

Ergebnisse der Mathematik und ihrer Grenzgebiete

Volume 64

3. Folge

A Series of Modern Surveys in Mathematics

Editorial Board

L. Ambrosio, Pisa V. Baladi, Paris
G.-M. Greuel, Kaiserslautern
M. Gromov, Bures-sur-Yvette
G. Huisken, Tübingen J. Jost, Leipzig
J. Kollár, Princeton G. Laumon, Orsay
U. Tillmann, Oxford J. Tits, Paris
D.B. Zagier, Bonn

For further volumes:

www.springer.com/series/728

Nihat Ay • Jürgen Jost • Hồng Vân Lê •
Lorenz Schwachhöfer

Information Geometry

Nihat Ay
Information Theory of Cognitive Systems
MPI for Mathematics in the Sciences
Leipzig, Germany
and

Santa Fe Institute, Santa Fe, NM, USA

Jürgen Jost
Geometric Methods and Complex Systems
MPI for Mathematics in the Sciences
Leipzig, Germany
and

Santa Fe Institute, Santa Fe, NM, USA

Hông Vân Lê
Mathematical Institute of ASCR
Czech Academy of Sciences
Praha 1, Czech Republic

Lorenz Schwachhöfer
Department of Mathematics
TU Dortmund University
Dortmund, Germany

ISSN 0071-1136

Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys
in Mathematics

ISBN 978-3-319-56477-7

DOI 10.1007/978-3-319-56478-4

ISSN 2197-5655 (electronic)

ISBN 978-3-319-56478-4 (eBook)

Library of Congress Control Number: 2017951855

Mathematics Subject Classification: 60A10, 62B05, 62B10, 62G05, 53B21, 53B05, 46B20, 94A15,
94A17, 94B27

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, 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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Information geometry is the differential geometric treatment of statistical models. It thereby provides the mathematical foundation of statistics. Information geometry therefore is of interest both for its beautiful mathematical structure and for the insight it provides into statistics and its applications. Information geometry currently is a very active field. For instance, Springer will soon launch a new topical journal “Information Geometry”. We therefore think that the time is appropriate for a monograph on information geometry that develops the underlying mathematical theory in full generality and rigor, that explores the connections to other mathematical disciplines, and that proves abstract and general versions of the classical results of statistics, like the Cramér–Rao inequality or Chentsov’s theorem. These, then, are the purposes of the present book, and we hope that it will become the standard reference for the field.

Parametric statistics as introduced by R. Fisher considers parametrized families of probability measures on some finite or infinite sample space Ω . Typically, one wishes to identify a parameter so that the resulting probability measure best fits the observation among the measures in the family. This naturally leads to quantitative questions, in particular, how sensitively the measures in the family depend on the parameter. For this, a geometric perspective is expedient. There is a natural metric, the Fisher metric introduced by Rao, on the space of probability measures on Ω . This metric is simply the projective or spherical metric obtained when one considers a probability measure as a non-negative measure with a scaling factor to render its total mass equal to unity. The Fisher metric thus is a Riemannian metric that induces a corresponding structure on parametrized families of probability measures as above. Furthermore, moving from one reference measure to another yields an affine structure as discovered by S.I. Amari and N.N. Chentsov. The investigation of these metric and affine structures is therefore called information geometry. Information-theoretical quantities like relative entropies (Kullback–Leibler divergences) then find a natural geometric interpretation.

Information geometry thus provides a way of understanding information-theoretic quantities, statistical models, and corresponding statistical inference methods in geometric terms. In particular, the Fisher metric and the Amari–Chentsov

structure are characterized by their invariance under sufficient statistics. Several geometric formalisms have been identified as powerful tools to this end and emphasize respective geometric aspects of probability theory. In this book, we move beyond the applications in statistics and develop both a functional analytic and a geometric theory that are of mathematical interest in their own right. In particular, the theory of dually affine structures turns out to be an analogue of Kähler geometry in a real as opposed to a complex setting.

Also, as the concept of Shannon information can be related to the entropy concepts of Boltzmann and Gibbs, there is also a natural connection between information geometry and statistical mechanics. Finally, information geometry can also be used as a foundation of important parts of mathematical biology, like the theory of replicator equations and mathematical population genetics.

Sample spaces could be finite, but more often than not, they are infinite, for instance, subsets of some (finite- or even infinite-dimensional) Euclidean space. The spaces of measures on such spaces therefore are infinite-dimensional Banach spaces. Consequently, the differential geometric approach needs to be supplemented by functional analytic considerations. One of the purposes of this book therefore is to provide a general framework that integrates the differential geometry into the functional analysis.

Acknowledgements

We would like to thank Shun-ichi Amari for many fruitful discussions. This work was mainly carried out at the Max Planck Institute for Mathematics in the Sciences in Leipzig. It has also been supported by the BSI at RIKEN in Tokyo, the ASSMS, GCU in Lahore-Pakistan, the VNU for Sciences in Hanoi, the Mathematical Institute of the Academy of Sciences of the Czech Republic in Prague, and the Santa Fe Institute. We are grateful for the excellent working conditions and financial support of these institutions during extended visits of some of us. In particular, we should like to thank Antje Vandenberg for her outstanding logistic support.

