}\, {}_1F_0\left[1 - k_{\text{off}};\ ;z\right]. \end{aligned}$$

The integral condition (Eq. (6.24))

$$\int_0^1 f_{\text{on}}(z)\, dz = \frac{k_{\text{on}}}{k_{\text{on}} + k_{\text{off}}}$$

gives us the particular solution

$$\begin{aligned} f_{\text{on}}(z) &= \frac{\Gamma(k_{\text{off}} + k_{\text{on}})}{\Gamma(k_{\text{off}})\Gamma(k_{\text{on}})}\, z^{k_{\text{on}}} (1-z)^{k_{\text{off}}-1} \\ &\equiv \frac{\Gamma(k_{\text{off}} + k_{\text{on}})}{\Gamma(k_{\text{off}})\Gamma(k_{\text{on}})}\, z^{k_{\text{on}}}\, {}_1F_0\left[1 - k_{\text{off}};\ ;z\right]. \end{aligned} \tag{7.6}$$

Using Eq. (7.5) it follows that

$$\begin{aligned} f_{\text{off}}(z) &= \frac{\Gamma(k_{\text{off}} + k_{\text{on}})}{\Gamma(k_{\text{off}})\Gamma(k_{\text{on}})}\, z^{k_{\text{on}}-1} (1-z)^{k_{\text{off}}} \\ &\equiv \frac{\Gamma(k_{\text{off}} + k_{\text{on}})}{\Gamma(k_{\text{off}})\Gamma(k_{\text{on}})}\, z^{k_{\text{on}}-1}\, {}_1F_0\left[-k_{\text{off}};\ ;z\right], \end{aligned}$$

and so

$$\begin{aligned} f_Z(z) &= f_{\text{on}}(z) + f_{\text{off}}(z) \\ &= \frac{\Gamma(k_{\text{off}} + k_{\text{on}})}{\Gamma(k_{\text{off}})\Gamma(k_{\text{on}})}\, z^{k_{\text{on}}-1} (1-z)^{k_{\text{off}}-1} \\ &\equiv \frac{\Gamma(k_{\text{off}} + k_{\text{on}})}{\Gamma(k_{\text{off}})\Gamma(k_{\text{on}})}\, z^{k_{\text{on}}-1}\, {}_1F_0\left[1 - k_{\text{off}};\ ;z\right]. \end{aligned}$$

Notice that using this method of direct substitution, we immediately obtain solutions in terms of ${}_1F_0\left[1 - k_{\text{off}};\ ;z\right]$ because we used the fact that the row sum of the transition

matrix K is equal to zero at the beginning.

Matrix method of solution shown in Section 6.4

Following the method shown in Section 6.4 for the cyclic model, write Eqs (7.4a)-(7.4b) in the following form:

$$\begin{pmatrix} z\frac{d}{dz} + 1 - k_{\text{on}} & k_{\text{off}} \\ k_{\text{on}} & z\frac{d}{dz} + 1 - k_{\text{off}} \end{pmatrix} \begin{pmatrix} f_{\text{off}} \\ f_{\text{on}} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \frac{d}{dz} \begin{pmatrix} f_{\text{off}} \\ f_{\text{on}} \end{pmatrix}.$$

Denote the matrix on the left of this last equation by $(\theta I + I + K)$, with $\theta := z\frac{d}{dz}$. Multiply both sides of the equation on the left by

$$\text{adj}(\theta I + I + K) = \begin{pmatrix} \theta + 1 - k_{\text{off}} & -k_{\text{off}} \\ -k_{\text{on}} & \theta + 1 - k_{\text{on}} \end{pmatrix}$$

to obtain

$$\det(\theta I + I + K) \begin{pmatrix} f_{\text{off}} \\ f_{\text{on}} \end{pmatrix} = \begin{pmatrix} 0 & -k_{\text{off}} \\ 0 & \theta + 1 - k_{\text{on}} \end{pmatrix} \frac{1}{z}\,\theta \begin{pmatrix} f_{\text{off}} \\ f_{\text{on}} \end{pmatrix}. \tag{7.7}$$

Recall that the determinant of a matrix is equal to the product of its eigenvalues; the eigenvalues of K are 0 and $-k_{\text{on}} - k_{\text{off}}$, so the second coordinate of Eq. (7.7) is

$$\begin{aligned} (\theta + 1)(\theta + 1 - k_{\text{on}} - k_{\text{off}}) f_{\text{on}} &= (\theta + 1 - k_{\text{on}}) \frac{1}{z}\,\theta\, f_{\text{on}} \\ &= \frac{1}{z}(\theta - k_{\text{on}})\,\theta\, f_{\text{on}}, \end{aligned}$$

or, when expanded out,

$$z(1-z)\frac{d^2 f_{\text{on}}}{dz^2} + [1 - k_{\text{on}} - (3 - k_{\text{on}} - k_{\text{off}})z]\,\frac{df_{\text{on}}}{dz} - (1 - k_{\text{on}} - k_{\text{off}}) f_{\text{on}} = 0.$$

This last equation is Euler's hypergeometric differential equation, which has general solution (Slater 1966)

$$\begin{aligned} f_{\text{on}}(z) &= c_0\, {}_2F_1 \left[\begin{matrix} 1 - k_{\text{off}} - k_{\text{on}}, & 1 \\ & 1 - k_{\text{on}} \end{matrix} ;\, z \right] \\ &\quad + c_1\, z^{k_{\text{on}}}\, {}_2F_1 \left[\begin{matrix} 1 - k_{\text{off}} - k_{\text{on}} + k_{\text{on}}, & 1 + k_{\text{off}} \\ & 1 + k_{\text{off}} \end{matrix} ;\, z \right] \\ &= c_0\, {}_2F_1 \left[\begin{matrix} 1 - k_{\text{off}} - k_{\text{on}}, & 1 \\ & 1 - k_{\text{on}} \end{matrix} ;\, z \right] + c_1\, z^{k_{\text{on}}}\, {}_1F_0 \left[1 - k_{\text{off}};\, ;\, z \right] \\ &= c_0\, {}_2F_1 \left[\begin{matrix} 1 - k_{\text{off}} - k_{\text{on}}, & 1 \\ & 1 - k_{\text{on}} \end{matrix} ;\, z \right] + c_1\, z^{k_{\text{on}}} (1 - z)^{k_{\text{off}} - 1}. \end{aligned}$$

Notice that the second ${}_2F_1$ reduced to a ${}_1F_0$ because of the equal parameters in the numerator and denominator. The remaining ${}_2F_1$ does not appear in the particular solution

because of the boundary condition at $z = 0$ (see Eq. (6.25b)). Using the conditions

$$f_{\text{on}}(0) = 0 \qquad \text{and} \qquad \int_0^1 f_{\text{on}}(z)\, dz = \frac{k_{\text{on}}}{k_{\text{on}} + k_{\text{off}}},$$

(see Eqs (6.25b) and (6.24)) we obtain

$$\begin{aligned} f_{\text{on}}(z) &= \frac{k_{\text{on}}}{k_{\text{on}} + k_{\text{off}}} \frac{\Gamma(k_{\text{on}} + k_{\text{off}} + 1)}{\Gamma(k_{\text{on}} + 1)\Gamma(k_{\text{off}})} z^{k_{\text{on}}} (1 - z)^{k_{\text{off}} - 1} \\ &= \frac{\Gamma(k_{\text{on}} + k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{off}})} z^{k_{\text{on}}} (1 - z)^{k_{\text{off}} - 1} \\ &= \frac{\Gamma(k_{\text{on}} + k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{off}})} z^{k_{\text{on}}}\, {}_1F_0\left[1 - k_{\text{off}};\ ;z\right]. \end{aligned}$$

Similarly,

$$\begin{aligned} f_{\text{off}}(z) &= \frac{k_{\text{off}}}{k_{\text{on}} + k_{\text{off}}} \frac{\Gamma(k_{\text{on}} + k_{\text{off}} + 1)}{\Gamma(k_{\text{on}})\Gamma(k_{\text{off}} + 1)} z^{k_{\text{on}} - 1} (1 - z)^{k_{\text{off}}} \\ &= \frac{\Gamma(k_{\text{on}} + k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{off}})} z^{k_{\text{on}} - 1} (1 - z)^{k_{\text{off}}} \\ &= \frac{\Gamma(k_{\text{on}} + k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{off}})} z^{k_{\text{on}} - 1}\, {}_1F_0\left[-k_{\text{off}};\ ;z\right], \end{aligned}$$

so finally

$$\begin{aligned} f_Z(z) &= f_{\text{on}}(z) + f_{\text{off}}(z) \\ &= \frac{\Gamma(k_{\text{on}} + k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{off}})} z^{k_{\text{on}} - 1} (1 - z)^{k_{\text{off}} - 1} \\ &= \frac{\Gamma(k_{\text{on}} + k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{off}})} z^{k_{\text{on}} - 1}\, {}_1F_0\left[1 - k_{\text{off}};\ ;z\right]. \end{aligned}$$

Notice that using this method, we obtained solutions in terms of ${}_2F_1$ first, as expected for a Fuchsian system of dimension two with three singularities. However, the solution eventually reduces to the form ${}_1F_0$, to be consistent with the solution (7.6) we obtain by direct substitution.

7.1.4 L-state models with two distinct transcription rates

Here we consider a generalisation of the two-state ON-OFF model by increasing the number of states, without increasing the number of distinct transcription rates. In other words, we have L gene states that can be separated into two mutually exclusive subsets according to transcription rate, which we will refer to as the inactive (or OFF) states, and the active (or ON) states.

As usual, we use a Möbius transformation to map the finite singularity corresponding to the inactive (resp. active) state to zero (resp. one). This L^{th}-order Fuchsian system

has three singularities (including the singularity at infinity). In general, corresponding Fuchsian equations have solutions that are linear combinations of functions of the form ${}_{L}F_{L-1}$ for $|z| < 1$ (Slater 1966), but since we have the extra constraint that the row sum of the transition matrix K is equal to zero, the solutions reduce to linear combinations of functions of the form ${}_{L-1}F_{L-2}$. One example has already been given in the derivation of the mixing density of the M-cycle (Section 6.4): the model has a total of $M+1$ states, and we found that $f_{\text{on}}(z)$ (and hence f_Z) is a linear combination of the functions ${}_{M}F_{M-1}$ (see Eqs (6.30) and (6.31)). The following example shows that it holds true for a different structure of the gene states.

Example 7.1–1 — Mixing density for the OFF-OFF-ON ladder model

Consider the OFF-OFF-ON "ladder" model as illustrated in Fig. 7.1a, and map the finite singularity μ/λ to one. The multistate Kramers-Moyal equations for this system, where f_1 and f_2 correspond to the gene states OFF_1 and OFF_2, respectively, and f_{on} corresponds to the ON state, are

$$\frac{d}{dz}\left[\left(\mathbf{e}_3\mathbf{e}_3^T - zI\right)\mathbf{f}(z)\right] = K\mathbf{f}(z),$$

where $\mathbf{e}_3 = (0,0,1)^T$, $\mathbf{f} = (f_1, f_2, f_{\text{on}})^T$, and

$$K = \begin{pmatrix} -k_{21} & k_{12} & 0 \\ k_{21} & -(k_{12}+k_{32}) & k_{23} \\ 0 & k_{32} & -k_{23} \end{pmatrix}.$$

Denoting $z\frac{d}{dz}$ by θ and following the method shown in Section 6.4, we obtain

$$f_Z = f_1 + f_2 + f_{\text{on}} = \frac{1}{z}\, f_{\text{on}},$$
$$f_1(1) = f_2(1) = f_{\text{on}}(0) = 0,$$

and

$$\det(\theta I + I + K)\,\mathbf{f} = \text{adj}(\theta I + I + K)\,\mathbf{e}_3\,\mathbf{e}_3^T\,\frac{1}{z}\,\theta\,\mathbf{f}.$$

The third coordinate of this equation is

$$\begin{aligned}(\theta+1+\nu_+)(\theta+1+\nu_-)(\theta+1)\,f_{\text{on}}(z) &= (\theta+1+\kappa_+)(\theta+1+\kappa_-)\,\frac{1}{z}\,\theta\,f_{\text{on}}(z) \\ &= \frac{1}{z}\,(\theta-\kappa_+)(\theta-\kappa_-)\,\theta\,f_{\text{on}}(z), \qquad (7.8)\end{aligned}$$

where $\nu_\pm$ are the eigenvalues of K and $\kappa_\pm$ are obtained when factorising the minor

$\mathrm{M}_{3,3}$ of the matrix $[\theta I + I + K]$:

$$\nu_\pm = \frac{1}{2}\left(-k_{21} - k_{12} - k_{32} - k_{23} \pm \sqrt{(k_{21}+k_{12}+k_{32}+k_{23})^2 - 4(k_{21}k_{32} + k_{21}k_{23} + k_{12}k_{23})}\right)$$
$$\kappa_\pm = \frac{1}{2}\left(k_{21} + k_{12} + k_{32} \pm \sqrt{(k_{21}+k_{12}+k_{32})^2 - 4k_{21}k_{32}}\right).$$

Equation (7.8) has the form of a generalised hypergeometric equation (Beukers 2009), so we can straightforwardly write down the general solution:

$$\begin{aligned}
f_{\text{on}}(z) &= c_0\, z^0\, {}_3F_2\left[\begin{matrix}1+\nu_+, & 1+\nu_-, & 1\\ 1-\kappa_+, & 1-\kappa_- & \end{matrix}\;;\; z\right] \\
&\quad + c_1\, z^{\kappa_+}\, {}_3F_2\left[\begin{matrix}1+\nu_++\kappa_+, & 1+\nu_-+\kappa_+, & 1+\kappa_+\\ & 1-\kappa_-+\kappa_+, & 1+\kappa_+ \end{matrix}\;;\; z\right] \\
&\quad + c_2\, z^{\kappa_-}\, {}_3F_2\left[\begin{matrix}1+\nu_++\kappa_-, & 1+\nu_-+\kappa_-, & 1+\kappa_-\\ & 1-\kappa_++\kappa_-, & 1+\kappa_- \end{matrix}\;;\; z\right] \\
&= c_0\, z^0\, {}_3F_2\left[\begin{matrix}1+\nu_+, & 1+\nu_-, & 1\\ 1-\kappa_+, & 1-\kappa_- & \end{matrix}\;;\; z\right] \\
&\quad + c_1\, z^{\kappa_+}\, {}_2F_1\left[\begin{matrix}1+\nu_++\kappa_+, & 1+\nu_-+\kappa_+\\ & 1-\kappa_-+\kappa_+ \end{matrix}\;;\; z\right] \\
&\quad + c_2\, z^{\kappa_-}\, {}_2F_1\left[\begin{matrix}1+\nu_++\kappa_-, & 1+\nu_-+\kappa_-\\ & 1-\kappa_++\kappa_- \end{matrix}\;;\; z\right].
\end{aligned}$$

Notice again that the ${}_3F_2$ functions reduced to ${}_2F_1$, and the boundary condition $f_{\text{on}}(0) = 0$ implies that $c_0 = 0$, so that

$$\begin{aligned}
f_Z(z) &= \frac{1}{z} f_{\text{on}}(z) \\
&= c_1\, z^{\kappa_+ - 1}\, {}_2F_1\left[\begin{matrix}1+\nu_++\kappa_+, & 1+\nu_-+\kappa_+\\ & 1-\kappa_-+\kappa_+ \end{matrix}\;;\; z\right] \\
&\quad + c_2\, z^{\kappa_- - 1}\, {}_2F_1\left[\begin{matrix}1+\nu_++\kappa_-, & 1+\nu_-+\kappa_-\\ & 1-\kappa_++\kappa_- \end{matrix}\;;\; z\right].
\end{aligned}$$

We could also obtain a differential equation for f_{on} via substitution by hand (as we did in Section 7.1.3), getting Euler's hypergeometric equation and hence the Gaussian (${}_2F_1$) hypergeometric functions directly.

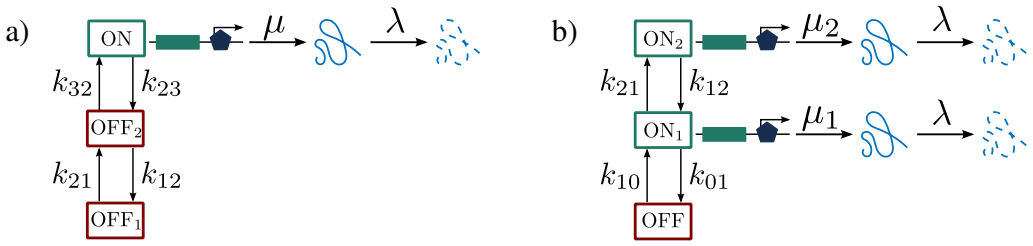

Figure 7.1: 3-state ladder models of gene transcription. *(a)* OFF-OFF-ON model, where the gene transitions between the active state (ON) and the inactive states (OFF_1 and OFF_2) with the rates shown. Transcription only takes place in the ON state, with rate constant μ. *(b)* OFF-ON-ON model, where the gene transitions between the inactive state (OFF) and the active states (ON_1 and ON_2) with the rates shown. Transcription only takes place in the ON_1 state with rate constant μ_1, and in the ON_2 state with rate constant $\mu_2 > \mu_1$. Degradation occurs as a first-order reaction with rate constant λ for both models, independently of gene state.

