

# Lecture Notes in Mathematics

1953

## **Editors:**

J.-M. Morel, Cachan

F. Takens, Groningen

B. Teissier, Paris

Khadiga A. Arwini · Christopher T.J. Dodson

# Information Geometry

Near Randomness  
and Near Independence

 Springer

## Authors

Khadiga A. Arwini  
Al-Fateh University  
Faculty of Sciences  
Mathematics Department  
Box 13496  
Tripoli, Libya  
arwini2001@yahoo.com

Christopher T.J. Dodson  
University of Manchester  
School of Mathematics  
Manchester M13 9PL, United Kingdom  
ctdodson@manchester.ac.uk

ISBN: 978-3-540-69391-8 e-ISBN: 978-3-540-69393-2  
DOI: 10.1007/978-3-540-69393-2

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

Library of Congress Control Number: 2008930087

Mathematics Subject Classification (2000): 53B50, 60D05, 62B10, 62P35, 74E35, 92D20

© 2008 Springer-Verlag Berlin Heidelberg

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 Publishing Services

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

---

## Preface

The main motivation for this book lies in the breadth of applications in which a statistical model is used to represent small departures from, for example, a Poisson process. Our approach uses information geometry to provide a common context but we need only rather elementary material from differential geometry, information theory and mathematical statistics. Introductory sections serve together to help those interested from the applications side in making use of our methods and results. We have available *Mathematica* notebooks to perform many of the computations for those who wish to pursue their own calculations or developments.

Some 44 years ago, the second author first encountered, at about the same time, differential geometry via relativity from Weyl's book [209] during undergraduate studies and information theory from Tribus [200, 201] via spatial statistical processes while working on research projects at Wiggins Teape Research and Development Ltd—cf. the Foreword in [196] and [170, 47, 58]. Having started work there as a student laboratory assistant in 1959, this research environment engendered a recognition of the importance of international collaboration, and a lifelong research interest in randomness and near-Poisson statistical geometric processes, persisting at various rates through a career mainly involved with global differential geometry. From correspondence in the 1960s with Gabriel Kron [4, 124, 125] on his Diakoptics, and with Kazuo Kondo who influenced the post-war Japanese schools of differential geometry and supervised Shun-ichi Amari's doctorate [6], it was clear that both had a much wider remit than traditionally pursued elsewhere. Indeed, on moving to Lancaster University in 1969, receipt of the latest *RAAG Memoirs Volume 4 1968* [121] provided one of Amari's early articles on information geometry [7], which subsequently led to his greatly influential 1985 Lecture Note volume [8] and our 1987 *Geometrization of Statistical Theory Workshop* at Lancaster University [10, 59].

Reported in this monograph is a body of results, and computer-algebraic methods that seem to have quite general applicability to statistical models admitting representation through parametric families of probability density

functions. Some illustrations are given from a variety of contexts for geometric characterization of statistical states near to the three important standard basic reference states: (Poisson) randomness, uniformity, independence. The individual applications are somewhat heuristic models from various fields and we incline more to terminology and notation from the applications rather than from formal statistics. However, a common thread is a geometrical representation for statistical perturbations of the basic standard states, and hence results gain qualitative stability. Moreover, the geometry is controlled by a metric structure that owes its heritage through maximum likelihood to information theory so the quantitative features—lengths of curves, geodesics, scalar curvatures etc.—have some respectable authority. We see in the applications simple models for galactic void distributions and galaxy clustering, amino acid clustering along protein chains, cryptographic protection, stochastic fibre networks, coupled geometric features in hydrology and quantum chaotic behaviour. An ambition since the publication by Richard Dawkins of *The Selfish Gene* [51] has been to provide a suitable differential geometric framework for dynamics of natural evolutionary processes, but it remains elusive. On the other hand, in application to the statistics of amino acid spacing sequences along protein chains, we describe in Chapter 7 a stable statistical qualitative property that may have evolutionary significance. Namely, to widely varying extents, all twenty amino acids exhibit greater clustering than expected from Poisson processes. Chapter 11 considers eigenvalue spacings of infinite random matrices and near-Poisson quantum chaotic processes.

The second author has benefited from collaboration (cf. [34]) with the group headed by Andrew Doig of the Manchester Interdisciplinary Biocentre, the University of Manchester, and has had long-standing collaborations with groups headed by Bill Sampson of the School of Materials, the University of Manchester (cf.eg. [73]) and Jacob Scharcanski of the Instituto de Informatica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil (cf.eg. [76]) on stochastic modelling. We are pleased therefore to have co-authored with these colleagues three chapters: titled respectively, Amino Acid Clustering, Stochastic Fibre Networks, Stochastic Porous Media and Hydrology.

The original draft of the present monograph was prepared as notes for short Workshops given by the second author at Centro de Investigaciones de Matematica (CIMAT), Guanajuato, Mexico in May 2004 and also in the Departamento de Xeometra e Topoloxa, Facultade de Matemáticas, Universidade de Santiago de Compostela, Spain in February 2005.

