Preface

Stochastic models for chemical and biochemical reactions have been around for a long time. The standard algorithm for simulating the dynamics of such processes on a computer (the "Gillespie algorithm") was published nearly 30 years ago (and most of the relevant theory was sorted out long before that). In the meantime there have been dozens of papers published on stochastic kinetics, and several books on stochastic processes in physics, chemistry and biology. Biological modelling and biochemical kinetic modelling have been around even longer. These distinct subjects have started to merge in recent years as technology has begun to give real insight into intra-cellular processes. Improvements in experimental technology are enabling quantitative real-time imaging of expression at the single-cell level, and improvement in computing technology is allowing modelling and stochastic simulation of such systems at levels of detail previously impossible. The message that keeps being repeated is that the kinetics of biological processes at the intra-cellular level are stochastic, and that cellular function can not be properly understood without building that stochasticity into *in silico* models. It was this message that first interested me in systems biology and in the many challenging statistical problems that follow naturally from this observation.

It was only when I came to try and teach this interesting view of computational systems biology to graduate students that I realised there was no satisfactory text on which to base the course. The papers assumed far too much background knowledge, the standard biological texts didn't cover stochastic modelling, and the stochastic processes texts were too far removed from the practical applications of systems biology, paying little attention to stochastic biochemical kinetics. Where stochastic models do crop up in the mainstream systems biology literature, they tend to be treated as an add-on or after-thought, in a slightly superficial way. As a statistician I see this as problematic. The stochastic processes formalism provides a beautiful, elegant, and coherent foundation for chemical kinetics, and there is a wealth of associated theory every bit as powerful and elegant as that for conventional continuous deterministic models. Given the increasing importance of stochastic models in systems biology, I thought it would be particularly appropriate to write an introductory text in this area from this perspective.

This book assumes a basic familiarity with what might be termed high school mathematics. That is, a basic familiarity with algebra and calculus. It is also helpful to have had some exposure to linear algebra and matrix theory, but not vital. Since the teaching of probability and statistics at school is fairly patchy, essentially nothing will be assumed, though obviously a good background in this area will be very helpful. Starting from here, the book covers everything that is necessary for a good

appreciation of stochastic kinetic modelling of biological networks in the systems biology context. There is an emphasis on the necessary probabilistic and stochastic methods, but the theory is rooted in the intended application, and no time is wasted covering interesting theory that is not necessary for stochastic kinetic modelling. On the other hand, more-or-less everything that is necessary is covered, and the text (at least up to Chapter 8) is intended to be self-contained. The final chapters are necessarily a little more technical in nature, as they concern the difficult problem of inference for stochastic kinetic models from experimental data. This is still an active research area, and so the main aim here is to give pointers to the existing literature and provide enough background information to render that literature more accessible to the non-specialist.

The decision to make the book practically oriented necessitated some technological choices that will not suit everyone. The two key technologies chosen for illustrating the theory in this book are SBML and R. I hope that the choice of the Systems Biology Markup Language (SBML) for model representation is not too controversial. It is the closest thing to a standard that exists in the systems biology area, and there are dozens of software tools that support it. Of course, most people using SBML are using it to encode continuous deterministic models. However, SBML Level 2 and beyond are perfectly capable of encoding discrete stochastic models, and so one of the reasons for using it in this text is to provide some working examples of SBML models constructed with discrete stochastic simulation in mind.

The other technological choice was the use of the statistical programming language, R. This is likely to be more controversial, as there are plenty of other languages that could have been used. It seemed to me that in the context of a textbook, using a very high level language was most appropriate. In the context of stochastic modelling, a language with good built-in mathematical and statistical support also seemed highly desirable. R stood out as being the best choice in terms of built-in language support for stochastic simulation and statistical analysis. It also has the great advantage over some of the other possible choices of being completely free open-source software, and therefore available to anyone reading the book. In addition, R is being used increasingly in bioinformatics and other areas of systems biology, so hopefully for this reason too it will be regarded as a positive choice.

This book is intended for a variety of audiences (advanced undergraduates, graduate students, postdocs, and academics from a variety of backgrounds), and exactly how it is read will depend on the context. The book is certainly suitable for a variety of graduate programs in computational biology. I will be using it as a text for a second semester course on a masters in bioinformatics programme that will cover much of Chapters 1, 2, 4, 5, 6, and 7 (most of the material from Chapter 3 will be covered in a first semester course). However, the book has also been written with self-study in mind, and here it is intended that the entire book be read in sequence, with some chapters skipped depending on background knowledge. It is intended to be suitable for computational systems biologists from a continuous deterministic background who would like to know more about the stochastic approach, as well as for statisticians who are interested in learning more about systems biology (though statisticians will probably want to skip most of Chapters 3, 4, and 5). It is worth pointing out that

Chapters 9 and 10 will be easier to read with a reasonable background in probability and statistics. Though it should be possible to read these chapters without such a background and still appreciate the key concepts and ideas, some of the technical details may be difficult to understand fully from an elementary viewpoint.

Writing this book has been more effort than I anticipated, and there were one or two moments when I doubted the wisdom of taking the project on. On the whole however, it has been an interesting and rewarding experience for me, and I am pleased with the result. I know it is a book that *I* will find useful and will refer to often, as it integrates a fairly diverse literature into a single convenient and notationally consistent source. I can only hope that others share the same view, and that the book will help make a stochastic approach to computational systems biology more widely appreciated.

Darren Wilkinson
Newcastle upon Tyne, October 2005