7.1.5 Three-state models with three distinct transcription rates

We showed in Section 7.1.2 that the multistate Kramers-Moyal equations for any Markov chain multistate model form a Fuchsian system. The finite regular singularities are given by the distinct transcription rates and infinity. Recall that the order of the corresponding Fuchsian equations reduce by one because the rows of K sum to zero. Hence, three-state models with three distinct transcription rates give rise to second order Fuchsian equations with four singularities, otherwise known as Heun's equation (Ronveaux 1995; Maier 2007). The canonical form of Heun's equation is

$$\frac{d^2y}{dz^2} + \left(\frac{\gamma}{z} + \frac{\delta}{z-1} + \frac{\epsilon}{z-a}\right)\frac{dy}{dz} + \frac{\alpha\beta z - q}{z(z-1)(z-a)}\,y = 0, \tag{7.9}$$

with the Fuchsian constraint

$$\gamma + \delta + \epsilon = \alpha + \beta + 1$$

to ensure that the singular point at $z = \infty$ is regular (Maier 2007).

Like the hypergeometric equation, solutions of Heun's equation are usually given as power-series solutions. However, the coefficients of the series are governed by three-term recursion relations so it is generally impossible to write down the series solution explicitly (Ronveaux 1995; Maier 2007). In some cases, the solution of Heun's equation can be written as a linear combination of Gauss hypergeometric (${}_2F_1$) functions (Craster and Hoàng 1998; Shanin and Craster 2002; Maier 2007; Ishkhanyan and Suominen 2009), or in integral form (Takemura 2008).

Due to the difficulties involved in solving Heun's equation, and the geometrical analysis of differential equations that would be required, obtaining the mixing density for a model with three distinct transcription rates goes beyond the scope of this thesis. However, we will show that we do indeed obtain Heun's equation for a relevant model of gene transcription.

Example 7.1–2 — Heun's equation for the OFF-ON-ON ladder model

The OFF-ON-ON "ladder" model (Fig. 7.1b) has been proposed to model interplay between the TATA box and the nucleosome (Tirosh and Barkai 2008), DNA condensation by nucleoid proteins or changes in DNA supercoiling (Sánchez *et al.* 2013), and variable polymerase initiation frequency (Senecal *et al.* 2014). Apply a Möbius transformation, so that the singularities corresponding to the OFF, ON_1 and ON_2 transcription rates are at $z = 0$, $z = a$, and $z = 1$, respectively, with $a < 1$. k_{ij} is the transition rate from gene state j to gene state i, as shown in Fig. 7.1. The multistate Kramers-Moyal equations for this system, where f_0 corresponds to the OFF state and f_1 and f_2 correspond to the gene states ON_1 and ON_2, respectively, are

$$\frac{d}{dz}[-zf_0] = -k_{10}f_0 + k_{01}f_1 \tag{7.10a}$$

$$\frac{d}{dz}[(a-z)f_1] = k_{10}f_0 - (k_{01}+k_{21})f_1 + k_{12}f_2 \tag{7.10b}$$

$$\frac{d}{dz}[(1-z)f_2] = k_{21}f_1 - k_{12}f_2. \tag{7.10c}$$

As we did for the cyclic model, we can show that

$$f_0(1) = f_1(0) = f_2(0) = 0,$$

hence if we sum the three equations (7.10) and integrate, the constant of integration will be equal to zero and we obtain

$$-zf_0 + (a-z)f_1 + (1-z)f_2 = 0. \tag{7.11}$$

We want to obtain an equation in terms of f_2 only, via substitution; the details are omitted, as they simply involve some tedious algebra. There are several routes, but for example one can use Eq. (7.11) to eliminate f_0 from Eq. (7.10b), and then Eq. (7.10c) to eliminate f_1, to obtain

$$\begin{aligned} 0 &= z(a-z)(1-z)\frac{d^2f_2}{dz^2} \\ &\quad + \{(k_{12}-2)z(a-z) + (1-z)\left[(k_{01}+k_{21}-1)z - k_{10}(a-z)\right]\}\frac{df_2}{dz} \\ &\quad + \{(k_{10}k_{21}+k_{10}k_{12}+k_{01}k_{12}-k_{10}-k_{01}-k_{21}-k_{12}+1)z \ldots \\ &\qquad\qquad -k_{10}(k_{12}m - m + k_{21})\}f_2(z) \\ &= \frac{d^2f_2}{dz^2} + \left\{\frac{-k_{10}}{z} + \frac{2-k_{12}}{z-1} + \frac{1-k_{01}-k_{21}}{z-a}\right\}\frac{df_2}{dz} + \frac{\alpha\beta z - q}{z(z-1)(z-a)}f_2, \end{aligned} \tag{7.12}$$

where

$$\alpha\beta = (k_{10}k_{21} + k_{10}k_{12} + k_{01}k_{12} - k_{10} - k_{01} - k_{21} - k_{12} + 1),$$
$$q = k_{10}(k_{12}m - m + k_{21}).$$

Comparison to Eq. (7.9) shows that Eq. (7.12) is indeed Heun's equation. An equation for f_0 can be obtained via a similar approach, and an equation for f_1 can be calculated by substitution of Eq. (7.12) into Eq. (7.10c).

7.2 ON-OFF models with non-Markovian promoter switching

The Markov chain two-state model is inadequate for describing systems with non-exponentially distributed waiting times in either gene state (Suter *et al.* 2011; Harper *et al.* 2011; Kandhavelu *et al.* 2012). So far, we have considered extensions of the two-state model that cater for such observations by increasing the number of ON and/or OFF states. The distribution for the total waiting time τ_{on} (resp. τ_{off}) spent in the active (resp. inactive) gene states then depends on the structure of the Markov chain gene state transition matrix K. However, extending the Markov chain two state model in this way constrains the possible waiting time distributions, as the time spent in each individual gene is still exponentially distributed. For example, for the M-cycle described in Chapter 6, the total waiting time τ_{off} spent in the inactive gene states before returning to the active state is the sum of M independent, exponentially-distributed waiting times (and hence τ_{off} is hypo-exponentially distributed (Smaili *et al.* 2013)).

A more general approach to account for non-exponential waiting times is to maintain only two gene states, but allow the waiting times τ_{on} and τ_{off} to be independent random variables drawn from any distribution we choose (Schwabe *et al.* 2012; Stinchcombe *et al.* 2012). In other words, we allow the jump process for the gene state to be non-Markovian.

Despite the simplicity of this system, we are not aware of any analytical expressions for the probability distribution $P(n)$ of the mRNA copy number that have been derived. The Poisson mixture result tells us that we need only obtain the mixing density f_X, but the corresponding multistate Kramers-Moyal-type equations require numerical solutions (Stinchcombe *et al.* 2012). Nevertheless, if we look further afield we find several useful results: calculating f_X for problems of this sort was of significant interest to electrical engineers in the 1950s–1980s, since the process X models the output of an RC filter driven by a binary random process (Wonham and Fuller 1958; McFadden 1959; Pawula and Rice 1986). More recently, Dubins-Freedman chains were shown to be special cases of RC filters (Mazza and Piau 2001). We can therefore draw directly from a fairly substantial body of existing results, including some exact solutions, elementary relations and integral equations for the conditional density of X given the gene state and the densities of X at the switching times, moment expressions and Laplace transforms, several approximation methods, and inevitably, indications as to which waiting time distributions might allow

us to obtain analytical solutions. We refer the interested reader to the excellent papers by Pawula and Rice (1986) and Mazza and Piau (2001) for an overview of the available results. For completeness we nevertheless state the exact solutions they obtained here, as they are instructive for suggesting amenable waiting time densities.

In the following, we assume without loss of generality that $\lambda = 1$ and we have already made the appropriate change of variable, e.g. $X = \mu Z$, so that the mixing density f_Z has support on $(0, 1)$ and its sample paths ζ satisfy

$$\frac{d\zeta}{dt} + \zeta(t) = \xi(t),$$

where $\xi(t) \in \{0, 1\}$ is the binary random process.

7.2.1 McFadden interval pdfs

Suppose the waiting times τ in both the active and the inactive gene states are distributed according to the McFadden interval pdf (McFadden 1959)

$$f_\tau(t) = \frac{e^{-at}\left(1 - e^{-t}\right)^{b-a-1}}{B(a, b-a)}, \qquad b > a,$$

where

$$B(\alpha, \beta) = \int_0^1 s^{\alpha-1}(1-s)^{\beta-1}\, ds$$

is the Beta function. Then f_Z is given by (Pawula and Rice 1986)

$$f_Z(z) = \frac{1 - I_z(b, a) - I_{1-z}(b, a)}{2[\psi(b) - \psi(a)]z(1-z)},$$

where

$$I_x(\alpha, \beta) = \frac{\int_0^x s^{\alpha-1}(1-s)^{\beta-1}\, ds}{B(\alpha, \beta)}$$

is the regularized incomplete beta function (Abramowitz and Stegun 1964), and

$$\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

is the logarithmic derivative of the gamma function, often referred to as the digamma function (Abramowitz and Stegun 1964). When $b - a \in \mathbb{N}$,

$$\psi(b) - \psi(a) = \sum_{k=0}^{b-a-1} \frac{1}{k+a},$$

and when $a, b \in \mathbb{N}$, f_Z reduces to

$$f_Z(z) = \frac{1 - \sum_{k=0}^{a-1} \binom{b+a-1}{k} \left[z^k(1-z)^{b+a-1-k} + z^{b+a-1-k}(1-z)^k\right]}{2z(1-z)\sum_{k=0}^{b-a-1} \frac{1}{k+a}}. \tag{7.13}$$

$b = a + 1$ and $b = a + 2$

Surprisingly, the first two cases of Eq. (7.13) give the same distribution for Z. When $b = a + 1$, the McFadden interval pdf reduces to the exponential distribution and we obtain the Beta distribution for Z, as expected. We obtain the same distribution when $b = a + 2$: for waiting time distributions

$$f_\tau(t) = ae^{-at} \qquad \text{or} \qquad f_\tau(t) = a(a+1)e^{-at}(1-e^{-t}),$$

the density of Z is given by the Beta distribution

$$f_Z(z) = \frac{z^{a-1}(1-z)^{a-1}}{B(a,a)}.$$

7.2.2 Gamma waiting time densities

Using integral equations adapted from McFadden 1959, in 1986 Pawula and Rice derived several differential equations for the conditional pdfs $f_{Z|\mathrm{ON}}$ and $f_{Z|\mathrm{OFF}}$, the densities of the local minima and maxima of Z, and f_Z. Taken together, once the waiting time densities are specified they can be used to obtain a differential equation in terms of f_Z only.

When the waiting times are gamma distributed with shape parameter 2 and rate parameter β, i.e.

$$f_\tau(t) = \beta^2 t e^{-\beta t},$$

the equation to solve is

$$\begin{aligned} 0 = {} & z^2(1-z)^2 f_Z''' - (5-2\beta)z(1-z)(2z-1)f_Z'' \\ & + [(\beta-2)^2 - 2(3\beta^2 - 11\beta + 12)z(1-z)]f_Z' \\ & + (3-2\beta)(\beta^2 - 2\beta + 2)(2z-1)f_Z. \end{aligned} \tag{7.14}$$

Pawula and Rice: $\tau \sim \Gamma(2, 1/2)$

When $\beta = 1/2$, Eq. (7.14) simplifies and can be solved:

$$f_Z(z) = \sqrt{\frac{2}{\pi^3}} \frac{\left|\Gamma\left(\frac{3-i}{4}\right)\right|^2}{\sqrt{1-z}} {}_2F_1\left[(1+i)/4, (1-i)/4\,;1/2\,;(2z-1)^2\right].$$

When $\beta \neq 1/2$, (Pawula and Rice 1986) instead provides recurrence formulae for the coefficients of the series expansion of f_Z.

Mazza and Piau: $\tau \sim \Gamma(2, \beta)$

In a veritable *tour de force*, Mazza and Piau investigated the properties of a certain kind of iterated random function (Mazza and Piau 2001), and in doing so obtained several results for RC filters of the kind we are interested in. Here we note only their most relevant, explicit results which extend those given above for $\Gamma(2, \beta)$-distributed waiting times.

First, they show how to find an explicit solution of Eq. (7.14), which are linear com-

binations of functions

$$[4z(1-z)]^{-n}\,{}_3F_2[a_1, a_2, a_3\,; b_1, b_2\,; 4z(1-z)]$$

for n satisfying an indicial equation which in this case gives $n = 0, 1-\beta$.

When $\beta = 1$, all the exponents of the indicial equation coincide and they show that f_Z has the form

$$f_Z(z) = c_1 \left({}_3F_2 \left[\begin{matrix} 2a, & 2b, & \frac{1}{2} \\ & 1, & 1 \end{matrix} ;\, z \right] + {}_3F_2 \left[\begin{matrix} 2a, & 2b, & \frac{1}{2} \\ & 1, & 1 \end{matrix} ;\, 1-z \right] \right)$$
$$+ c_2\, {}_2F_1 \left[\begin{matrix} a, & b \\ & 1 \end{matrix} ;\, z \right] {}_2F_1 \left[\begin{matrix} a, & b \\ & 1 \end{matrix} ;\, 1-z \right],$$

where

$$a := \frac{1-i}{4}, \qquad b := \frac{1+i}{4}.$$

7.3 Discussion

As far as we are aware, the Fuchsian property of the multistate Kramers-Moyal equations for Markov chain multistate models has not yet been explored, but given the insights gained here with only minimal tools from the theory of Fuchsian systems and equations, the subject merits deeper investigations. Beyond the obvious advantage of obtaining solutions for more Markov chain multistate models, the Fuchsian properties can assist with understanding potential qualitative behaviour of the solutions. For example, Zhang *et al.* 2012 searched for multimodal behaviour in $P(n)$ for the 2-cycle, whereas we know that the mixing density f_{X_t} for any M-cycle has only two finite singularities, and therefore $P(n)$ can have a maximum of two (non-adjacent) modes. The potential insights that could be gained from applications of more sophisticated Fuchsian theorems remain to be uncovered.

Similarly, in Section 7.2 we drew upon on the associated results from just two different fields to obtain some solutions for multistate models with non-exponential waiting times in each gene state. Given the vast number of fields for which a random variable X_t satisfies a first-order random differential equation, there are likely more results for the taking.

Chapter 8

Quantification of timescales

As mentioned in Chapters 1 and 2, it is widely accepted that changes in promoter architecture, or other kinds of "gene state switching", cause the transcription rate in single cells to change stochastically in time (Sánchez *et al.* 2013). Multistate promoter models describe these gene state transitions explicitly, giving rise to several timescales: *(i)* the degradation rate of the mRNA molecules λ, which we are assuming here to be independent of the gene state, *(ii)* the transcription rates μ_i, $i = 1, \ldots, L$, one for each gene state, and *(iii)* the transition rates k_{ij}, $i, j = 1, \ldots, L$, $i \neq j$. The dynamical behaviour of the mRNA copy number in single cells, and the properties of the solution $P(n)$ of the master equation for the model depend on the balance of these timescales relative to each other.

Parameter regimes with timescale separations between the gene state switching rates and the transcription rates have been well studied. First, if gene state switching is sufficiently slow the copy number distribution $P(n)$ can be multimodal. For some organisms, each mode can correspond to a different phenotypic state (Choi *et al.* 2008; Gupta *et al.* 2011; Ozbudak *et al.* 2004). It has been shown that stochastic transitioning between multiple phenotypic states can be an effective survival strategy for cell populations in fluctuating or unpredictable environments (Balaban *et al.* 2004; Kussell and Leibler 2005; Acar *et al.* 2005; Thattai and Oudenaarden 2004). On the other hand, in other cases timescales of gene switching are often assumed to be in "rapid pre-equilibrium" (Ge *et al.* 2015), i.e., much faster than the timescales for transcription, to take advantage of the simplifications that follow for the mathematical analysis of a model (Shahrezaei and Swain 2008; Hu *et al.* 2011; Thomas *et al.* 2012).