The authors have benefited at different times from discussions with many people but we mention in particular Shun-ichi Amari, Peter Jupp, Patrick Laycock, Hiroshi Matsuzoe, T. Subba Rao and anonymous referees. However, any overstatements in this monograph will indicate that good advice may have been missed or ignored, but actual errors are due to the authors alone.

*Khadiga Arwini, Department of Mathematics, Al-Fateh University, Libya*  
*Kit Dodson, School of Mathematics, the University of Manchester, England*

---

## Contents

<b>Preface</b> .....	V
<b>1 Mathematical Statistics and Information Theory</b> .....	1
1.1 Probability Functions for Discrete Variables .....	2
1.1.1 Bernoulli Distribution .....	3
1.1.2 Binomial Distribution .....	3
1.1.3 Poisson Distribution .....	4
1.2 Probability Density Functions for Continuous Variables .....	6
1.2.1 Uniform Distribution .....	8
1.2.2 Exponential Distribution .....	8
1.2.3 Gaussian, or Normal Distribution .....	9
1.3 Joint Probability Density Functions .....	9
1.3.1 Bivariate Gaussian Distributions .....	10
1.4 Information Theory .....	11
1.4.1 Gamma Distribution .....	13
<b>2 Introduction to Riemannian Geometry</b> .....	19
2.0.2 Manifolds .....	20
2.0.3 Tangent Spaces .....	20
2.0.4 Tensors and Forms .....	22
2.0.5 Riemannian Metric .....	25
2.0.6 Connections .....	26
2.1 Autoparallel and Geodesic Curves .....	29
2.2 Universal Connections and Curvature .....	29
<b>3 Information Geometry</b> .....	31
3.1 Fisher Information Metric .....	32
3.2 Exponential Family of Probability Density Functions .....	33
3.3 Statistical $\alpha$ -Connections .....	34
3.4 Affine Immersions .....	35
3.4.1 Weibull Distributions: Not of Exponential Type .....	36

## VIII Contents

3.5	Gamma 2-Manifold $\mathcal{G}$ . . . . .	37
3.5.1	Gamma $\alpha$ -Connection . . . . .	38
3.5.2	Gamma $\alpha$ -Curvatures . . . . .	39
3.5.3	Gamma Manifold Geodesics . . . . .	40
3.5.4	Mutually Dual Foliations . . . . .	42
3.5.5	Gamma Affine Immersion . . . . .	42
3.6	Log-Gamma 2-Manifold $\mathcal{L}$ . . . . .	42
3.6.1	Log-Gamma Random Walks . . . . .	45
3.7	Gaussian 2-Manifold . . . . .	45
3.7.1	Gaussian Natural Coordinates . . . . .	47
3.7.2	Gaussian Information Metric . . . . .	47
3.7.3	Gaussian Mutually Dual Foliations . . . . .	48
3.7.4	Gaussian Affine Immersions . . . . .	48
3.8	Gaussian $\alpha$ -Geometry . . . . .	49
3.8.1	Gaussian $\alpha$ -Connection . . . . .	49
3.8.2	Gaussian $\alpha$ -Curvatures . . . . .	50
3.9	Gaussian Mutually Dual Foliations . . . . .	50
3.10	Gaussian Submanifolds . . . . .	51
3.10.1	Central Mean Submanifold . . . . .	51
3.10.2	Unit Variance Submanifold . . . . .	52
3.10.3	Unit Coefficient of Variation Submanifold . . . . .	52
3.11	Gaussian Affine Immersions . . . . .	52
3.12	Log-Gaussian Manifold . . . . .	53
<b>4</b>	<b>Information Geometry of Bivariate Families . . . . .</b>	<b>55</b>
4.1	McKay Bivariate Gamma 3-Manifold $M$ . . . . .	55
4.2	McKay Manifold Geometry in Natural Coordinates . . . . .	58
4.3	McKay Densities Have Exponential Type . . . . .	59
4.3.1	McKay Information Metric . . . . .	59
4.4	McKay $\alpha$ -Geometry . . . . .	60
4.4.1	McKay $\alpha$ -Connection . . . . .	60
4.4.2	McKay $\alpha$ -Curvatures . . . . .	61
4.5	McKay Mutually Dual Foliations . . . . .	64
4.6	McKay Submanifolds . . . . .	65
4.6.1	Submanifold $M_1$ . . . . .	65
4.6.2	Submanifold $M_2$ . . . . .	68
4.6.3	Submanifold $M_3$ . . . . .	69
4.7	McKay Bivariate Log-Gamma Manifold $\tilde{M}$ . . . . .	71
4.8	Generalized McKay 5-Manifold . . . . .	72
4.8.1	Bivariate 3-Parameter Gamma Densities . . . . .	72
4.8.2	Generalized McKay Information Metric . . . . .	73
4.9	Freund Bivariate Exponential 4-Manifold $\mathcal{F}$ . . . . .	74
4.9.1	Freund Fisher Metric . . . . .	75
4.10	Freund Natural Coordinates . . . . .	76

