

# Lecture Notes in Mathematics

2012

**Editors:**

J.-M. Morel, Cachan

B. Teissier, Paris

**Subseries:**

École d'Été de Probabilités de Saint-Flour

## Saint-Flour Probability Summer School



The Saint-Flour volumes are reflections of the courses given at the Saint-Flour Probability Summer School. Founded in 1971, this school is organised every year by the Laboratoire de Mathématiques (CNRS and Université Blaise Pascal, Clermont-Ferrand, France). It is intended for PhD students, teachers and researchers who are interested in probability theory, statistics, and in their applications.

The duration of each school is 13 days (it was 17 days up to 2005), and up to 70 participants can attend it. The aim is to provide, in three high-level courses, a comprehensive study of some fields in probability theory or Statistics. The lecturers are chosen by an international scientific board. The participants themselves also have the opportunity to give short lectures about their research work.

Participants are lodged and work in the same building, a former seminary built in the 18th century in the city of Saint-Flour, at an altitude of 900 m. The pleasant surroundings facilitate scientific discussion and exchange.

The Saint-Flour Probability Summer School is supported by:

- Université Blaise Pascal
- Centre National de la Recherche Scientifique (C.N.R.S.)
- Ministère délégué à l'Enseignement supérieur et à la Recherche

For more information, see back pages of the book and  
<http://math.univ-bpclermont.fr/stflour/>

Jean Picard  
Summer School Chairman  
Laboratoire de Mathématiques  
Université Blaise Pascal  
63177 Aubière Cedex  
France

Alison Etheridge

# Some Mathematical Models from Population Genetics

École d'Été de Probabilités  
de Saint-Flour XXXIX-2009

 Springer

Alison Etheridge  
University of Oxford  
Department of Statistics  
1 South Parks Road  
Oxford OX1 3TG  
United Kingdom  
[etheridg@stats.ox.ac.uk](mailto:etheridg@stats.ox.ac.uk)

ISBN: 978-3-642-16631-0 e-ISBN: 978-3-642-16632-7  
DOI: 10.1007/978-3-642-16632-7  
Springer Heidelberg Dordrecht London New York

Lecture Notes in Mathematics ISSN print edition: 0075-8434  
ISSN electronic edition: 1617-9692

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Cover design:* SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

These notes on mathematical models from population genetics reflect the 16 h of lectures that I delivered in St Flour in July 2009. Other than minor corrections and clarifications, they have changed very little in the year that has elapsed since then. Although it was tempting to add more material, I concluded that not only would this lead to unacceptable delays, but it would also be redundant. Whereas there are few references of which I am aware that present the material covered here in a self-contained way, there are now many texts that cover, for example, coalescent theory in more detail.

The notes are intended for graduate students in mathematics. They aim to introduce the reader to a range of mathematical models that have their origins in theoretical population genetics. Some date right back to the origins of the subject and some were introduced in the last few years. All share a rich mathematical structure. Research on the more recent models, notably the  $\Lambda$ -coalescents and their spatial analogues, is progressing at a breathtaking speed and so it is impossible to provide a comprehensive survey of what is known. Instead I have aimed to explain some of the reasons that such models are interesting biologically and to equip the reader with enough background to be able to browse the literature as it appears.

There are many people to whom I owe thanks. My interest in population genetics stems from a collaboration with Nick Barton (IST Austria and the University of Edinburgh). Working with Nick over many years has been a privilege and a pleasure and almost all the material covered here I first learned about through conversations with him. I am grateful to Leif Döring, Bjarki Eldon, Bob Griffiths, Habib Saadi and the many others who read and commented on parts of the manuscript. Special thanks are due to Amandine Véber, who went through several iterations of the whole document in tremendous detail and undoubtedly improved the notes beyond recognition. I was fortunate to spend the first four months of 2009 visiting Université Paris Sud in Orsay. My thanks go to everyone there, especially Yves Le Jan for making that possible. While I was in Orsay, Jean-François Le Gall persuaded me to give a masters course as a dry run for (at least part of) this course. The experience was extremely valuable and my thanks go to Jean-François and to the enthusiastic audience. Jean Picard quietly ensured that everything at St Flour ran extremely smoothly and the

participants were tremendous. I simply had a lot of fun. Finally, as always, I thank Lionel, Charlotte and Matthew Mason for all their support and understanding.

Oxford, August 2010

*Alison Etheridge*

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Mutation and Random Genetic Drift</b>	<b>5</b>
2.1	The Wright–Fisher Model and the Kingman Coalescent	5
2.2	The Cannings Model	11
2.3	Selfing	13
2.4	Adding Mutations	14
2.5	Inferring Genealogies From Data	17
2.6	Some Properties of Kingman’s Coalescent	18
2.7	Genealogies and Pedigrees	21
2.8	The Moran Model	23
2.9	The Site Frequency Spectrum	25
2.10	The Lookdown Process	27
2.11	A More Simplistic Limit	30
<b>3</b>	<b>One Dimensional Diffusions</b>	<b>33</b>
3.1	Diffusions	33
3.2	Convergence to Diffusions	36
3.3	Speed and Scale	38
3.4	Hitting Probabilities and Feller’s Boundary Classification	40
3.5	Green’s Functions	42
3.6	Stationary Distributions and Reversibility	46
<b>4</b>	<b>More than Two Types</b>	<b>53</b>
4.1	Multi-Dimensional Diffusion Models	53
4.2	The Poisson–Dirichlet and GEM Distributions	57
4.3	Ewens Sampling Formula	61
<b>5</b>	<b>Selection</b>	<b>65</b>
5.1	Genetic Diversity	65
5.2	Wright–Fisher Model with Selection	65

5.3	Selection in a Diploid Population .....	69
5.4	The Ancestral Selection Graph.....	70
5.5	Adding Structure to the Coalescent .....	74
5.6	Selective Sweeps.....	77
5.7	Coalescents with Multiple Mergers .....	82
<b>6</b>	<b>Spatial Structure .....</b>	<b>89</b>
6.1	Subdivided Populations and the Structured Coalescent.....	89
6.2	Duality .....	91
6.3	Collapse of Structure .....	95
6.4	Evolution in a Spatial Continuum and the Pain in the Torus .....	99
6.5	The Spatial $\Lambda$ -Fleming–Viot Process .....	102
6.6	More General Models .....	106
	<b>References .....</b>	<b>109</b>
	<b>List of Participants .....</b>	<b>113</b>
	<b>Index .....</b>	<b>117</b>