Preface

Information geometry provides the mathematical sciences with a new framework for analysis. This framework is relevant to a wide variety of domains, and it has already been usefully applied to several of these, providing them with a new perspective from which to view the structure of the systems which they investigate. Nevertheless, the development of the field of information geometry can only be said to have just begun.

Information geometry began as an investigation of the natural differential geometric structure possessed by families of probability distributions. As a rather simple example, consider the set $S$ of normal distributions with mean $\mu$ and variance $\sigma^2$:

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}.$$  

By specifying $(\mu, \sigma)$ we determine a particular normal distribution, and hence $S$ may be viewed as a 2-dimensional space (manifold) which has $(\mu, \sigma)$ as a coordinate system. However, this is not a Euclidean space, but rather a Riemannian space with a metric which naturally follows from the underlying properties of probability distributions. In particular, when $S$ is a family of normal distributions, it is a space of constant negative curvature. The underlying characteristics of probability distributions lead not only to this Riemannian structure; an investigation of the structure of probability distributions leads to a new concept within differential geometry: that of mutually dual affine connections. In addition, the structure of dual affine connections naturally arises in the framework of affine differential geometry, and has begun to attract the attention of mathematicians researching differential geometry.

Probability distributions are the fundamental element over which fields such as statistics, stochastic processes, and information theory are developed. Hence not only is the natural dualistic differential geometric structure of the space of probability distributions beautiful, but it must also play a fundamental role in these information sciences. In fact, considering statistical estimation from a differential geometric viewpoint has provided statistics with a new analytic tool which has allowed several previously open problems to be solved; information geometry has already established itself within the field of statistics. In the fields of information theory, stochastic processes, and systems, information geometry
is being currently applied to allow the investigation of hitherto unexplored possibilities.

The utility of information geometry, however, is not limited to these fields. It has, for example, been productively applied to areas such as statistical physics and the mathematical theory underlying neural networks. Further, dualistic differential geometric structure is a general concept not inherently tied to probability distributions. For example, the interior method for linear programming may be analyzed from this point of view, and this suggests its relation to completely integrable dynamical systems. Finally, the investigation of the information geometry of quantum systems may lead to even further developments.

This book presents for the first time the entirety of the emerging field of information geometry. To do this requires an understanding of at least the fundamental concepts in differential geometry. Hence the first three chapters contain an introduction to differential geometry and the recently developed theory of dual connections. An attempt has been made to develop the fundamental framework of differential geometry as concisely and intuitively as possible. It is hoped that this book may serve generally as an introduction to differential geometry. Although differential geometry is said to be a difficult field to understand, this is true only of those texts written by mathematicians for other mathematicians, and it is not the case that the principal ideas in differential geometry are hard. Nevertheless, this book introduces only the amount of differential geometry necessary for the remaining chapters, and endeavors to do so in a manner which, while consistent with the conventional definitions in mathematical texts, allows the intuition underlying the concepts to be comprehended most immediately.

On the other hand, a comprehensive treatment of statistics, system theory, and information theory, among others, from the point of view of information geometry is for each distinct, relying on properties unique to that particular theory. It was beyond the scope of this book to include a thorough description of these fields, and inevitably, many of the relevant topics from these areas are rather hastily introduced in the latter half of the book. It is hoped that within these sections the reader will simply gather the flavor of the research being done, and for a more complete analysis refer to the corresponding papers. To complement this approach, many topics which are still incomplete and perhaps consist only of vague ideas have been included.

Nothing would make us happier than if this book could serve as an invitation for other researches to join in the development of information geometry.
Information geometry provides a new method applicable to various areas including information sciences and physical sciences. It has emerged from investigating the geometrical structures of the manifold of probability distributions, and has applied successfully to statistical inference problems. However, it has been proved that information geometry opens a new paradigm useful for elucidation of information systems, intelligent systems, control systems, physical systems, mathematical systems, and so on.

There have been remarkable progresses recently in information geometry. For example, in the field of neurocomputing, a set of neural networks forms a neuro-manifold. Information geometry has become one of fundamental methods for analyzing neurocomputing and related areas. Its usefulness has also been recognized in multiterminal information theory and portfolio, in nonlinear systems and nonlinear prediction, in mathematical programming, in statistical inference and information theory of quantum mechanical systems, and so on. Its mathematical foundations have also shown a remarkable progress.

In spite of these developments, there were no expository textbooks covering the methods and applications of information geometry except for statistical ones. Although we published a booklet to show the wide scope of information geometry in 1993, it was unfortunately written in Japanese. It is our great pleasure to see its English translation. Mr. Daishi Harada has achieved an excellent work of translation.

In addition to correction of many misprints and errors found in the Japanese edition, we have made revision and rearrangement throughout the manuscript to make it as readable as possible. Also we have added several new topics, and even new sections and a new chapter such as §2.5, §3.2, §3.5, §3.8 and Chapter 7. The bibliography and the guide to it have largely been extended as well. These works were done by the authors after receiving the original translation, and it is the authors, not the translator, who should be responsible for the English writing of these parts.

This is a small booklet, however. We have presented a concise but comprehensive introduction to the mathematical foundation of information geometry in the first three chapters, while the other chapters are devoted to an overview
of wide areas of applications. Even though we could not show detailed and comprehensive explanations for many topics, we expect that the readers feel its flavor and prosperity from the description. It is our pleasure if the book would play a key role for further developments of information geometry.

Year 2000

Shun-ichi Amari
Hiroshi Nagaoka