However, true understanding and practical applications of results under timescale separations suffer from two limitations. First, there is no exact quantification or measure for "fast" or "slow" switching regimes, so it is difficult to judge how applicable results obtained under such assumptions are for real data. Second, timescale assumptions are usually based on mean waiting times, and there has been little investigation into the effects of the waiting time *distribution* on parameter regime estimates. It is plausible that these effects could be important in intermediate switching regimes in particular. These issues are becoming more pertinent now that more and more time course data is emerging showing switching in intermediate switching regimes (Li and Xie 2011; Taniguchi *et al.* 2010; Choi *et al.* 2008; Gupta *et al.* 2011; Ozbudak *et al.* 2004), the appearance of non-exponential waiting times (Suter *et al.* 2011; Harper *et al.* 2011; Kandhavelu *et al.* 2012), and the emergence of exact solutions for models that are opening the door to the analysis of models with

non-exponential waiting times (Hornos *et al.* 2005; Huang *et al.* 2014; Stinchcombe *et al.* 2012; Zhang *et al.* 2012).

We will address the first limitation using the mean-first-passage time to reach one stable state from another to characterise parameter regimes quantitatively (Szabo *et al.* 1980). We then apply our results to parameter sets that have been inferred from published experiments in gene expression (Golding *et al.* 2005; Zenklusen *et al.* 2008; Suter *et al.* 2011; Molina *et al.* 2013; Zoller *et al.* 2015), and show that most genes fall within intermediate regimes where approximations based on timescale separations are invalid. Finally, we make some preliminary remarks on the effects of waiting time distributions, based in part on our observations from our classification of the parameter sets from published experiments.

8.1 The mean-first-passage time between stable states

The dynamical responses of multistate models can be understood as follows. An L-state promoter model is a hybrid dynamical system with steady states corresponding to distinct values of $\{\mu_i/\lambda\}_{i=1}^{L}$. When there is a state transition, there is a 'transience period' during which the probability distribution of the mRNA evolves in time approaching the new steady state. The dominance of this transient behavior thus depends on the amount of time spent in the transience periods, compared to the time spent in each gene state before a switch occurs. In 'slow switching' regimes, for example, the transience time is negligible compared to the waiting time in each state, whereas in 'fast switching' regimes the transience time is far longer than the waiting times in the gene states, so the steady states are never reached.

In order to quantitatively compare parameter regimes and distinguish between 'fast', 'slow' and 'intermediate' promoter switching, we need to first define the characteristic times of the transients. For simplicity, assume we have an inactive (OFF) state with no transcription, and an active (ON) state with constant transcription rate μ. The degradation rate is a constant, λ, independently of gene state. At steady state in the active (resp. inactive) gene state, the expected value of the mRNA copy number is μ/λ (resp. zero). Hence, we define the time it takes for a sample path to transition from the inactive steady state to the active steady state by its first passage time to $\lceil \frac{\mu}{\lambda} \rceil$ when starting at zero (in the absence of gene switches). These transition times are outcomes from a random variable that we denote by T_{up}. Similarly, the random variable T_{down} denotes the time it takes for a sample path of the mRNA copy number to reach zero when the gene is inactive, having started with $\lceil \frac{\mu}{\lambda} \rceil$ transcripts. See Fig. 8.1 for an example outcome of (a) T_{up} and (b) T_{down}. We characterise the transient times by the mean-first-passage times $\mathbb{E}(T_{\text{up}})$ and $\mathbb{E}(T_{\text{down}})$.

Notice that $\mathbb{E}(T_{\text{up}})$ and $\mathbb{E}(T_{\text{down}})$ are completely specified by the transcription and degradation rate constants μ and λ. In the following section we derive their exact expressions.

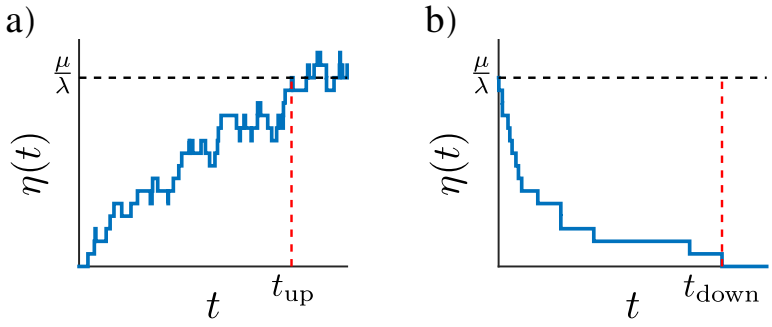

Figure 8.1: Characteristic timescales of the switching transients. *(a)* Outcome of the random variable T_{up}, which describes the time it takes for a sample path of the mRNA copy number to reach the value $\lceil \frac{\mu}{\lambda} \rceil$ while the gene is active, from starting value zero. *(b)* Outcome of the random variable T_{down}, which describes the time it takes for a sample path of the mRNA copy number to reach the value zero while the gene is inactive, from starting value $\lceil \frac{\mu}{\lambda} \rceil$.

8.2 Derivation of $\mathbb{E}(T_{\text{up}})$ and $\mathbb{E}(T_{\text{down}})$

Denote the mean of the active steady state by

$$m := \left\lceil \frac{\mu}{\lambda} \right\rceil .$$

We will use the theory of absorbing Markov chains to show that

$$\mathbb{E}(T_{\text{up}}) = \frac{1}{\mu} \sum_{i=1}^{m} \frac{1}{i} \frac{m!}{(m-i)!} \frac{1}{m^{i-1}},$$

$$\mathbb{E}(T_{\text{down}}) = \frac{1}{\lambda} \sum_{i=1}^{m} \frac{1}{i}.$$

Note that these two transient times are non-trivial functions of μ, λ, and $m := \lceil \mu/\lambda \rceil$.

Since T_{up} and T_{down} are calculated in the absence of gene state switching, $\eta : t \mapsto \eta(t)$ is a continuous-time absorbing Markov chain with $m+1$ states $\{0, 1, \ldots, m\}$. In the calculation of T_{up}, $\eta(0) = 0$ and has reached its absorbing state when $\eta(t) = m$, whereas in the calculation of T_{down}, $\eta(0) = m$ and has reached its absorbing state when $\eta(t) = 0$. The Markov process can be represented by the infinitesimal generator:

$$Q = \begin{pmatrix} S & \mathbf{S_0} \\ \mathbf{0} & 0 \end{pmatrix},$$

where S is an $m \times m$ matrix such that $S_{ii} < 0$ for $i = 1, \ldots, m$, $S_{ij} \geq 0$ for $i \neq j$, and $S\,\mathbf{1}_{m\times 1} + \mathbf{S_0} = \mathbf{0}_{m\times 1}$. We wish to calculate the expected time until absorption in the state represented by the $(m+1)^{\text{st}}$ row. The time T to absorption is phase-type distributed, with probability distribution corresponding to the initial probability vector $(\boldsymbol{\alpha}_{1\times m}, \alpha_{m+1})$ given by (Neuts 1981):

$$F(t) = 1 - \boldsymbol{\alpha} \exp(St)\mathbf{1}, \quad t \geq 0.$$

The expectation is thus given by

$$\mathbb{E}(T) = -\boldsymbol{\alpha} S^{-1} \mathbf{1}_{m\times 1}.$$

In our case, $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$ so we need only calculate the sum of the first row of $-S^{-1}$ to obtain the result.

To avoid confusion, in the calculation of $\mathbb{E}(T_{\text{up}})$ (resp. $\mathbb{E}(T_{\text{down}})$) we denote the matrix S by S_{up} (resp. S_{down}).

Inversion of S_{down}^{-1} to obtain $\mathbb{E}(T_{\text{down}})$

For a Markov chain with absorbing state 0 represented by the $(m+1)^{\text{st}}$ row of the infinitesimal generator Q, S_{down}^{-1} is

$$S_{\text{down}} = \begin{pmatrix} -m\lambda & m\lambda & 0 & & & & \\ 0 & -(m-1)\lambda & (m-1)\lambda & & & & \\ & 0 & -(m-2)\lambda & (m-2)\lambda & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & 0 & -2\lambda & 2\lambda \\ & & & 0 & 0 & -\lambda \end{pmatrix}.$$

T_{down} is given by the sum of the first row of $-S_{\text{down}}^{-1}$.

To invert S_{down}, write it in the form $S_{\text{down}} = -\lambda D(I - N)$, where $D = \text{diag}\{m, m-1, \ldots, 1\}$, I is the $m \times m$ identity matrix, and N is a nilpotent matrix with ones along the upper diagonal and zeros everywhere else, i.e. $N_{i,i+1} = 1$ for $i = 1, \ldots, m-1$, and $N_{ij} = 0$ otherwise. Then

$$\begin{aligned} S_{\text{down}}^{-1} &= [-\lambda D(I-N)]^{-1} \\ &= -\frac{1}{\lambda}\left(I + \sum_{k=1}^{m-1} N^k\right) D^{-1} \\ &= -\frac{1}{\lambda}\begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & 1 & 1 & \ldots & 1 \\ 0 & 0 & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{m} & & & & \\ & \frac{1}{m-1} & & & \\ & & \ddots & & \\ & & & \frac{1}{2} & \\ & & & & 1 \end{pmatrix}. \end{aligned}$$

The expression for $\mathbb{E}(T_{\text{down}})$ follows immediately:

$$\mathbb{E}(T_{\text{down}}) = \frac{1}{\lambda}\sum_{i=1}^{m}\frac{1}{i}.$$

Inversion of S_{up}^{-1} to obtain $\mathbb{E}(T_{\text{up}})$

For a Markov chain with absorbing state m represented by the $(m+1)^{\text{st}}$ row of the infinitesimal generator Q, S_{up} is given by

$$S_{\text{up}} = \begin{pmatrix} -\mu & \mu & & & & & \\ \lambda & -\mu-\lambda & \mu & & & & \\ & 2\lambda & -\mu-2\lambda & \mu & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & (m-2)\lambda & -\mu-(m-2)\lambda & \mu & \\ & & & & (m-1)\lambda & -\mu-(m-1)\lambda & \end{pmatrix}$$

To invert S_{up} we use the results from Usmani 1994 to calculate each element of the first row of $[\sigma_{ij}] \coloneqq S_{\text{up}}^{-1}$.

Denote the lower diagonal elements of S_{up} by a_i, the diagonal elements by b_i and the upper diagonal elements by c_i. We have

$$\begin{aligned} a_i &= (i-1)\lambda, \\ b_i &= -\mu-(i-1)\lambda, \\ c_i &= \mu. \end{aligned}$$

Let Θ_n denote the $n \times n$ submatrix formed from the first n rows and columns of S_{up}, $\theta_n \coloneqq \det(\Theta_n)$, and define the sequence $\{\phi_i\}_{i=1}^{n}$ as follows:

$$\begin{aligned} \phi_i &= b_i\phi_{i+1} - a_{i+1}c_i\phi_{i+2}, \quad i = n, n-1, \ldots, 2, 1, \\ \phi_{n+1} &= 1, \\ \phi_{n+2} &= 0. \end{aligned}$$

It can easily be shown by induction that $\theta_n = (-\mu)^n$ for $n = 0, 1, \ldots, m$. So, using the identity

$$\sigma_{1j} = (-1)^{j+1}c_1c_2\ldots c_{j-1}\frac{\phi_{j+1}}{\theta_n}, \qquad j \neq 1,$$

(Eq. (4.3) in Usmani 1994) directly gives the j^{th} element of the first row of S_{up}^{-1}:

$$\begin{aligned} \sigma_{1j} &= \frac{(-1)^{j+1}\mu^{j-1}\phi_{j+1}}{(-\mu)^m} \\ &= -\frac{1}{\mu}\frac{(-\mu)^j\phi_{j+1}}{(-\mu)^m}. \end{aligned}$$

Now define

$$F_j \coloneqq \frac{(-\mu)^j\phi_{j+1}}{(-\mu)^m}.$$

Then $\mathbb{E}\left(T_{\text{up}}\right) = (1/\mu)\sum_{j=1}^{m} F_j$, and using the identity

$$\theta_k \phi_{k+1} - a_{k+1} c_k \theta_{k-1} \phi_{k+2} = \theta_n, \qquad k = n, n-1, \ldots, 1,$$

(Lemma 2 in Usmani 1994) we have the recurrence relation

$$F_j = 1 + \frac{j}{m} F_{j+1}; \qquad F_m = 1.$$

It can then be verified that

$$F_j = \frac{1}{(j-1)!} \sum_{s=0}^{m-j} \frac{(j+s-1)!}{m^s},$$

so that

$$\begin{aligned}
\sum_{j=1}^{m} F_j &= \sum_{j=1}^{m} \frac{1}{(j-1)!} \sum_{s=0}^{m-j} \frac{(j+s-1)!}{m^s} \\
&= \sum_{s=0}^{m-1} \frac{1}{m^s} \sum_{j=1}^{m-s} \frac{(j+s-1)!}{(j-1)!} \\
&= \sum_{i=1}^{m} \frac{1}{m^{i-1}} \sum_{j=1}^{m+1-i} \frac{(i+j-2)!}{(j-1)!} \\
&= \sum_{i=1}^{m} \frac{1}{m^{i-1}} \sum_{j=1}^{m+1-i} (i+j-2)^{\underline{i-1}}, \qquad (8.1)
\end{aligned}$$

where $(x)^{\underline{i}} \coloneqq x(x-1)\ldots(x-i+1) = x!/(x-m)!$ is the falling factorial.

Now using the recurrence relation $i(x)^{\underline{i-1}} = (x+1)^{\underline{i}} - (x)^{\underline{i}}$ (Roman 1984), we have

$$\begin{aligned}
\sum_{j=1}^{m+1-i} (i+j-2)^{\underline{i-1}} &= (i-1)^{\underline{i-1}} + \frac{1}{i} \sum_{j=2}^{m+1-i} \left[(i+j-1)^{\underline{i}} - (i+j-2)^{\underline{i}} \right] \\
&= \frac{(i-1)!}{0!} - \frac{1}{i}\frac{i!}{0!} + \frac{1}{i}\frac{m!}{(m-i)!} \\
&= \frac{1}{i}\frac{m!}{(m-i)!}. \qquad (8.2)
\end{aligned}$$

Finally, using Eqs. (8.1) and (8.2) we have

$$\mathbb{E}\left(T_{\text{up}}\right) = \frac{1}{\mu}\sum_{j=1}^{m} F_j = \frac{1}{\mu}\sum_{i=1}^{m} \frac{1}{i}\frac{m!}{(m-i)!}\frac{1}{m^{i-1}},$$

as required.

8.3 Regime characterisation using expected waiting times

In order to characterise parameter regimes in terms of state switching timescales, we need to compare the characteristic transient times $\mathbb{E}(T_{\text{up}})$ and $\mathbb{E}(T_{\text{down}})$ with the waiting times in each state. The most straightforward way to characterise the waiting times is to use $\mathbb{E}(T_{\text{on}})$ and $\mathbb{E}(T_{\text{off}})$, the expected waiting times in the active and inactive states, respectively. For Markov chain multistate promoter models, the expected waiting times can usually be easily determined by the transition rates k_{ij}. For example, if k is the (constant) transition rate from the inactive to the active state, $T_{\text{off}} \sim \text{Exp}(1/k)$ so that $\mathbb{E}(T_{\text{off}}) = 1/k$. For the M-cycle we presented in Chapter 6, T_{off} is hypoexponentially distributed with $\mathbb{E}(T_{\text{off}}) = \sum 1/k_i$.

Slow switching regimes

If $\mathbb{E}(T_{\text{off}}) \gg \mathbb{E}(T_{\text{down}})$ and $\mathbb{E}(T_{\text{on}}) \gg \mathbb{E}(T_{\text{up}})$, or equivalently $\mathbb{E}(T_{\text{off}})/\mathbb{E}(T_{\text{down}}) \gg 1$ and $\mathbb{E}(T_{\text{on}})/\mathbb{E}(T_{\text{up}}) \gg 1$, the system displays *slow-switching* behavior, i.e., the system has sufficient time to reach the steady states at $N = 0$ and $N = \mu/\lambda$ when the promoter is inactive and active, respectively, and the average time spent in each promoter state before switching is longer than the transience periods where the system moves between the two steady states (Fig. 8.2a, top-right corner). Hence steady states around $N = 0$ and $N = \mu/\lambda$ are reached, leading to bimodal copy number distributions as seen in gene expression with stochastic transitioning between two phenotypic states, or transcriptional bursting with long silent periods.

