
Preface

This *lange megillah* is concerned with establishing properties of a mathematical model in population genetics that some might regard, in light of what is being modeled, as entirely obvious. Once written down, however, a mathematical model has a life of its own; it must be addressed in its own terms, and understood without reference to its origins.

The models we consider are phrased as partial differential equations, which arise as limits of finite Markov chains. The existence of solutions to these partial differential equations and their properties are suggested by the physical, economic, or biological systems under consideration, but logically speaking, are entirely independent of them. What is far from obvious are the regularity properties of these solutions, and, as is so often the case, the existence of solutions actually hinges on these very subtle properties. Using a Schauder method, we *prove* the existence of solutions to a class of degenerate parabolic and elliptic equations that arise in population genetics and mathematical finance.

The archetypes for these equations arise as infinite population limits of the Wright-Fisher models in population genetics. These describe the prevalence of a mutant allele, in a population of fixed size, under the effects of genetic drift, mutation, migration and selection. The formal generator of the infinite population limit acts on functions defined on $[0, 1]$ (the space of frequencies) and is given by

$$L_{WF} = x(1-x)\partial_x^2 + [b_0(1-x) - b_1x + sx(1-x)]\partial_x. \quad (1)$$

Processes defined by such operators were studied by Feller in the early 1950s and used to great effect by Kimura et al. in the 1960s and '70s to give quantitative answers to a wide range of questions in population genetics. Notwithstanding, a modern appreciation of the analytic properties of these equations is only now coming into focus.

In this monograph we provide analytic foundations for equations of this type and their natural higher dimensional generalizations. We call these operators generalized Kimura diffusions. They act on functions defined on generalizations of convex polyhedra, which are called manifolds with corners. We provide the basic Hölder space-type estimates for operators in this class with which we establish the existence of solutions. These operators satisfy a strong form of the maximum principle, which implies uniqueness.

The partial differential operators we consider are degenerate and the underlying manifolds with corners are themselves singular. This inevitably produces significant technical challenges in the analysis of such equations, and explains, in part, the length

of this text. The Markov processes defined by these operators provide fundamental models for many Biological and Economic situations, and it is for this reason that we feel that these operators merit such a detailed and laborious treatment.

A large portion of this book is devoted to a careful exploration of operators of the form

$$L_{\mathbf{b},m} = \sum_{i=1}^n [x_i \partial_{x_i}^2 + b_i \partial_{x_i}] + \sum_{l=1}^m \partial_{y_l}^2, \quad (2)$$

acting on functions defined in $\mathbb{R}_+^n \times \mathbb{R}^m$. Here the coefficients $\{b_i\}$ are non-negative constants. These operators are interesting in their own right, arising as models in population genetics, but for us, they are largely building blocks for the analysis of general Kimura diffusions. These are the analogues, in the present context, of the “constant coefficient elliptic operators” in classical elliptic theory. A notable feature of this family is that, because the coefficient of $\partial_{x_i}^2$ vanishes at $x_i = 0$, we need to retain the first order transverse term. The value of $\mathbf{b} = (b_1, \dots, b_n)$ has a pervasive effect on the behavior of the solution. Much of our analysis rests upon explicit formulæ for the solutions of the initial value problems:

$$\partial_t v - L_{\mathbf{b},m} v = 0 \text{ with } v(\mathbf{x}, \mathbf{y}, 0) = f(\mathbf{x}, \mathbf{y}). \quad (3)$$

Using these models, we have succeeded in developing a rather complete existence and regularity theory for general Kimura diffusions with Hölder continuous data. This in turn suffices to prove the existence of a C_0 -semi-group acting on continuous functions, and shows that the associated martingale problem has a unique solution. The existence of a strong Markov process, which in applications to genetics describes the statistical behavior of individual populations, follows from this.

In special situations, such results have been established by other authors, but without either the precise control on the regularity of solutions, or the generality considered herein. As long as it is, this text just begins to scratch the surface of this very rich field. We have largely restricted our attention to the solutions with the best possible regularity properties, which leads to considerable simplifications. For applications it will be important to consider solutions with more complicated boundary behavior; we hope that this text will provide a solid foundation for these investigations.

ACKNOWLEDGMENTS

We would like to acknowledge the generous financial support and unflagging personal support provided by Ben Mann and the DARPA FunBio project. It is certainly the case that without Ben’s encouragement, we would never have undertaken this project. We would like to thank our FunBio colleagues¹ who provided us with the motivation

¹Phil Benfey, Michael Deem, Richard Lenski, Jack Morava, Lior Pachter, Herbert Edelsbrunner, John Harer, Jim Damon, Peter Bates, Joshua Weitz, Konstantin Mischaikow, Gunnar Carlsson, Bernd Sturmfels, Tim Buchman, Ary Goldberger, Jonathan Eisen, Olivier Porquie, Thomas Fink, Ned Wingreen, Jonathan Dushoff, Peter Nara, *inter alios*, and research staff: Mark Filipkowski, Shauna Koppel, Rachel Scholz, Matthew Clement, Traci Kiesling, Traci Pals, *inter alios*.

and knowledge base to pursue this project, and Simon Levin for his leadership and inspiration. CLE would like to thank Josh Plotkin, Warren Ewens, Todd Parsons, and Ricky Der, from whom he has learned what little he knows about population genetics.

We would both like to thank Charlie Fefferman for showing us an explicit formula for $k_t^0(x, y)$, which set us off in the very fruitful direction pursued herein. We would also like to thank Dan Stroock for his help with connections to Probability Theory. We would like to thank Camelia Pop for providing a long list of corrections.

CLE would like to acknowledge the financial support of DARPA through grants HR0011-05-1-0057 and HR0011-09-1-0055, the support of NSF through grant DMS-0603973, and the hospitality and support provided by Leslie Greengard and the Courant Institute, where this text was completed. He would also like to thank his wife Jane for doing the hard work to get us there and back.

RM would like to acknowledge the financial support of DARPA through grants HR0011-05-1-0057 and HR0011-09-1-0055, and the support of NSF through grants DMS-0805529 and DMS-1105050. He is also immensely grateful to his wife Laurence for her general forbearance and support, and to his daughters Sabine and Sophie for making it all worthwhile.