The research of J.J. leading to his book contribution has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 267087. The research of H.V.L. is supported by RVO: 67985840.

Contents

1	Introduction	1
1.1	A Brief Synopsis	1
1.2	An Informal Description	6
1.2.1	The Fisher Metric and the Amari–Chentsov Structure for Finite Sample Spaces	7
1.2.2	Infinite Sample Spaces and Functional Analysis	8
1.2.3	Parametric Statistics	10
1.2.4	Exponential and Mixture Families from the Perspective of Differential Geometry	14
1.2.5	Information Geometry and Information Theory	15
1.3	Historical Remarks	17
1.4	Organization of this Book	20
2	Finite Information Geometry	25
2.1	Manifolds of Finite Measures	25
2.2	The Fisher Metric	29
2.3	Gradient Fields	39
2.4	The m - and e -Connections	42
2.5	The Amari–Chentsov Tensor and the α -Connections	47
2.5.1	The Amari–Chentsov Tensor	47
2.5.2	The α -Connections	50
2.6	Congruent Families of Tensors	52
2.7	Divergences	68
2.7.1	Gradient-Based Approach	68
2.7.2	The Relative Entropy	70
2.7.3	The α -Divergence	73
2.7.4	The f -Divergence	76
2.7.5	The q -Generalization of the Relative Entropy	78

2.8	Exponential Families	79
2.8.1	Exponential Families as Affine Spaces	79
2.8.2	Implicit Description of Exponential Families	84
2.8.3	Information Projections	91
2.9	Hierarchical and Graphical Models	100
2.9.1	Interaction Spaces	101
2.9.2	Hierarchical Models	108
2.9.3	Graphical Models	112
3	Parametrized Measure Models	121
3.1	The Space of Probability Measures and the Fisher Metric	121
3.2	Parametrized Measure Models	135
3.2.1	The Structure of the Space of Measures	139
3.2.2	Tangent Fibration of Subsets of Banach Manifolds	140
3.2.3	Powers of Measures	143
3.2.4	Parametrized Measure Models and k -Integrability	150
3.2.5	Canonical n -Tensors of an n -Integrable Model	164
3.2.6	Signed Parametrized Measure Models	168
3.3	The Pistone–Sempi Structure	170
3.3.1	e -Convergence	170
3.3.2	Orlicz Spaces	172
3.3.3	Exponential Tangent Spaces	176
4	The Intrinsic Geometry of Statistical Models	185
4.1	Extrinsic Versus Intrinsic Geometric Structures	185
4.2	Connections and the Amari–Chentsov Structure	189
4.3	The Duality Between Exponential and Mixture Families	201
4.4	Canonical Divergences	210
4.4.1	Dual Structures via Divergences	210
4.4.2	A General Canonical Divergence	213
4.4.3	Recovering the Canonical Divergence of a Dually Flat Structure	215
4.4.4	Consistency with the Underlying Dualistic Structure	217
4.5	Statistical Manifolds and Statistical Models	219
4.5.1	Statistical Manifolds and Isostatistical Immersions	220
4.5.2	Monotone Invariants of Statistical Manifolds	223
4.5.3	Immersion of Compact Statistical Manifolds into Linear Statistical Manifolds	226
4.5.4	Proof of the Existence of Isostatistical Immersions	228
4.5.5	Existence of Statistical Embeddings	238

5	Information Geometry and Statistics	241
5.1	Congruent Embeddings and Sufficient Statistics	241
5.1.1	Statistics and Congruent Embeddings	244
5.1.2	Markov Kernels and Congruent Markov Embeddings	253
5.1.3	Fisher–Neyman Sufficient Statistics	261
5.1.4	Information Loss and Monotonicity	263
5.1.5	Chentsov’s Theorem and Its Generalization	268
5.2	Estimators and the Cramér–Rao Inequality	277
5.2.1	Estimators and Their Bias, Mean Square Error, Variance	277
5.2.2	A General Cramér–Rao Inequality	281
5.2.3	Classical Cramér–Rao Inequalities	286
5.2.4	Efficient Estimators and Consistent Estimators	287
6	Fields of Application of Information Geometry	295
6.1	Complexity of Composite Systems	295
6.1.1	A Geometric Approach to Complexity	296
6.1.2	The Information Distance from Hierarchical Models	298
6.1.3	The Weighted Information Distance	307
6.1.4	Complexity of Stochastic Processes	317
6.2	Evolutionary Dynamics	327
6.2.1	Natural Selection and Replicator Equations	328
6.2.2	Continuous Time Limits	333
6.2.3	Population Genetics	336
6.3	Monte Carlo Methods	348
6.3.1	Langevin Monte Carlo	350
6.3.2	Hamiltonian Monte Carlo	351
6.4	Infinite-Dimensional Gibbs Families	354
	Appendix A Measure Theory	361
	Appendix B Riemannian Geometry	367
	Appendix C Banach Manifolds	381
	References	387
	Index	397
	Nomenclature	403