In the limits $\mathbb{E}(T_{\text{off}})/\mathbb{E}(T_{\text{down}}) \to \infty$ and $\mathbb{E}(T_{\text{on}})/\mathbb{E}(T_{\text{up}}) \to \infty$, the dynamical behavior can be described as a mixture of the two stationary distributions $\text{Poi}(0) = \delta_{0,n}$ and $\text{Poi}(\mu/\lambda)$. The limiting distribution can easily be shown more rigorously using the results from Chapter 3. For example, the stationary distribution of the Markov chain two-state (random telegraph) model in the Poisson mixture form is (Raj *et al.* 2006)

$$P(n) = \int_0^1 \frac{e^{-\frac{\mu}{\lambda}z}}{n!} \left(\frac{\mu}{\lambda}z\right)^n f_Z(z)\, dz, \tag{8.3}$$

where

$$Z \sim \text{Beta}\left(\frac{k_{\text{on}}}{\lambda}, \frac{k_{\text{off}}}{\lambda}\right),$$

$$\text{i.e.} \qquad f_Z(z) = \frac{\Gamma\left(\frac{k_{\text{on}}}{\lambda} + \frac{k_{\text{off}}}{\lambda}\right)}{\Gamma\left(\frac{k_{\text{on}}}{\lambda}\right)\Gamma\left(\frac{k_{\text{off}}}{\lambda}\right)} z^{\frac{k_{\text{on}}}{\lambda}-1}(1-z)^{\frac{k_{\text{off}}}{\lambda}-1},$$

and k_{on} and k_{off} are the transition rates to the active and inactive states, respectively. The limits $\mathbb{E}(T_{\text{off}})/\mathbb{E}(T_{\text{down}}) \to \infty$ and $\mathbb{E}(T_{\text{on}})/\mathbb{E}(T_{\text{up}}) \to \infty$ correspond to the limits $k_{\text{on}} \to 0$ and $k_{\text{off}} \to 0$, and the Beta distribution becomes concentrated at 0 and 1 with weights $k_{\text{off}}/(k_{\text{on}} + k_{\text{off}})$ and $k_{\text{on}}/(k_{\text{on}} + k_{\text{off}})$, respectively. Substituting into Eq. (8.3), in

this slow-switching limit $P(n)$ becomes

$$\begin{aligned} P(n) &= \int_0^1 \frac{e^{-\frac{\mu}{\lambda}z}}{n!} \left(\frac{\mu}{\lambda}z\right)^n \left[\frac{k_{\text{off}}}{k_{\text{on}}+k_{\text{off}}}\delta(z) + \frac{k_{\text{on}}}{k_{\text{on}}+k_{\text{off}}}\delta(z-1)\right] dz \\ &= \frac{k_{\text{off}}}{k_{\text{on}}+k_{\text{off}}}\,\delta_{0,n} + \frac{k_{\text{on}}}{k_{\text{on}}+k_{\text{off}}}\frac{e^{-\frac{\mu}{\lambda}}}{n!}\left(\frac{\mu}{\lambda}\right)^n \\ &= \frac{\mathbb{E}\left(T_{\text{off}}\right)}{\mathbb{E}\left(T_{\text{on}}\right)+\mathbb{E}\left(T_{\text{on}}\right)}\,\delta_{0,n} + \frac{\mathbb{E}\left(T_{\text{on}}\right)}{\mathbb{E}\left(T_{\text{on}}\right)+\mathbb{E}\left(T_{\text{on}}\right)}\frac{e^{-\frac{\mu}{\lambda}}}{n!}\left(\frac{\mu}{\lambda}\right)^n, \end{aligned}$$

where δ is the Dirac delta function and $\delta_{0,n}$ is the Kronecker delta function.

Fast-switching regimes

If $\mathbb{E}\left(T_{\text{off}}\right)/\mathbb{E}\left(T_{\text{down}}\right) \ll 1$ and $\mathbb{E}\left(T_{\text{on}}\right)/\mathbb{E}\left(T_{\text{up}}\right) \ll 1$, the system has no time to reach the steady states before promoter switching occurs (Fig. 8.2a, bottom-left corner). We refer to this as the *fast-switching* regime, although it is sometimes referred to as *rapid pre-equilibrium* (Ge *et al.* 2015).

In the limiting case where $\mathbb{E}\left(T_{\text{off}}\right)/\mathbb{E}\left(T_{\text{down}}\right) \to 0$ and $\mathbb{E}\left(T_{\text{on}}\right)/\mathbb{E}\left(T_{\text{up}}\right) \to 0$, the distribution of N becomes a scaled Poisson distribution. Again using the random telegraph model as an example, $\mathbb{E}\left(T_{\text{off}}\right)/\mathbb{E}\left(T_{\text{down}}\right) \to 0$ and $\mathbb{E}\left(T_{\text{on}}\right)/\mathbb{E}\left(T_{\text{up}}\right) \to 0$ correspond to $k_{\text{on}}/\lambda \to \infty$ and $k_{\text{off}}/\lambda \to \infty$, so that the Beta distribution tends to a Dirac delta distribution centred at $k_{\text{on}}/(k_{\text{on}}+k_{\text{off}})$. Substitution into Eq. (8.3) leads to

$$\begin{aligned} P(n) &= \int_0^1 \frac{e^{-\frac{\mu}{\lambda}z}}{n!}\left(\frac{\mu}{\lambda}z\right)^n \delta\left(z - \frac{k_{\text{on}}}{k_{\text{on}}+k_{\text{off}}}\right) dz \\ &= \frac{e^{-\frac{\mu}{\lambda}\frac{k_{\text{on}}}{k_{\text{on}}+k_{\text{off}}}}}{n!}\left(\frac{\mu}{\lambda}\frac{k_{\text{on}}}{k_{\text{on}}+k_{\text{off}}}\right)^n, \end{aligned}$$

i.e. $N \sim \text{Poi}((\mu/\lambda)(k_{\text{on}}/(k_{\text{on}}+k_{\text{off}})))$.

Intermediate, transient-dominated regimes

Finally, for parameter sets leading to the intermediate regimes, i.e. $\mathbb{E}\left(T_{\text{off}}\right)/\mathbb{E}\left(T_{\text{down}}\right) \sim 1$ and $\mathbb{E}\left(T_{\text{on}}\right)/\mathbb{E}\left(T_{\text{up}}\right) \sim 1$, the system is dominated by transients: the promoter remains active for just long enough so that on average the sample paths of the mRNA copy number reach values close to or just past μ/λ, at which point the promoter switches to the inactive state for just long enough for them to return close to zero (Fig. 8.2a, centre). These are the regimes where transient dynamics become significant aspects of the system, and the full distribution $P(n)$ is required.

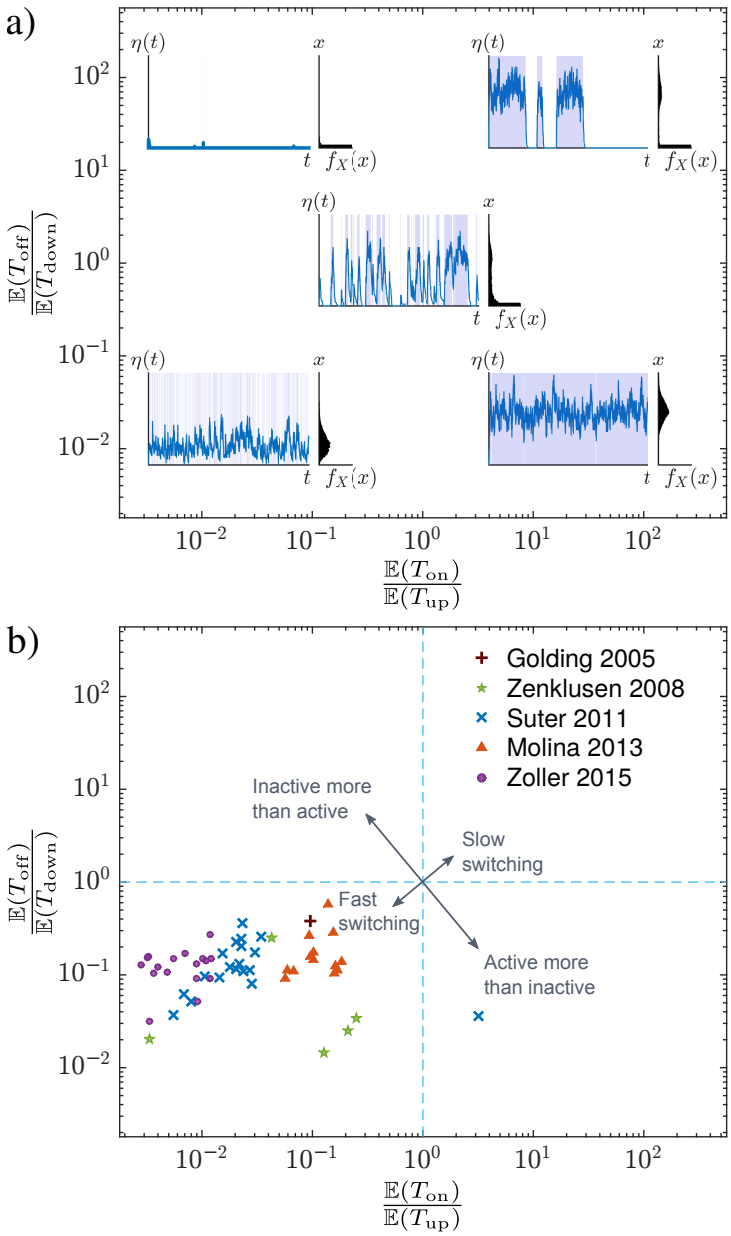

Figure 8.2: Classification of parameter regimes in a normalised space that quantitatively delineates between "fast" and "slow" switching dynamics. *(a)* Sample paths of the gene state switching and mRNA copy number, and corresponding time-average probability distributions for five illustrative points in the normalised regime space. Light blue shading indicates an active gene state. The horizontal axis measures the value of $\mathbb{E}(T_{\text{on}})$, the expected time spent in the active state, relative to $\mathbb{E}(T_{\text{up}})$, the mean-first-passage time for the mRNA copy number to transition from zero to $\lceil \frac{\mu}{\lambda} \rceil$. Similarly, the vertical axis measures the value of $\mathbb{E}(T_{\text{off}})$, the expected time spent in the inactive state, relative to $\mathbb{E}(T_{\text{down}})$, the mean-first-passage time for the mRNA copy number to transition from $\lceil \frac{\mu}{\lambda} \rceil$ to zero. Thus parameter regimes falling in the upper right-hand corner display slow promoter switching dynamics, and those in the lower left-hand corner display fast switching dynamics. Transient-dominated gene-state switching fall around the centre of the figure. *(b)* Characterisation of the parameter sets reported in published data sets. All but one of the points lie in the "fast"-switching quadrant of the regime space.

8.4 Quantification of parameter regimes from published experiments

When working with experimental data, it would be useful to characterise the parameter regime in order to judge whether approximate solutions of an ON-OFF model would suffice for the description of the system. For example, an important parameter regime is that of *instantaneous bursts*, where one can assume that all the mRNA molecules that are transcribed during an active period of the gene state appear at the same time. In this case the model can be described fully in terms of just three parameters: the 'burst frequency', the 'burst size', and the degradation rate (Cai *et al.* 2006; Raj *et al.* 2006; Dar *et al.* 2012). However, without a quantitative characterisation of the relevant timescales, parameter values are described in vague relational terms such as "when x is much larger than y" (Popović *et al.* 2015; Munsky *et al.* 2012; Raj and Oudenaarden 2009).

Using the characterisation that we described in the previous sections, we can quantitatively check where in parameter space real data lie, and whether the timescale assumptions that are often used to analyse them are valid. We have used parameter values $(k_{\text{on}} = 1/\mathbb{E}\left(T_{\text{off}}\right), k_{\text{off}} = 1/\mathbb{E}\left(T_{\text{on}}\right), \mu, \lambda)$ from the literature that were estimated by fitting the two-state random telegraph model to time-lapse data (Golding *et al.* 2005; Suter *et al.* 2011; Molina *et al.* 2013) and snapshot data (Zenklusen *et al.* 2008). We have also used the parameter values estimated by Zoller *et al.* 2015, where timeseries data from Suter *et al.* 2011 were fit to M-cycle models. The results are shown in Fig. 8.2b.

We can see that all but one of the parameter regimes lie in the lower left quadrant of the plane, corresponding to 'fast' switching between both states. Here we use 'fast' in the sense of the fast-switching paragraph in the previous section: the gene state switches over timescales that are too short for the mRNA numbers to reach either steady state around $N = 0$ or $N = \mu/\lambda$. The only gene in the lower right quadrant, where the gene state is active most of the time, is the NcKap1 gene[1] from Suter *et al.* 2011.

With the exception of the data point for the Golding *et al.* 2005 parameter values, which used data from *E. coli*, the other studies obtained data from eukaryotes. Interestingly, the parameter regime for the prokaryote imply that the gene was inactive more than it was active, and the switching occurred on slower timescales than reported for most of the other genes. As expected for eukaryotic genes, most of the data lie above the line $y = x$, indicating that the gene was inactive more than it was active. The genes furthest above the line $y = x$ are the ones that display the most "bursty" behaviour, where transcription only takes place during short, infrequent time intervals. However, none of the parameter regimes are in the upper quadrant, where the 'instantaneous bursts' regime would lie. Since none of the genes lie in the upper right quadrant, which would be indicative of slow switching, we can deduce immediately that none of the fitted models displayed bimodality. A quick check of the figures in the references (Golding *et al.* 2005; Suter *et al.* 2011; Molina *et al.* 2013; Zoller *et al.* 2015) confirms the validity of this deduction.

[1] However, when the same data was evaluated in Zoller *et al.* 2015 using the M-cycle rather than the two-state random telegraph model, the NcKap1 gene appears in a regime with short waiting times in the active state. We discuss this effect briefly in the following section.

8.5 Final remarks: Effects of waiting time densities

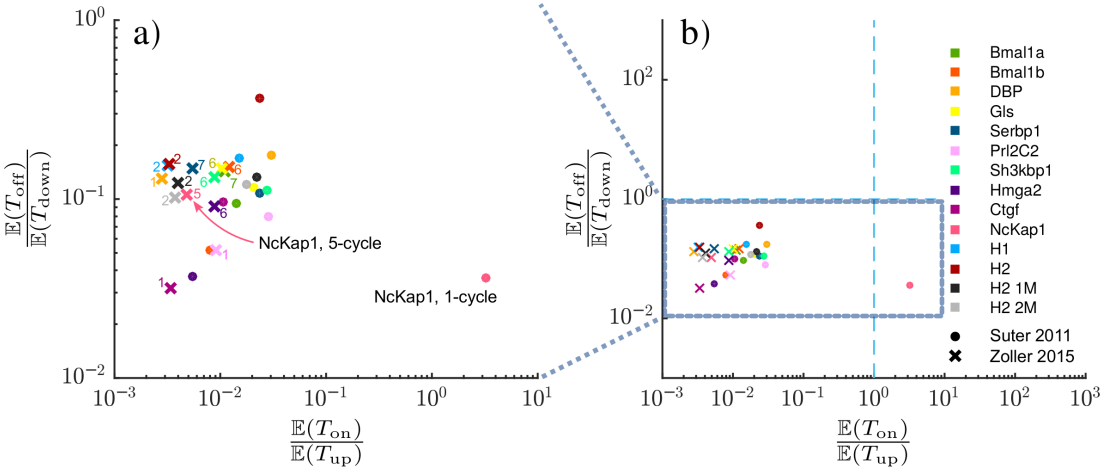

Figure 8.3: Comparison of parameter regimes implied by fitting the 1-cycle (in Suter *et al.* 2011) and the M-cycle (in Zoller *et al.* 2015) to the same data. Different colours correspond to different genes; circles correspond to results from Suter *et al.* 2011, and crosses correspond to results from Zoller *et al.* 2015. The legend is valid for both subfigures. *(a)* Zoomed-in view of the region in regime space where the scatter points lie to compare results from fitting data to models with different waiting time densities in each state. Circles show fitting to the 1-cycle, crosses show fitting to the M-cycle, where M is indicated by a number next to each cross. *(b)* All the scatter points lie in the lower-left region of regime space.

In Fig. 8.2 the expected waiting times were plotted to characterise regimes and make comparisons between different genes and datasets. Clearly, differences in experimental and parameter estimation procedures will bias comparisons between datasets, but it is interesting to see that despite using the same data, the parameter characterisation from the Suter *et al.* 2011 and Zoller *et al.* 2015 papers are markedly different (note the logarithmic scale). The difference between the parameter estimation for the two studies is that in Suter *et al.* 2011 all data parameters were fit to the random telegraph model (1-cycle), whereas in Zoller *et al.* 2015 both parameter estimation and model selection from M-cycles were performed. To allow for a more direct comparison, Fig. 8.3 includes only the 'Suter 2011' and 'Zoller 2015' scatter points from Fig. 8.2 where the same data was analysed in both papers. The number of OFF states used to model each gene in Zoller *et al.* 2015 are indicated next to each data point in subfigure (a).

Notwithstanding differences in estimation procedures, one might expect the regime characterisations to be different since here we have only considered the expected waiting times $\mathbb{E}(T_{\text{on}})$ and $\mathbb{E}(T_{\text{off}})$. Recalling how the "1-cycle equivalent" model (where for the 5-cycle we used the effective rate $1/k_{\text{off}} = \sum 1/k_i$) performed so poorly in the MLE estimation Section 6.3, perhaps the effects are similar: when fitting to the 1-cycle some parameters overcompensate the others. Indeed, Suter *et al.* 2011 found that parameter estimation overestimated k_{off} and underestimated k_{on} when fitting the 1-cycle to 2-cycle

data, even when fitting using time-series rather than snapshot data (Suter *et al.* 2011, supplementary material).