4.11	Freund $a$ -Geometry	77
4.11.1	Freund $a$ -Connection	77
4.11.2	Freund $a$ -Curvatures	78
4.12	Freund Foliations	80
4.13	Freund Submanifolds	81
4.13.1	Independence Submanifold $F_1$	81
4.13.2	Submanifold $F_2$	82
4.13.3	Submanifold $F_3$	83
4.13.4	Submanifold $F_4$	84
4.14	Freund Affine Immersion	87
4.15	Freund Bivariate Log-Exponential Manifold	87
4.16	Bivariate Gaussian 5-Manifold $\mathcal{N}$	88
4.17	Bivariate Gaussian Fisher Information Metric	89
4.18	Bivariate Gaussian Natural Coordinates	90
4.19	Bivariate Gaussian $a$ -Geometry	91
4.19.1	$a$ -Connection	91
4.19.2	$a$ -Curvatures	94
4.20	Bivariate Gaussian Foliations	98
4.21	Bivariate Gaussian Submanifolds	99
4.21.1	Independence Submanifold $\mathcal{N}_1$	99
4.21.2	Identical Marginals Submanifold $\mathcal{N}_2$	101
4.21.3	Central Mean Submanifold $\mathcal{N}_3$	103
4.21.4	Affine Immersion	105
4.22	Bivariate Log-Gaussian Manifold	106
<b>5</b>	<b>Neighbourhoods of Poisson Randomness, Independence, and Uniformity</b>	109
5.1	Gamma Manifold $\mathcal{G}$ and Neighbourhoods of Randomness	110
5.2	Log-Gamma Manifold $\mathcal{L}$ and Neighbourhoods of Uniformity	111
5.3	Freund Manifold $\mathcal{F}$ and Neighbourhoods of Independence	112
5.3.1	Freund Submanifold $F_2$	113
5.4	Neighbourhoods of Independence for Gaussians	114
<b>6</b>	<b>Cosmological Voids and Galactic Clustering</b>	119
6.1	Spatial Stochastic Processes	120
6.2	Galactic Cluster Spatial Processes	121
6.3	Cosmological Voids	125
6.4	Modelling Statistics of Cosmological Void Sizes	126
6.5	Coupling Galaxy Clustering and Void Sizes	130
6.6	Representation of Cosmic Evolution	132
<b>7</b>	<b>Amino Acid Clustering</b>	
	<b>With A.J. Doig</b>	139
7.1	Spacings of Amino Acids	139
7.2	Poisson Spaced Sequences	141



7.3	Non-Poisson Sequences as Gamma Processes . . . . .	142
7.3.1	Local Geodesic Distance Approximations . . . . .	145
7.4	Results . . . . .	148
7.5	Why Would Amino Acids Cluster? . . . . .	151
<b>8</b>	<b>Cryptographic Attacks and Signal Clustering . . . . .</b>	<b>153</b>
8.1	Cryptographic Attacks . . . . .	153
8.2	Information Geometry of the Log-gamma Manifold . . . . .	154
8.3	Distinguishing Nearby Unimodular Distributions . . . . .	155
8.4	Difference From a Uniform Distribution . . . . .	157
8.5	Gamma Distribution Neighbourhoods of Randomness . . . . .	157
<b>9</b>	<b>Stochastic Fibre Networks</b>	
	<b>With W.W. Sampson . . . . .</b>	<b>161</b>
9.1	Random Fibre Networks . . . . .	161
9.2	Random Networks of Rectangular Fibres . . . . .	164
9.3	Log-Gamma Information Geometry for Fibre Clustering . . . . .	168
9.4	Bivariate Gamma Distributions for Anisotropy . . . . .	169
9.5	Independent Polygon Sides . . . . .	171
9.5.1	Multiplanar Networks . . . . .	175
9.6	Correlated Polygon Sides . . . . .	179
9.6.1	McKay Bivariate Gamma Distribution . . . . .	182
9.6.2	McKay Information Geometry . . . . .	184
9.6.3	McKay Information Entropy . . . . .	188
9.6.4	Simulation Results . . . . .	191
<b>10</b>	<b>Stochastic Porous Media and Hydrology</b>	
	<b>With J. Scharcanski and S. Felipussi . . . . .</b>	<b>195</b>
10.1	Hydrological Modelling . . . . .	195
10.2	Univariate Gamma Distributions and Randomness . . . . .	196
10.3	McKay Bivariate Gamma 3-Manifold . . . . .	196
10.4	Distance Approximations in the McKay Manifold . . . . .	198
10.5	Modelling Stochastic Porous Media . . . . .	200
10.5.1	Adaptive Tomographic Image Segmentation . . . . .	201
10.5.2	Mathematical Morphology Concepts . . . . .	203
10.5.3	Adaptive Image Segmentation and Representation . . . . .	209
10.5.4	Soil Tomographic Data . . . . .	214
<b>11</b>	<b>Quantum Chaology . . . . .</b>	<b>223</b>
11.1	Introduction . . . . .	223
11.2	Eigenvalues of Random Matrices . . . . .	226
11.3	Deviations . . . . .	229
	<b>References . . . . .</b>	<b>235</b>
	<b>Index . . . . .</b>	<b>247</b>