A possible cause is that any property other than the expected waiting times are lost. For example, the waiting time in the inactive state is exponential for the 1-cycle, with zero mode, whereas the total waiting time in the inactive state τ_{off} for the M-cycle is hypoexponentially distributed (Neuts 1981), and can be peaked away from zero. Furthermore, the coefficient of variation of an exponentially distributed random variable is equal to one, whereas for a hypoexponentially distributed random variable it is less than one (Neuts 1981).

Estimated parameter values and the corresponding timescales form an essential component of data analysis of gene expression since we can not observe all the processes at work. With increasing evidence to show that the dynamical properties of waiting time distributions can not be accounted for by the random telegraph model (1-cycle), investigations into the effects of waiting time distributions on parameter estimation is becoming more pertinent. The timescale quantification introduced in this chapter, combined with model fitting using time-lapse data with consistent estimation procedures, could be a step forward in this direction.

Part IV

Further directions and discussion

Part IV: Further directions and discussion

Introduction

Having derived and evaluated the Poisson mixture result in Part II, and utilised it more systematically in Part III, this part can be viewed as a two-part discussion of the potential gains of working within the Poisson mixture framework, how it fits within previous work which followed similar courses, and future directions to extend the scope of the work.

Chapter 9 shows how we can extend some of our approaches to be applicable to certain types of feedback model. Suggestions are given for how we can compare and contrast the initial results with what we know about non-feedback models, to make progress towards greater understanding and facility with feedback models. Chapter 10 concludes the thesis with a short summary.

Chapter 9

Further directions

Several sequential steps, requiring various biomolecules, are involved in the transcription of a gene. Mathematically, this means that a realistic model of gene transcription can not simply assume constant transcription and degradation rates. Indeed, the resulting Poisson distributions fail to describe much of the gene expression data that has been collected from a broad range of species, and thousands of genes. For this reason, we derived the results of Chapter 3 starting from a theoretical framework that allows the transcription and degradation rates in the population to have general, time-dependent distributions.

However, in addition to the process of transcription itself, it is becoming increasingly clear that gene expression is controlled by large regulatory networks (Lee *et al.* 2002; Austin *et al.* 2006). Autoregulation of gene expression has been identified in many cellular processes (Becskei and Serrano 2000; Thattai and Oudenaarden 2001; Rosenfeld *et al.* 2002; Austin *et al.* 2006; Süel *et al.* 2007), and is an important control mechanism for the cell. The derivation of the Poisson mixture result does not account for transcription and degradation rates that depend on the current state of the system itself, i.e., the number of mRNA transcripts present in the cell. However, it was recently shown that a simple autoregulation model has a Poisson mixture solution (Iyer-Biswas and Jayaprakash 2014). In light of this, the purpose of this section is to discuss some of the connections between feedback models and Poisson mixtures. Several papers presenting the solution of a certain type of autoregulation model with first-order feedback terms have appeared recently (Hornos *et al.* 2005; Visco *et al.* 2008; Grima *et al.* 2012; Vandecan and Blossey 2013; Pendar *et al.* 2013; Kumar *et al.* 2014; Huang *et al.* 2014; Iyer-Biswas and Jayaprakash 2014), so we show how to derive differential equations for the mixing density for models with n^{th}-order feedback terms. Using the understanding we have about Poisson mixture solutions for non-feedback models presented in previous chapters, we mention the nuances of the Poisson mixture approach for feedback models compared to non-feedback models, and pose some questions that naturally follow for further investigation.

9.1 ON-OFF feedback models

Autoregulation models for transcription of a single gene usually take the form of a two-state, ON-OFF model where a feedback mechanism regulates the switching rates between the two gene states. There are two main justifications for this. First, transcription is usually assumed to be a zero-order reaction because the mRNA molecules are not directly involved in the transcription process, and the transcriptional 'machinery' – the number of

DNA molecules, and the number and location of promoter sites, for example – does not change. Second, autoregulation is often effectuated in an indirect manner, by activating or repressing the production of transcription factors or other proteins that are necessary for or promote transcription initiation (Navarro *et al.* 2012; Fidalgo *et al.* 2012). Mathematically, this autoregulatory effect can be described by allowing the propensities of the gene state switching to depend on the mRNA copy number (see Fig. 9.1).

Under the assumption of such ON-OFF feedback models, sample paths of the transcription rate are random binary processes. At stationarity, the waiting times in each gene state have some distribution, at least one of which will be non-exponential. Examples of ON-OFF models with non-exponential waiting times were studied in Part III under the Poisson mixture framework, where we could say *a priori* that we could find a mixing density that would completely specify the full distribution $P(n,t)$ for the mRNA copy number. It thus seems reasonable to assume that the feedback models have solutions that can be written in Poisson mixture form, and then search for the mixing densities. In fact, Gardiner and Chaturvedi 1977 proposed such an approach for handling any chemical master equation, including those for multi-molecular reactions and systems of several species (although they do not appear to have studied feedback models). They showed that any discrete probability distribution could be represented in what they termed the *Poisson representation*, as long as we allow the *quasiprobability distribution* (the equivalent of our mixing density f_X) to be complex.

Iyer-Biswas and Jayaprakash 2014 recently used the Poisson representation to obtain an exact solution for an ON-OFF autoregulation model with a linear feedback term. They found that the solution for the autoactivation model can always be written as a Poisson mixture, but the weight function for the autorepression model is complex for certain parameter regimes (see Section 9.4). When the weight function is complex, the mRNA distribution $P(n)$ can be *sub-Poissonian*, i.e., the Fano factor is less than one. As shown in Chapter 5, Poisson mixtures have a minimum Fano factor of 1, and therefore autorepression is able to minimize noise to an extent that is not possible for autoactivation or non-feedback models. The form of the weight function in this case is simple enough to infer the relevant parameter combinations that determine the shape of the distribution. As such, the authors were able to justify using the Poisson representation for these simple models that had already been solved for $P(n)$ explicitly several times (Hornos *et al.* 2005; Visco *et al.* 2008; Grima *et al.* 2012; Vandecan and Blossey 2013; Pendar *et al.* 2013; Huang *et al.* 2014).

Although it is extremely useful to be able to demarcate parameter space according to qualitatively different properties of the solution, in the context of the properties of Poisson mixtures presented in this thesis, it seems that far more could be gained from comparing the properties of the Poisson representation for feedback and non-feedback models. For example:

- We showed that all models without feedback can be written in Poisson-mixture form, where the mixing density f_{X_t} is a true density on the positive real line, and the extrinsic random variable X_t has clear physical interpretations. What, if anything, does the weight function physically represent for feedback models?

- Sample paths of the transcription rates for ON-OFF autoregulation models can produce sub-Poissonian mRNA distributions, which is not possible for models without feedback. This phenomenon is due to sufficient feedback control of the gene switch times in such a way as to reduce variability of the mRNA copy numbers. Consider the following thought experiment. Suppose we observe the cellular drives in each cell of a stationary population, which, due to feedback effects, has a sub-Poissonian probability distribution $P(n)$. Then feed the same drives to a cell population with identical initial conditions, but no capacity for feedback. The non-feedback population would still produce super-Poissonian behaviour. The density f_{X_t} of the extrinsic random variable X_t would be the same in both systems, so we can immediately deduce that the weight function is different to f_{X_t}. However, are f_{X_t} and the weight function related?

- Since the weight function for the autoactivation model with a linear feedback term is always a mixing density, can we work forwards to derive the Poisson mixture result for this model, and others with the same property, as we did for non-feedback models?

Answering these questions falls beyond the scope of this thesis. However, one way to work towards answering the first and second bullet points is to derive equations for the weight functions, and compare them to the multistate Kramers-Moyal equations we obtain for models without feedback. Since such a discussion was not included in the recent paper that gave the Poisson mixture solution for the model with a linear feedback term (Iyer-Biswas and Jayaprakash 2014), we take the first steps towards addressing the idea below. We show how to derive equations for the weight function for polynomial feedback terms, since the method was omitted from Iyer-Biswas and Jayaprakash 2014 and they only gave the equations for linear feedback. We will then make some comparisons between the equations for the weight functions, the forms of the weight functions themselves for these ON-OFF feedback models, and the mixing density for non-feedback models.

9.2 Obtaining equations for f_{X_t}

We consider two-state ON-OFF models where one or both of the gene switching rates is a function of the number of transcripts present in the cell. Denote the transition rate from ON to OFF (resp. OFF to ON) by the function $a(n)$ (resp. $b(n)$). The transcription rate constant is μ in the ON state, and zero in the OFF state, and the degradation rate constant is λ, irrespective of gene state (see Fig. 9.1a).

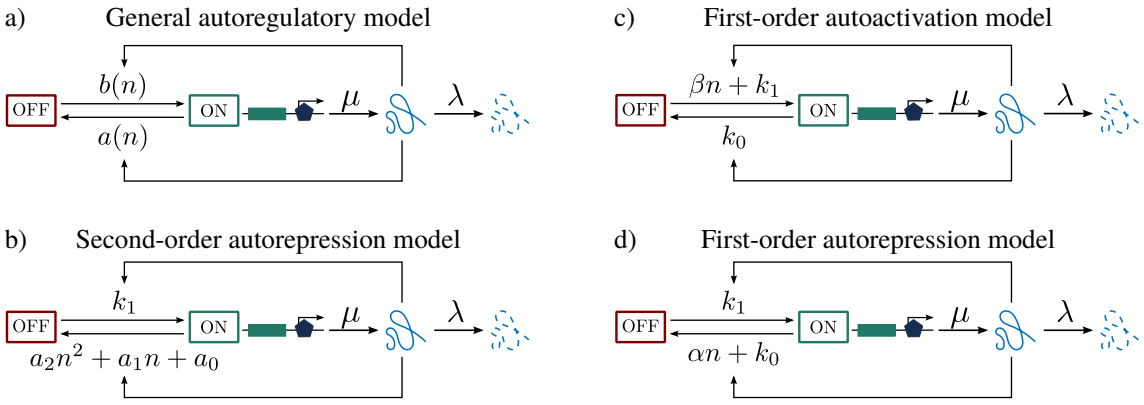

Figure 9.1: ON-OFF feedback models of gene transcription. The gene transitions between the active (ON) state and the inactive (OFF) state with rates that may depend on n, the number of mRNA molecules in the system. Transcription only takes place in the ON state, with rate μ. Degradation occurs as a first-order reaction with rate constant λ, independently of gene state. *(a)* General autoregulatory model where transition rates are undefined functions of n. *(b)* Autorepression model with second-order feedback terms. *(c)* Autoactivation model with linear feedback term. *(d)* Autorepression model with linear feedback term.

Similar to our approach when writing down multistate Kramers-Moyal equations for f_{X_t}, use the law of total probability to write

$$P(n,t) = P_{\text{on}}(n,t) + P_{\text{off}}(n,t),$$

where P_{on} and P_{off} denote the probability that there are $n \in \mathbb{N}$ mRNA molecules, and the gene is ON and OFF, respectively. The master equations for P_{on} and P_{off} are

$$\begin{aligned}\frac{P_{\text{on}}}{dt} &= \mu[P_{\text{on}}(n-1,t) - P_{\text{on}}(n,t)] + \lambda[(n+1)P_{\text{on}}(n+1,t) - nP_{\text{on}}(n,t)] \\ &\quad - a(n)P_{\text{on}}(n,t) + b(n)P_{\text{off}}(n,t) \qquad (9.1a)\\ \frac{P_{\text{off}}}{dt} &= \lambda[(n+1)P_{\text{off}}(n+1,t) - nP_{\text{off}}(n,t)] + a(n)P_{\text{on}}(n,t) - b(n)P_{\text{off}}(n,t). \qquad (9.1b)\end{aligned}$$

For models without feedback we can use the Poisson mixture result to infer that P_{on} and P_{off} can be written as Poisson mixtures, but for feedback models we must *assume* that we can write

$$\begin{aligned}P_{\text{on}}(n,t) &= \int \frac{x^n}{n!} e^{-x} f_{\text{on}}(x,t)\,dx, \qquad \text{and} \qquad (9.2a)\\ P_{\text{off}}(n,t) &= \int \frac{x^n}{n!} e^{-x} f_{\text{off}}(x,t)\,dx, \qquad (9.2b)\end{aligned}$$

where f_{on} and f_{off} are functions of the continuous variables x and t, not necessarily density functions. Thus we refer f_{on} and f_{off} as *weight functions* rather than mixing densities.

To obtain equations for the weight functions, we substitute the assumed expressions (9.2) for P_{on} and P_{off} into the master equations (9.1), and attempt to consolidate the weight

functions under the integral sign. Below we take each term that can appear in a master equation individually, and show how to consolidate the weight function. For clarity, we drop the subscripts and dependence on t for the purpose of these illustrations. Throughout we will use the fact that the weight functions have zero boundary conditions.

Degradation terms

Degradation terms always appear in the form $(n+1)P(n+1) - nP(n)$. We have

$$\begin{aligned}(n+1)P(n+1) - nP(n) &= \int \left[\frac{x^{n+1}}{n!}\, e^{-x}\, f(x) - \frac{x^n}{(n-1)!}\, e^{-x}\, f(x)\right] dx \\ &= \int \left[\frac{x^n}{n!}\, e^{-x} - \frac{x^{n-1}}{(n-1)!}\, e^{-x}\right] x f(x)\, dx \\ &= \int \frac{x^n}{n!}\, e^{-x}\, \frac{d}{dx}[x f(x)]\, dx, \end{aligned} \tag{9.3}$$

where we used integration by parts to obtain the final line.

Transcription terms

Transcription terms always appear in the form $P(n-1) - P(n)$. We have

$$\begin{aligned}P(n-1) - P(n) &= \int \left[\frac{x^{n-1}}{(n-1)!}\, e^{-x} - \frac{x^n}{n!}\, e^{-x}\right] f(x)\, dx \\ &= \int \frac{x^n}{n!}\, e^{-x}\, \frac{d}{dx}\left[-f(x)\right]\, dx,\end{aligned}$$

again using integration by parts.

Feedback terms

When $a(n)$ and $b(n)$ are linear combinations of $n(n-1)\ldots(n-k),\ k < n,\ k \in \mathbb{N}$ (and hence linear combinations of $n^i,\ i \in \mathbb{N}$), we can consolidate the weight functions so that they are independent of n using integration by parts. For linear feedback terms, we have

$$\begin{aligned}nP(n) &= \int \frac{x^{n-1}}{(n-1)!}\, e^{-x}\, x\, f(x)\, dx \\ &= \int \left[\frac{x^{n-1}}{(n-1)!}\, e^{-x} - \frac{x^n}{n!}\, e^{-x}\right] x\, f(x)\, dx + \int \frac{x^n}{n!}\, e^{-x}\, x\, f(x)\, dx \\ &= -\int \frac{x^n}{n!}\, e^{-x}\, \frac{d}{dx}\left[x\, f(x)\right]\, dx + \int \frac{x^n}{n!}\, e^{-x}\, x\, f(x)\, dx \\ &= \int \frac{x^n}{n!}\, e^{-x} \left(x\, f(x) - \frac{d}{dx}\left[x\, f(x)\right]\right) dx.\end{aligned}$$

For terms of the form $n(n-1)P(n)$, we have

$$n(n-1)P(n) = \int \frac{x^{n-2}}{(n-2)!}\, e^{-x}\, x^2\, f(x)\, dx$$

$$
\begin{aligned}
&= \int \left[\frac{x^{n-2}}{(n-2)!}\, e^{-x} - \frac{x^{n-1}}{(n-1)!}\, e^{-x} \right] x^2\, f(x)\, dx \\
&\qquad\qquad + \int \frac{x^{n-1}}{(n-1)!}\, e^{-x}\, x^2\, f(x)\, dx \\
&= \int \frac{x^{n-1}}{(n-1)!}\, e^{-x} \left(x^2\, f(x) - \frac{d}{dx}\left[x^2\, f(x) \right] \right) dx \\
&= \int \left[\frac{x^{n-1}}{(n-1)!}\, e^{-x} - \frac{x^n}{n!}\, e^{-x} \right] \left(x^2\, f(x) - \frac{d}{dx}\left[x^2\, f(x) \right] \right) dx \\
&\qquad\qquad + \int \frac{x^n}{n!}\, e^{-x} \left(x^2\, f(x) - \frac{d}{dx}\left[x^2\, f(x) \right] \right) dx \\
&= -\int \frac{x^n}{n!}\, e^{-x}\, \frac{d}{dx} \left(x^2\, f(x) - \frac{d}{dx}\left[x^2\, f(x) \right] \right) dx \\
&\qquad\qquad + \int \frac{x^n}{n!}\, e^{-x} \left(x^2\, f(x) - \frac{d}{dx}\left[x^2\, f(x) \right] \right) dx \\
&= \int \frac{x^n}{n!}\, e^{-x} \left\{ x^2\, f(x) - \frac{d}{dx}\left[x^2\, f(x) \right] \right. \\
&\qquad\qquad \left. - \frac{d}{dx} \left(x^2\, f(x) - \frac{d}{dx}\left[x^2\, f(x) \right] \right) \right\} dx.
\end{aligned}
$$

The process of iterated integration by parts yields expressions for terms of the form $n(n-1)\dots(n-k),\ k < n,\ k \in \mathbb{N}$, and hence any polynomial feedback terms can be deduced. For example, for an autoregression model with ON to OFF transition rate $a_2 n^2 + a_1 n + a_0$, we have

$$a(n) = a_2 n^2 + a_1 n + a_0 = a_2 n(n-1) + (a_2 + a_1) n + a_0,$$

so with OFF to ON transition rate k_1, transcription rate μ, and degradation rate parameter λ (Fig. 9.1b), so the equations for f_{on} and f_{off} are:

$$
\begin{aligned}
\frac{\partial f_{\text{on}}}{\partial t} &= -\frac{\partial}{\partial x}\left[(\mu - \lambda x) f_{\text{on}} \right] + k_1 f_{\text{off}} \\
&\quad - a_2 \left\{ x^2 f_{\text{on}} - \frac{\partial}{\partial x}\left[x^2 f_{\text{on}} \right] - \frac{\partial}{\partial x}\left(x^2 f(x) - \frac{\partial}{\partial x}\left[x^2 f_{\text{on}} \right] \right) \right\} \\
&\quad - (a_2 + a_1) \left\{ x f_{\text{on}} - \frac{\partial}{\partial x}\left[x f_{\text{on}} \right] \right\} - a_0 f_{\text{on}} \\
\frac{\partial f_{\text{off}}}{\partial t} &= -\frac{\partial}{\partial x}\left[-\lambda x f_{\text{off}} \right] - k_1 f_{\text{off}} \\
&\quad + a_2 \left\{ x^2 f_{\text{on}} - \frac{\partial}{\partial x}\left[x^2 f_{\text{on}} \right] - \frac{\partial}{\partial x}\left(x^2 f(x) - \frac{\partial}{\partial x}\left[x^2 f_{\text{on}} \right] \right) \right\} \\
&\quad + (a_2 + a_1) \left\{ x f_{\text{on}} - \frac{\partial}{\partial x}\left[x f_{\text{on}} \right] \right\} + a_0 f_{\text{on}},
\end{aligned}
$$

or, factorising:

$$\frac{\partial f_{\text{on}}}{\partial t} = -\frac{\partial}{\partial x}\left[(\mu - \lambda x) f_{\text{on}} \right] + k_1 f_{\text{off}}$$

$$- a_2 \left\{ \left(\frac{\partial}{\partial x} - 1 \right)^2 x^2 f_{\text{on}} \right\} + (a_2 + a_1) \left\{ \left(\frac{\partial}{\partial x} - 1 \right) x f_{\text{on}} \right\} - a_0 f_{\text{on}}$$

$$\begin{aligned} \frac{\partial f_{\text{off}}}{\partial t} = & -\frac{\partial}{\partial x} \left[-\lambda x f_{\text{off}} \right] - k_1 f_{\text{off}} \\ & + a_2 \left\{ \left(\frac{\partial}{\partial x} - 1 \right)^2 x^2 f_{\text{on}} \right\} - (a_2 + a_1) \left\{ \left(\frac{\partial}{\partial x} - 1 \right) x f_{\text{on}} \right\} - a_0 f_{\text{on}}. \end{aligned}$$

9.3 First order autoactivation model

Consider the ON-OFF autoactivation model with linear n-dependence in the transition rate from the inactive to the active state, i.e. $b(n) = \beta n + k_1$ (Fig. 9.1c). The transition rate from the active to the inactive state is a constant, k_0. The master equations for P_{on} and P_{off} are then

$$\begin{aligned} \frac{\partial P_{\text{on}}}{\partial t} = & \mu[P_{\text{on}}(n-1,t) - P_{\text{on}}(n,t)] + \lambda[(n+1)P_{\text{on}}(n+1,t) - nP_{\text{on}}(n,t)] \\ & - k_0 P_{\text{on}}(n,t) + (\beta n + k_1) P_{\text{off}}(n,t) \\ \frac{\partial P_{\text{off}}}{\partial t} = & \lambda[(n+1)P_{\text{off}}(n+1,t) - nP_{\text{off}}(n,t)] \\ & + k_0 P_{\text{on}}(n,t) - (\beta n + k_1) P_{\text{off}}(n,t), \end{aligned}$$

and using the expressions in the previous subsection for each term, we obtain the following equations for the weight functions:

$$\begin{aligned} \frac{\partial f_{\text{on}}}{\partial t} = & -\frac{\partial}{\partial x} \left[(\mu - \lambda x) f_{\text{on}} \right] - k_0 f_{\text{on}} + k_1 f_{\text{off}} + \beta x f_{\text{off}} - \frac{\partial}{\partial x} \left[\beta x f_{\text{off}} \right] \\ \frac{\partial f_{\text{off}}}{\partial t} = & -\frac{\partial}{\partial x} \left[-\lambda x f_{\text{off}} \right] + k_0 f_{\text{on}} - k_1 f_{\text{off}} - \beta x f_{\text{off}} + \frac{\partial}{\partial x} \left[\beta x f_{\text{off}} \right]. \end{aligned}$$

These are the multistate Kramers-Moyal equations for the usual non-feedback ON-OFF model, with the addition of the terms involving β.

As could be expected, the autoactivation term β increases the propensity of the gene state to switch from OFF to ON, in proportion to x. We also see from the $\frac{\partial}{\partial x} [\beta x f_{\text{off}}]$ term that feedback has an effect on f_{off} in the same way as degradation does; we can interpret this part of the feedback effect as an increase in the rate of degradation from λ to $\lambda + \beta$.

If one prefers, interpretation can be made on the system as written in the form

$$\begin{aligned} \frac{\partial f_{\text{on}}}{\partial t} + \frac{\partial}{\partial x} \left[(\mu - \lambda x) f_{\text{on}} \right] = & -k_0 f_{\text{on}} + k_1 f_{\text{off}} + \beta(x-1) f_{\text{off}} - \beta x \frac{\partial f_{\text{off}}}{\partial x} \\ \frac{\partial f_{\text{off}}}{\partial t} + \frac{\partial}{\partial x} \left[-\lambda x f_{\text{off}} \right] = & k_0 f_{\text{on}} - k_1 f_{\text{off}} - \beta(x-1) f_{\text{off}} + \beta x \frac{\partial f_{\text{off}}}{\partial x}. \end{aligned}$$

In this way we see that autoactivation adds

$$\begin{aligned} & -\beta x \tfrac{\partial f_{\text{off}}}{\partial x}, && \text{an advection term, and} \\ & \beta(x-1) f_{\text{off}}, && \text{a source term} \end{aligned}$$

to the equation for f_{on}, and similar interpretations can be made for f_{off}.

These coupled equations can easily be solved at stationarity (Iyer-Biswas and Jayaprakash 2014), to obtain for $f_X \coloneqq f_{\text{on}} + f_{\text{off}}$

$$f_X(x) = C\, e^{\frac{\beta}{\beta+\lambda}x} x^{\frac{k_1}{\beta+\lambda}-1} (\mu - \lambda x)^{\frac{k_0}{\beta+\lambda}-1},$$

where C is a normalisation constant. Comparison to the solution of the non-feedback (random telegraph) model, which has the form $\tilde{C}\, x^{k_1/\lambda-1}(\mu - \lambda x)^{k_0/\lambda-1}$, where $\tilde{C}$ is a constant, shows explicitly that the linear feedback effectively increased the "degradation rate" from λ to $\lambda+\beta$, as predicted. The increasing propensity to switch from OFF to ON when x is increasing manifests itself in the exponential term.

9.4 First order autorepression model

We can proceed in the same way to investigate the effects of adding a term αn to the rate of switching from ON to OFF, to produce a negative motif (Fig. 9.1d). With $a(n) = k_0 + \alpha n$, the equations for the weight functions are

$$\begin{aligned}
\frac{\partial f_{\text{on}}}{\partial t} &= -\frac{\partial}{\partial x}\left[(\mu - \lambda x) f_{\text{on}}\right] - k_0 f_{\text{on}} + k_1 f_{\text{off}} - \alpha x f_{\text{on}} + \frac{\partial}{\partial x}\left[\alpha x f_{\text{on}}\right] \\
\frac{\partial f_{\text{off}}}{\partial t} &= -\frac{\partial}{\partial x}\left[-\lambda x f_{\text{off}}\right] + k_0 f_{\text{on}} - k_1 f_{\text{off}} + \alpha x f_{\text{on}} - \frac{\partial}{\partial x}\left[\alpha x f_{\text{on}}\right],
\end{aligned}$$

and the solution for f_X at stationarity is

$$f_X(x) = C\, e^{\frac{\alpha}{\alpha+\lambda}x} x^{\frac{k_1}{\lambda}-1} (\mu - \lambda x)^{\frac{a\mu}{(\alpha+\lambda)^2} + \frac{k_0}{\alpha+\lambda} - \frac{\alpha k_1}{\lambda(\alpha+\lambda)} - 1}.$$

Interestingly, the effects of the feedback term are harder to interpret here: the exponential term still increases with x, the index of x is still $k_1/\lambda - 1$, not $k_1/(\lambda+\alpha) - 1$, and the index of $\mu - \lambda x$ is quite different to the form taken for the autoactivation model, and the random telegraph (non-feedback) model. In particular, the index of $\mu - \lambda x$ here can take values below -1, which is not the case for the autoactivation or non-feedback models.

According to Iyer-Biswas and Jayaprakash 2014, when the index of $\mu - \lambda x$ takes values below -1, $P(n)$ is sub-Poissonian so f_X has support on the complex plane. An explanation that aids some intuition and physical interpretation for such results could lead to significant benefits for this field; analytical investigations into several very similar models have been conducted (Hornos *et al.* 2005; Visco *et al.* 2008; Grima *et al.* 2012; Pendar *et al.* 2013; Kumar *et al.* 2014; Huang *et al.* 2014; Liu *et al.* 2015), which, individually, do not do much to further our tools or understanding for quantitative analysis of gene expression. The hope is that, in the same way as we were able to show how the results and methods presented in this thesis follow from working within a general framework, the same could be done for feedback models.

9.5 Further remarks

With the exception of Iyer-Biswas and Jayaprakash 2014, to our knowledge all other solutions for master equation feedback models have been obtained using the probability generating function (Hornos *et al.* 2005; Visco *et al.* 2008; Grima *et al.* 2012; Pendar *et al.* 2013; Kumar *et al.* 2014; Huang *et al.* 2014; Liu *et al.* 2015). We note that in the same way as the equations for f_{on} and f_{off} were derived here, we can also obtain corresponding equations for the probability generating functions, which, if solved, would give us an expression for $P(n)$ directly. However, our preference lies with the weight function in the interests of gaining physical intuition and further insights into the whole class of ON-OFF feedback models.

Chapter 10

Summary and discussion

This thesis set out to address two obstacles faced when using master equation models both to understand the process and regulation of gene transcription, and to make inferences from mRNA count data. First, few models have been solved analytically; even then the vast majority of those for which we do have solutions were obtained using the probability generating function at stationarity. Second, with the field quickly becoming data-rich, we need to move past a culture of only solving specific models one-by-one, so that our findings can be understood and placed within a wider context. These issues were addressed by first solving a transcription-degradation model under a general framework, allowing for non-stationarity of the processes involved. Only with the solution for the general model in hand, and its properties, did we then consider specific models. By doing so our understanding of these models was enhanced, leading to greater physical insight.

Under this framework, the Poisson mixture form of the solution of the general model was derived in Chapter 3, which then provides access to numerous analytical tools. This approach of solving the general model first, and then capitalising on its Poisson mixture form, brings with it two broad advantages. The first is more directly pragmatic: Since the mixing density is the only model-specific component of the Poisson mixture solution, we have identified another approach for solving models of this kind, namely, obtaining an expression for the mixing density f_{X_t}. As described in Chapter 4, the random variable X_t is continuous and satisfies a simple random differential equation that has been well-studied. We can thus draw upon the rich theory and results that already exist in the literature to help us calculate f_{X_t}, even for non-stationary models. If we are unable to determine f_{X_t} analytically, obtaining a numerical solution via stochastic simulations of X_t is far quicker than simulating the mRNA copy number directly, and standard ODE or PDE solvers give further options for solving the differential equations for f_{X_t}. As a result, solutions of several previously-unsolved models were derived in Chapters 4-7. In Chapter 5 we saw how the Poisson mixture form allows us to derive moment expressions in terms of the mixing density that are common to all gene transcription models, including the results for 'intrinsic' and 'extrinsic' noise (Swain *et al.* 2002), which simplify and explain analysis of noise in data.

The second broad advantage of this more general approach is more fundamental, more profound: By studying a model via the natural decoupling of the solution into its discrete, Poisson component and continuous, mixing component, we gain a far deeper intuition for, and understanding of, the model's structure and the properties of its solution. Through the

derivation of the Poisson mixture solution in Chapter 3, we understood that all extrinsic fluctuations, that is, uncertainty or variability in the transcription and degradation rates, take effect exclusively via a certain combination of the parameters that we used to define the extrinsic variable X_t. Moreover, in Chapter 4 we saw how X_t has clear physical interpretations. Both of these realisations clarify how extrinsic fluctuations combine to affect the final probability distribution of the mRNA copy number, and we start to gain some intuition for the sources of, and reasons for, solution behaviour and characteristics. Stripping the model down to its extrinsic component by solving for f_{X_t} provides us with additional understanding of the structure of a model. We utilised this understanding in Chapter 6 to navigate parameter space and identify regimes with qualitative characteristics of our choosing, and in Chapter 7 to deduce properties of models without first needing to solve them.

Continuing with the philosophy that biomathematicians should aim to develop tools and derive results that are applicable to whole *classes* of situations or models, and which aid comprehension of models or data within a wider context, Chapter 8 proposes a quantitative characterisation of timescales in the ubiquitous ON-OFF gene switching models. Parameter regimes can then be classified objectively in a normalised space that quantitatively delineates between "fast" and "slow" switching dynamics.

Finally, it would be remiss of me not to mention the "Poisson representation" for handling master equations (Gardiner and Chaturvedi 1977). While Gardiner and Chaturvedi recognised the utility of working within a Poisson representation framework, their intent was different: They showed that their Poisson representation is a basis for solutions of master equation models, if we allow the weight function (or quasiprobability distribution, in their nomenclature) to have support on the complex plane. They thus proposed to *assume* that a solution of the master equation can be written in terms of the Poisson representation, in order to derive the weight function and in so doing, obtain an exact solution of the master equation. On the other hand, we arrived at the Poisson mixture form organically for the transcription-degradation models in our framework, where the weight function is a true density with support on the positive real line and with strong physical interpretations, without making any prior assumptions.

Armed with our understanding of the Poisson mixture solution for transcription-degradation models without feedback, as discussed in this thesis, better comprehension of the Poisson representation for feedback models may well be forthcoming. Some preliminary results and initial questions to prompt investigations were discussed in Chapter 9, but a more complete comprehension of the transition from Poisson mixture solutions of non-feedback models to Poisson representations of the solutions of feedback models may well yield fruitful extensions of the results presented in this thesis.

Bibliography

Abramowitz, M. and I. A. Stegun (1964). *Handbook of Mathematical Functions*. Dover.

Acar, M., A. Becskei, and A. van Oudenaarden (2005). Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435, pp. 228–232.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19.6, pp. 716–723.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter (2008). *Molecular Biology of the Cell*. 5th ed. Garland Science.

Austin, D., M. Allen, J. McCollum, R. Dar, J. Wilgus, G. Sayler, N. Samatova, C. Cox, and M. Simpson (2006). Gene network shaping of inherent noise spectra. *Nature* 439, pp. 608–611.

Bahar Halpern, K., S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg, and S. Itzkovitz (2015). Bursty gene expression in the intact mammalian liver. *Molecular Cell* 58, pp. 147–156.

Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler (2004). Bacterial persistence as a phenotypic switch. *Science* 305.5690, pp. 1622–1625.

Bartholomay, A. F. (1958). Stochastic models for chemical reactions: I. Theory of the unimolecular reaction process. *The Bulletin of Mathematical Biophysics* 20.3, pp. 175–190.

Becskei, A. and L. Serrano (2000). Engineering stability in gene networks by autoregulation. *Nature* 405, pp. 590–593.

Beguerisse-Díaz, M., B. Wang, R. Desikan, and M. Barahona (2012). Squeeze-and-breathe evolutionary Monte Carlo optimization with local search acceleration and its application to parameter fitting. *Journal of the Royal Society Interface* 9.73, pp. 1925–1933.

Bengtsson, M., A. Ståhlberg, P. Rorsman, and M. Kubista (2005). Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Research* 15, pp. 1388–1392.

Bertsimas, D. and I. Popescu (2005). Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* 15.3, pp. 780–804.

Bertrand, E., P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long (1998). Localization of ASH1 mRNA particles in living yeast. *Molecular Cell* 2.4, pp. 437–445.

Berg, O. G. (1978). A model for the statistical fluctuations of the protein numbers in a microbial population. *Journal of Theoretical Biology* 71.4, pp. 587–603.

Beukers, F. (2009). Notes on differential equations and hypergeometric functions. `http://www.staff.science.uu.nl/~beuke106/HypergeometricFunctions/HGFcourse2009.pdf`. [Online; accessed 20-September-2015].

Bieler, J., R. Cannavo, K. Gustafson, C. Gobet, D. Gatfield, and F. Naef (2014). Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Molecular Systems Biology* 10.7.

Blake, W. J., M. Kærn, C. R. Cantor, and J. Collins (2003). Noise in eukaryotic gene expression. *Nature* 422, pp. 633–637.

Blake, W. J., G. Balázsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, and J. J. Collins (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular Cell* 24, pp. 853–865.

Brémaud, P. (2001). *Markov chains, Gibbs fields, Monte Carlo Simulation and Queues.* Springer.

Burger, R. (1999). Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* 153, pp. 1055–1069.

Cai, L., N. Friedman, and X. S. Xie (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature* 440, pp. 358–362.

Cao, Y., D. T. Gillespie, and L. R. Petzold (2006). Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics* 124, 044109.

Carey, L. B., D. van Dijk, P. M. A. Sloot, J. A. Kaandorp, and E. Segal (2013). Promoter sequence determines the relationship between expression level and noise. *PLoS Biol* 11.4, pp. 1–15.

Choi, P. J., L. Cai, K. Frieda, and X. S. Xie (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322.5900, pp. 442–446.

Chong, S., C. Chen, H. Ge, and X. S. Xie (2014). Mechanism of transcriptional bursting in bacteria. *Cell* 158, pp. 314–326.

Chubb, J. R., T. Trcek, S. M. Shenoy, and R. H. Singer (2006). Transcriptional pulsing of a developmental gene. *Current Biology* 16, pp. 1018–1025.

Corrigan, A. M. and J. R. Chubb (2014). Regulation of transcriptional bursting by a naturally oscillating signal. *Current Biology* 24, pp. 205–211.

Coulon, A., O. Gandrillon, and G. Beslon (2010). On the spontaneous stochastic dynamics of a single gene: Complexity of the molecular interplay at the promoter. *BMC Systems Biology* 4.2.

Coulon, A., C. C. Chow, R. H. Singer, and D. R. Larson (2013). Eukaryotic transcriptional dynamics: From single molecules to cell populations. *Nature Reviews Genetics* 14.8, pp. 572–584.

Coulon, A., M. L. Ferguson, V. de Turris, M. Palangat, C. C. Chow, and D. R. Larson (2014). Kinetic competition during the transcription cycle results in stochastic RNA processing. *eLife* 3, e03939.

Crandall, S. H. and W. Q. Zhu (1983). Random vibration: A survey of recent developments. *Journal of Applied Mechanics* 50, pp. 953–962.

Craster, R. V. and V. H. Hoàng (1998). Applications of Fuchsian differential equations to free boundary problems. *Proceedings of the Royal Society A* 454, pp. 1241–1252.

Dar, R. D., B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences* 109.43, pp. 17454–17459.

Davis, M. H. A. (1984). Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)* 46.3, pp. 353–388.

Denisov, S. I., W. Horsthemke, and P. Hänggi (2008). Steady-state Lévy flights in a confined domain. *Phys. Rev. E* 77, 061112 (6).

Denisov, S. I., W. Horsthemke, and P. Hänggi (2009). Generalized Fokker-Planck equation: Derivation and exact solutions. *The European Physical Journal B* 68.4, pp. 567–575.

Elowitz, M. B. and S. Leibler (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, pp. 335–339.

Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (2002). Stochastic gene expression in a single cell. *Science* 297, pp. 1183–1186.

Erdelyi, A. (1953). *Higher Transcendental Functions*. Vol. I. Bateman manuscript project. McGraw-Hill.

Fano, U. (1947). Ionization yield of radiations. II. The fluctuations of the number of ions. *Physical Review* 72.1, pp. 26–29.

Feillet, C., P. Krusche, F. Tamanini, R. C. Janssens, M. J. Downey, P. Martin, M. Teboul, S. Saito, F. A. Lévi, T. Bretschneider, G. T. J. van der Horst, F. Delaunay, and D. A. Rand (2014). Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. *Proceedings of the National Academy of Sciences* 111.27, pp. 9828–9833.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. Vol. I. Wiley Series in Probability and Mathematical Statistics. Wiley.

Femino, A. M., F. S. Fay, K. Fogarty, and R. H. Singer (1998). Visualization of single RNA transcripts in situ. *Science* 280.5363, pp. 585–590.

Fidalgo, M., F. Faiola, C.-F. Pereira, J. Ding, A. Saunders, J. Gingold, C. Schaniel, I. R. Lemischka, J. C. R. Silva, and J. Wang (2012). Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. *Proceedings of the National Academy of Sciences* 109.40, pp. 16202–16207.

Fitzhugh, R. (1983). Statistical properties of the asymmetric random telegraph signal, with applications to single-channel analysis. *Mathematical Biosciences* 64.1, pp. 75–89.

Forterre, P. (2010). Defining life: The virus viewpoint. *Origins of Life and Evolution of Biospheres* 40.2, pp. 151–160.

Francesconi, M. and B. Lehner (2014). The effects of genetic variation on gene expression dynamics during development. *Nature* 505, pp. 208–211.

Gardiner, C. W., K. J. McNeil, D. F. Walls, and I. S. Matheson (1976). Correlations in stochastic theories of chemical reactions. *Journal of Statistical Physics* 14.4, pp. 307–331.

Gardiner, C. W. and S. Chaturvedi (1977). The Poisson representation. I. A new technique for chemical master equations. *Journal of Statistical Physics* 17.6, pp. 429–468.

Gardiner, C. W. (1985). *Handbook of Stochastic Methods*. 2nd. Springer.

Ge, H., H. Qian, and X. S. Xie (2015). Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. *Physical Review Letters* 114, 078101.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* 22, pp. 403–434.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* 81, pp. 2340–2361.

Gillespie, D. T. (1992a). A rigorous derivation of the chemical master equation. *Physica A* 188, pp. 404–425.

Gillespie, D. T. (1992b). *Markov Processes: An Introduction for Physical Scientists*. Academic Press Inc.

Golding, I. and E. C. Cox (2004). RNA dynamics in live Escherichia coli cells. *Proceedings of the National Academy of Sciences* 101.31, pp. 11310–11315.

Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox (2005). Real-time kinetics of gene activity of individual bacteria. *Cell* 123, pp. 1025–1036.

Grima, R., P. Thomas, and A. V. Straube (2011). How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? *The Journal of Chemical Physics* 135, 084103.

Grima, R., D. R. Schmidt, and T. J. Newman (2012). Steady-state fluctuations of a genetic feedback loop: An exact solution. *The Journal of Chemical Physics* 137, 035104.

Grossmann, S. (1976). Langevin forces in chemically reacting multicomponent fluids. *The Journal of Chemical Physics* 65.5, pp. 2007–2012.

Gupta, P. B., C. M. Fillmore, G. Jiang, S. D. Shapira, K. Tao, C. Kuperwasser, and E. S. Lander (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146.4, pp. 633–644.

Haldane, J. B. S. (1932). The time of action of genes, and its bearing on some evolutionary problems. *The American Naturalist* 66.702, pp. 5–24.

Harper, C. V., B. Finkenstadt, D. J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashall, D. G. Spiller, J. J. Mullins, D. A. Rand, J. R. E. Davis, and M. R. H. White (2011). Dynamic analysis of stochastic transcription cycles. *PLoS Biology* 9.4.

Hilfinger, A. and J. Paulsson (2011). Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proceedings of the National Academy of Sciences* 108.29, pp. 12167–12172.

Hornos, J. E. M., D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes (2005). Self-regulating gene: An exact solution. *Physical Review E* 72.5, 051907.

Hornung, G., R. Bar-Ziv, D. Rosin, N. Tokuriki, D. S. Tawfik, M. Oren, and N. Barkai (2012). Noise–mean relationship in mutated promoters. *Genome Research* 22.12, pp. 2409–2417.

Horn, R. A. and C. R. Johnson (1999). *Matrix Analysis*. Cambridge University Press.

Hu, B., D. A. Kessler, W.-J. Rappel, and H. Levine (2011). Effects of input noise on a simple biochemical switch. *Physical Review Letters* 107.14, 148101.

Huang, L., Z. Yuan, P. Liu, and T. Zhou (2014). Feedback-induced counterintuitive correlations of gene expression noise with bursting kinetics. *Physical Review E* 90.5, 052702.

Ilyashenko, Y. and S. Yakovenko (2008). *Lectures on Analytic Differential Equations*. Vol. 86. Graduate Studies in Mathematics. American Mathematical Society.

Ishkhanyan, A. and K.-A. Suominen (2009). New solutions of Heun's general equation. arXiv:0909.1684 [math-ph].

Iyer-Biswas, S., F. Hayot, and C. Jayaprakash (2009). Stochasticity of gene products from transcriptional pulsing. *Physical Review E* 79.3, 031911.

Iyer-Biswas, S. and C. Jayaprakash (2014). Mixed Poisson distributions in exact solutions of stochastic autoregulation models. *Physical Review E* 90.5, 052712.

Jacob, F. and J. Monod (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal Of Molecular Biology* 3, pp. 318–356.

Jacob, F. and J. Monod (1962). On the regulation of gene activity. *Cold Spring Harbor Symposia on Quantitative Biology* 26, pp. 193–211.

Ji, N., T. C. Middelkoop, R. A. Mentink, M. C. Betist, S. Tonegawa, D. Mooijman, H. C. Korswagen, and A. van Oudenaarden (2013). Feedback control of gene expression variability in the *Caenorhabditis elegans* Wnt pathway. *Cell* 155.4, pp. 869–880.

Jones, D. L., R. C. Brewster, and R. Phillips (2014). Promoter architecture dictates cell-to-cell variability in gene expression. *Science* 346.6216, pp. 1533–1536.

Junker, J. P. and A. van Oudenaarden (2014). Every cell is special: Genome-wide studies add a new dimension to single-cell biology. *Cell* 157.1, pp. 8–11.

Kærn, M., T. C. Elston, W. J. Blake, and J. J. Collins (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature reviews* 6, pp. 451–464.

Kampen, N. G. van (1976). Fluctuations in continuous systems. *AIP Conference Proceedings* 27.1, pp. 153–186.

Kampen, N. G. van (1992). *Stochastic Processes in Physics and Chemistry*. 2nd. Elsevier.

Kandhavelu, M., A. Häkkinen, O. Yli-Harja, and A. S. Ribeiro (2012). Single-molecule dynamics of transcription of the lar promoter. *Physical Biology* 9.2, 026004.

Keizer, J. (1975). Concentration fluctuations in chemical reactions. *The Journal of Chemical Physics* 63.11, pp. 5037–5043.

Keizer, J. (1976). Dissipation and fluctuations in nonequilibrium thermodynamics. *The Journal of Chemical Physics* 64.11, pp. 1679–1687.

Keizer, J. (1977). Master equations, Langevin equations, and the effect of diffusion on concentration fluctuations. *The Journal of Chemical Physics* 67.4, pp. 1473–1476.

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov Chain. *The Annals of Mathematical Statistics* 24.3, pp. 338–354.

Kim, H. D. and E. K. O'Shea (2008). A quantitative model of transcription factor–activated gene expression. *Nature Structural & Molecular Biology* 15, pp. 1192–1198.

Ko, M. S. H., H. Nakauchi, and N. Takahashi (1990). The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO Journal* 9.9, pp. 2835–2842.

Ko, M. S. H. (1991). A stochastic model for gene induction. *Journal of Theoretical Biology* 153, pp. 181–194.

Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 7.4, pp. 284–304.

Kueh, H. Y., A. Champhekar, S. L. Nutt, M. B. Elowitz, and E. V. Rothenberg (2013). Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* 341.6146, pp. 670–673.

Kumar, N., T. Platini, and R. V. Kulkarni (2014). Exact distributions for stochastic gene expression models with bursting and feedback. *Physical Review Letters* 113, 268105.

Kuramoto, Y. (1975). Self-entrainment of a population of coupled non-linear oscillators. In: *International Symposium on Mathematical Problems in Theoretical Physics*. Vol. 39. Lecture Notes in Physics, pp. 420–422.

Kussell, E. and S. Leibler (2005). Phenotypic diversity, population growth and information in fluctuating environments. *Science* 309, pp. 2075–2077.

Langevin, P. (1908). Sur la théorie du mouvement brownien. *C. R. Acad. Sci. Paris* 146, pp. 530–533.

Larson, D. R., C. Fritzsch, L. Sun, X. Meng, D. S. Lawrence, and R. H. Singer (2013). Direct observation of frequency modulated transcription in single cells using light activation. *eLife* 2, e00750.

Lasserre, J. B. (2002). Bounds on measures satisfying moment conditions. *The Annals of Applied Probability* 12.3, pp. 1114–1137.

Lee, T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* 298, pp. 799–804.

Lestas, I., G. Vinnicombe, and J. Paulsson (2010). Fundamental limits of the suppression of molecular fluctuations. *Nature* 467, pp. 174–178.

Levine, M., C. Cattoglio, and R. Tjian (2014). Looping back to leap forward: Transcription enters a new era. *Cell* 157, pp. 13–25.

Li, G.-W. and X. S. Xie (2011). Central dogma at the single-molecule level in living cells. *Nature* 475, pp. 308–315.

Lin, Y., C. H. Sohn, C. K. Dalal, L. Cai, and M. B. Elowitz (2015). Combinatorial gene regulation by modulation of relative pulse timing. *Nature* 527, pp. 54–58.

Little, S. C., M. Tikhonov, and T. Gregor (2013). Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell* 154.4, pp. 789–800.

Liu, P., Z. Yuan, L. Huang, and T. Zhou (2015). Roles of factorial noise in inducing bimodal gene expression. *Physical Review E* 91.6, 062706.

Lück, S., K. Thurley, P. F. Thaben, and P. O. Westermark (2014). Rhythmic degradation explains and unifies circadian transcriptome and proteome data. *Cell Reports* 9, pp. 741–751.

Lück, S. and P. O. Westermark (2015). Circadian mRNA expression: Insights from modeling and transcriptomics. *Cellular and Molecular Life Sciences*, pp. 1–25.

Lyubimova, A., S. Itzkovitz, J. P. Junker, Z. P. Fan, X. Wu, and A. van Oudenaarden (2013). Single-molecule mRNA detection and counting in mammalian tissue. *Nature Protocols* 8, pp. 1743–1758.

Maier, R. S. (2007). The 192 solutions of the Heun equation. *Mathematics of Computation* 76.258, pp. 811–843.

Mazza, C. and D. Piau (2001). Dubins-Freedman processes and RC filters. *The Annals of Applied Probability* 11.4, pp. 1330–1352.

McFadden, J. A. (1959). The probability density of the output of an RC filter when the input is a binary random process. *IRE Transactions on Informatino Theory* 5.4, pp. 174–178.

McQuarrie, D. (1967). Stochastic approach to chemical kinetics. *Journal of Applied Probability*, pp. 413–478.

Miller, A. R. and R. B. Paris (2011). Euler-type transformations for the generalized hypergeometric function. *Zeitschrift für angewandte Mathematik und Physik* 62.1, pp. 31–45.

Milo, R. and R. Phillips (2015). *Cell Biology by the Numbers*. Garland Science.

Molina, N., D. M. Suter, R. Cannavo, B. Zoller, I. Gotic, and F. Naef (2013). Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proceedings of the National Academy of Sciences* 110.51, pp. 20563–20568.

Montroll, E. W. and K. E. Shuler (1957). *The Application of the Theory of Stochastic Processes to Chemical Kinetics*. Vol. 1. John Wiley & Sons, Inc., pp. 361–399.

Moyal, J. E. (1949). Stochastic Processes and Statistical Physics. *Journal of the Royal Statistical Society. Series B (Methodological)* 11.2, pp. 150–210.

Muller, H. J. (1922). Variation due to change in the individual gene. *The American Naturalist* 56, pp. 32–50.

Munsky, B. and M. Khammash (2006). The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics* 124, 044104.

Munsky, B., G. Neuert, and A. van Oudenaarden (2012). Using gene expression noise to understand gene regulation. *Science* 336.6078, pp. 183–187.

Murphy, S. A. and A. W. van der Vaart (2000). On profile likelihood. *Journal of the American Statistical Association* 95.450, pp. 449–465.

Muramoto, T., D. Cannona, M. Gierliński, A. Corrigan, G. J. Barton, and J. R. Chubb (2012). Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. *Proceedings of the National Academy of Sciences* 109.19, pp. 7350–7355.

Navarro, P., N. Festuccia, D. Colby, A. Gagliardi, N. P. Mullin, W. Zhang, V. Karwacki-Neisius, R. Osorno, D. Kelly, M. Robertson, and I. Chambers (2012). OCT4/SOX2-

independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. *The EMBO Journal* 31.24, pp. 4547–4562.

Neuts, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Dover Publications Inc.

Nevozhay, D., R. M. Adams, K. F. Murphy, K. Josić, and G. Bálazsi (2009). Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences* 106.13, pp. 5123–5128.

Nitzan, A., P. Ortoleva, J. Deutch, and J. Ross (1974). Fluctuations and transitions at chemical instabilities: The analogy to phase transitions. *The Journal of Chemical Physics* 61.3, pp. 1056–1074.

Novick, A. and M. Weiner (1957). Enzyme induction as an all-or-nothing phenomenon. *Proceedings of the National Academy of Sciences* 43, pp. 553–566.

Olson, E. J., L. A. Hartsough, B. P. Landry, R. Shroff, and J. J. Tabor (2014). Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nature Methods* 11, pp. 449–455.

Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden (2002). Regulation of noise in the expression of a single gene. *Nature Genetics* 31, pp. 69–73.

Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden (2004). Multistability in the lactose utilization of Esherichia coli. *Nature* 427, pp. 737–740.

Paszek, P. (2007). Modeling stochasticity in gene regulation: Characterization in the terms of the underlying distribution function. *Bulletin of Mathematical Biology* 69, pp. 1567–1601.

Paulsson, J. (2004). Summing up the noise in gene networks. *Nature* 427, pp. 415–419.

Paulsson, J. (2005). Models of stochastic gene expression. *Physics of Life Reviews* 2, pp. 157–175.

Pawula, R. and S. O. Rice (1986). On filtered binary processes. *IEEE Transactions on Information Theory* 32.1, pp. 63–72.

Pawula, R. (1967). Generalizations and extensions of the Fokker-Planck-Kolmogorov equations. *IEEE Transactions on Information Theory* 13.1, pp. 33–41.

Pawula, R. (1970). The transition probability density function of the low-pass filtered random telegraph signal. *International Journal of Control* 12.1, pp. 25–32.

Peccoud, J. and B. Ycart (1995). Markovian modeling of gene-product synthesis. *Theoretical Population Biology* 48.2, pp. 222–234.

Pedraza, J. M. and J. Paulsson (2008). Effects of molecular memory and bursting on fluctuations in gene expression. *Science* 319, pp. 339–343.

Pendar, H., T. Platini, and R. V. Kulkarni (2013). Exact protein distributions for stochastic models of gene expression using partitioning of Poisson processes. *Physical Review E* 87.4, 042720.

Plemelj, J. (1964). *Problems in the Sense of Riemann and Klein*. Vol. 16. Interscience Tracts in Pure and Applied Mathematics. Interscience Publishers.

Plyukhin, A. V. (2008). Generalized Fokker-Planck equation, Brownian motion, and ergodicity. *Physical Review E* 77.6, 061136.

Pontecorvo, G. (1968). Hermann Joseph Muller. *Annual Review of Genetics* 2, pp. 1–10.

Popović, N., C. Marr, and P. S. Swain (2015). A geometric analysis of fast-slow models for stochastic gene expression. *Journal of Mathematical Biology*, pp. 1–36.

Rainville, D. (1960). *Special Functions*. Macmillan.

Raj, A., C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biology* 4.10, pp. 1707–1719.

Raj, A. and A. van Oudenaarden (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135.2, pp. 216–226.

Raj, A. and A. van Oudenaarden (2009). Single-molecule approaches to stochastic gene expression. *Annual Review of Biophysics* 38, pp. 255–70.

Raser, J. M. and E. J. O'Shea (2004). Control of stochasticity in eukaryotic gene expression. *Science* 304.5678, pp. 1811–1814.

Raser, J. M. and E. K. O'Shea (2005). Noise in gene expression: Origins, consequences and control. *Science* 309, pp. 2010–2013.

Remenyi, A., H. Scholer, and M. Wilmanns (2004). Combinatorial control of gene expression. *Nature Structural and Molecular Biology* 11, pp. 812–815.

Riordan, J. (1937). Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions. *The Annals of Mathematical Statistics* 8.2, pp. 103–111.

Risken, H. (1989). *The Fokker-Planck equation - Methods of Solutions and Applications*. 2nd. Springer.

Rockman, M. V. and L. Kruglyak (2006). Genetics of global gene expression. *Nature Reviews Genetics* 7, pp. 862–872.

Rolski, T., H. Schmidli, V. Schmidt, and J. Teugels (2009). *Stochastic Processes for Insurance and Finance*. John Wiley & sons.

Roman, S. (1984). *The Umbral Calculus*. Dover Publications Inc.

Ronveaux, A. (1995). *Heun's Differential Equations*. Oxford University Press.

Rosenfeld, N., M. B. Elowitz, and U. Alon (2002). Negative autoregulation speeds up the response times of transcription networks. *Journal Of Molecular Biology* 323, pp. 785–793.

Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz (2005). Gene regulation at the single-cell level. *Science* 307, pp. 1962–1965.

Rybakova, K. N., F. J. Bruggeman, A. Tomaszewska, M. J. Moné, C. Carlberg, and H. V. Westerhoff (2015). Multiplex eukaryotic transcription (in)activation: Timing, bursting and cycling of a ratchet clock mechanism. *PLoS Computational Biology* 11.4, e1004236.

Sagan, C. (1979). Can we know the universe? Reflections on a grain of salt. In: *Broca's Brain*. Random House. Chap. 2.

Sánchez, A., H. G. Garcia, D. Jones, R. Phillips, and J. Kondev (2008). Effect of promoter architecture on the cell-to-cell variability in gene expession. *PLoS Computational Biology* 105.13, pp. 5081–5086.

Sánchez, A. and J. Kondev (2008). Transcriptional control of noise in gene expression. *Proceedings of the National Academy of Sciences* 105.13, pp. 5081–5086.

Sánchez, A., S. Choubey, and J. Kondev (2013). Regulation of noise in gene expression. *Annual Review of Biophysics* 42, pp. 469–491.

Sánchez, A. and I. Golding (2013). Genetic determinants and cellular constraints in noisy gene expression. *Science* 342, pp. 1188–1193.

Sanchez-Osorio, I., F. Ramos, P. Mayorga, and E. Dantan (2014). Foundations for modeling the dynamics of gene regulatory networks: A multilevel-perspective review. *Journal of Bioinformatics and Computational Biology* 12.1, 1330003.

Satija, R. and A. K. Shalek (2014). Heterogeneity in immune responses: From populations to single cells. *Trends in Immunology* 352.

Schwanhäusser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach (2011). Global quantification of mammalian gene expression control. *Nature* 473.7347, pp. 337–42.

Schwabe, A., K. N. Rybakova, and F. J. Bruggeman (2012). Transcription stochasticity of complex gene regulation models. *Biophysical Journal* 103.6, pp. 1152–1161.

Schrödinger, E. (1948). *What is life? The physical aspect of the living cell.* Based on lectures delivered under the auspices of the Institute at Trinity College, Dublin, in February 1943. Cambridge University Press.

Schwartz, D. (1962). Genetic studies on mutant enzymes in maize. III Control of gene action in the synthesis of pH 7.5 esterase. *Genetics* 47, pp. 1609–1615.

Senecal, A., B. Munsky, F. Proux, N. Ly, F. E. Braye, C. Zimmer, F. Mueller, and X. Darzacq (2014). Transcription factors modulate c-Fos transcriptional bursts. *Cell Reports* 8, pp. 75–83.

Shanin, A. V. and R. V. Craster (2002). Removing false singular points as a method of solving ordinary differential equations. *European Journal of Applied Mathematics* 13, pp. 617–639.

Shahrezaei, V. and P. S. Swain (2008). Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences* 105.45, pp. 17256–17261.

Shen-Orr, S. S., R. Milo, S. Mangan, and U. Alon (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* 31, pp. 64–68.

Shimoga, V., J. T. White, Y. Li, E. Sontag, and L. Bleris (2013). Synthetic mammalian transgene negative autoregulation. *Molecular Systems Biology* 9.670.

Sigal, A., R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, and U. Alon (2006). Variability and memory of protein levels in human cells. *Nature* 444, pp. 643–646.

Singer, Z. S., J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz (2014). Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells. *Molecular Cell* 55.2, pp. 319–331.

Slater, L. J. (1966). *Generalized Hypergeometric Functions*. Cambridge University Press.

Smaili, K., T. Kadri, and S. Kadry (2013). Hypoexponential distribution with different parameters. *Applied Mathematics* 4, pp. 624–631.

Smiley, M. W. and S. R. Proulx (2010). Gene expression dynamics in randomly varying environments. *Journal of Mathematical Biology* 61, pp. 231–251.

Smolen, P., D. A. Baxter, and J. H. Byrne (2000). Modeling transcriptional control in gene networks—methods, recent results, and future directions. *Bulletin of Mathematical Biology* 62.2, pp. 247–292.

Soong, T. T. (1973). *Random Differential Equations in Science and Engineering*. Vol. 103. Mathematics in Science and Engineering. Academic Press.

Soyfer, V. N. (2001). The consequences of political dictatorship for Russian science. *Nature Reviews Genetics* 2.3, pp. 723–729.

Spellman, P. T., G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher (1998). Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Molecular Biology of the Cell* 9, pp. 3273–3297.

Spudich, J. L. and D. E. Koshland Jr (1976). Non-genetic individuality: Chance in the single cell. *Nature* 262, pp. 467–471.

Ståhlberg, A., C. Thomsen, D. Ruff, and P. Aman (2012). Quantitative PCR analysis of DNA, RNAs, and proteins in the same single cell. *Clinical Chemistry*.

Stevense, M., T. Muramoto, I. Müller, and J. R. Chubb (2010). Digital nature of the immediate-early transcriptional response. *Development* 137, pp. 579–584.

Stinchcombe, A. R., C. S. Peskin, and D. Tranchina (2012). Population density approach for discrete mRNA distributions in generalized switching models for stochastic gene expression. *Physical Review E* 85.6, 061919.

Strogatz, S. H. (2000). From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena* 143.1–4, pp. 1–20.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98, pp. 1–4.

Süel, G. M., R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz (2007). Tunability and noise dependence in differentiation dynamics. *Science* 315, pp. 1716–1719.

Suprunenko, Y. F., P. T. Clemson, and A. Stefanovska (2013). Chronotaxic systems: A new class of self-sustained nonautonomous oscillators. *Physical Review Letters* 111.2, 024101.

Suter, D. M., N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332, pp. 472–474.

Swain, P. S., M. B. Elowitz, and E. D. Siggia (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences* 99.20, pp. 12795–12800.

Szabo, A., K. Schulten, and Z. Schulten (1980). First passage time approach to diffusion controlled reactions. *The Journal of Chemical Physics* 72, pp. 4350–4357.

Takemura, K. (2008). Integral representation of solutions to Fuchsian system and Heun's equation. *Journal of Mathematical Analysis and Applications* 342, pp. 52–69.

Tanaka, M. M., C. T. Bergstrom, and B. R. Levin (2003). The evolution of mutator genes in bacterial populations: The roles of environmental change and timing. *Genetics* 164, pp. 843–854.

Taniguchi, Y., P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329.5991, pp. 533–538.

Thattai, M. and A. van Oudenaarden (2001). Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences* 98.15, pp. 8614–8619.

Thattai, M. and A. van Oudenaarden (2004). Stochastic gene expression in fluctuating environments. *Genetics* 167, pp. 523–530.

Thomas, P., R. Grima, and A. V. Straube (2012). Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators. *Physical Review E* 86, 041110.

Tirosh, I. and N. Barkai (2008). Two strategies for gene regulation by promoter nucleosomes. *Genome Research* 18, pp. 1084–1091.

Usmani, R. (1994). Inversion of Jacobi's tridiagonal matrix. *Computers & Mathematics with Applications* 27.8, pp. 59–66.

Vandecan, Y. and R. Blossey (2013). Self-regulatory gene: An exact solution for the gene gate model. *Physical Review E* 87.4, 042705.

Visco, P., R. J. Allen, and M. R. Evans (2008). Exact solution of a model DNA-inversion genetic switch with orientational control. *Physical Review Letters* 101.11, 118104.

Volfson, D., J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty (2006). Origins of extrinsic variability in eukaryotic gene expression. *Nature* 439, pp. 861–864.

Voliotis, M., P. Thomas, R. Grima, and C. G. Bowsher (2015). Stochastic simulation of biomolecular networks in dynamic environments. arXiv: 1511.01268 [q-bio.QM].

Walczak, A. M., A. Mugler, and C. H. Wiggins (2009). A stochastic spectral analysis of transcriptional regulatory cascades. *Proceedings of the National Academy of Sciences* 106.16, pp. 6529–6534.

Warren, L., D. Bryder, I. Weissman, and S. Quake (2006). Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proceedings of the National Academy of Sciences* 103, pp. 17807–17812.

Weiss, N. A. (2005). *A Course in Probability*. Addison-Wesley.

Wikipedia (2014). *Cellular noise — Wikipedia, The Free Encyclopedia*. [Online; accessed 09-Nov-2015]. URL: https://en.wikipedia.org/wiki/Cellular_noise.

Wilkinson, D. J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics* 10, pp. 122–133.

Wonham, W. M. and A. T. Fuller (1958). Probability densities of the smoothed random telegraph signal. *Journal of Electronics and Control* 4, pp. 567–576.

Yunger, S., L. Rosenfeld, Y. Garini, and Y. Shav-Tal (2010). Single-allele analysis of transcription kinetics in living mammalian cells. *Nature methods* 7.8, pp. 631–633.

Zenklusen, D., D. R. Larson, and R. H. Singer (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural and Molecular Biology* 15.12, pp. 1263–1271.

Zenobi, R. (2013). Single-cell metabolomics: Analytical and biological perspectives. *Science* 342.6163.

Zhang, J., L. Chen, and T. Zhou (2012). Analytical distribution and tunability of noise in a model of promoter progress. *Biophysical Journal* 102, pp. 1247–1257.

Zoller, B., D. Nicolas, N. Molina, and F. Naef (2015). Structure of silent transcription intervals and noise characteristics of mammalian genes. *Molecular Systems Biology* 11.823.

Zon, R. van, S. Ciliberto, and E. G. D. Cohen (2004). Power and heat fluctuation theorems for electric circuits. *Physical Review Letters* 92, 130601 (13).

A. Summary of permission for third party copyright works

Page 13: Figure 2.2

Source: Molecular Cell, 58, K. Bahar Halpern *et al.*, Bursty gene expression in the intact mammalian liver, 147-156.

Copyright holder and year: Elsevier 2015

Permission to reproduce granted.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Nov 29, 2015

This is a License Agreement between Justine Dattani ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Justine Dattani
Customer address	Department of Mathematics
	London, SW7 2AZ
License number	3754140585846
License date	Nov 22, 2015
Licensed content publisher	Elsevier
Licensed content publication	Molecular Cell
Licensed content title	Bursty Gene Expression in the Intact Mammalian Liver
Licensed content author	Keren Bahar Halpern,Sivan Tanami,Shanie Landen,Michal Chapal,Liran Szlak,Anat Hutzler,Anna Nizhberg,Shalev Itzkovitz
Licensed content date	2 April 2015
Licensed content volume number	58
Licensed content issue number	1
Number of pages	10
Start Page	147
End Page	156
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	2
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	figures 1A, 1C, 2C
Title of your thesis/dissertation	Exact solutions of master equations for the analysis of gene transcription models
Expected completion date	Dec 2015
Estimated size (number of pages)	150
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 GBP
VAT/Local Sales Tax	0.00 GBP / 0.00 GBP

Total 0.00 GBP

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students

registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version.
Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
    - via their non-commercial person homepage or blog
    - by updating a preprint in arXiv or RePEc with the accepted manuscript
    - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
    - directly by providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
    - via non-commercial hosting platforms such as their institutional repository
    - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions
You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.
Terms & Conditions applicable to all Open Access articles published with Elsevier:
Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.
The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.
If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:
CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.
CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.
CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee. Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.8

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.