Information Geometry (Math 676)

Fall 2022

Technicalities

Instructor: Clayton Shonkwiler (`clayton.shonkwiler@colostate.edu`)
Office: Weber 206C
Course web page: `http://www.math.colostate.edu/∼clayton/teaching/m676f22/`
Time/Location: 2:00–2:50 MWF, Eddy 111.
Office Hours: By appointment.

Summary of the Course

Information geometry has (slightly poetically) been described as "the geometry of decision making"; it is the story of a natural differential geometric structure possessed by families of probability distributions, and provides a coherent conceptual framework for parameter estimation, including performance bounds like the classical Cramér–Rao bound.

The first main goal of this course is to understand the Fisher information metric, which is a Riemannian metric on the manifold parametrizing a family of probability distributions induced by the Fisher information matrix. In turn, this metric induces an intrinsic distance, sometimes called the Fisher–Rao distance, which in general has desirable features (like symmetry) that other notions of distance between probability distributions (like the Kullback–Leibler divergence) lack.

From there, we will go deeper on the special structure of information manifolds, particularly conjugate connections, divergences, and the Legendre transformation. This will set us up for a differential geometric approach to statistical inference (e.g., trying to determine which distribution in a family your observed data came from).

The main running example throughout all of this is the family of multivariate Gaussian distributions, which is simple and concrete enough to allow for explicit calculations while still being of practical relevance. In this case one can see that the corresponding manifold is hyperbolic space, and as distance goes to zero the (symmetrized) Kullback–Leibler divergence limits to the intrinsic distance.

In general, the manifolds coming from exponential families (which are the focus of most of the literature) have global coordinates, though this is not true for general manifolds. So an optimistic potential outcome of the class is some insight into the situation when parameters satisfy manifold constraints, including orthogonality constraints.

There is no official text for the course, but the following books and papers will be useful resources:

- Shun-ichi Amari and Hiroshi Nagaoka, *Methods of Information Geometry*, Providence, RI, 2000.
- Shun-ichi Amar, *Information Geometry and Its Applications*, Springer, Tokyo, 2016.
- Nihat Ay, Jürgen Jost, Hông Vân Lê, and Lorenz Schwachhöfer, *Information Geometry*, Springer, Cham, 2017.
- Frank Nielsen, An elementary introduction to information geometry, *Entropy* 22(10):1100, 2020.

1 Grading

The goal of this course is for all of us to explore these connections between differential geometry and statistical inference and hopefully to learn some things that none of us knows yet. Therefore, the most important thing you can do is to actively participate in this project. Of course that means regular participation in our in-class discussions, and will also sometimes mean doing homework problems to develop concepts that require deeper reflection. I am required to assign you a grade, which will be based on:

Homework: 50%
Participation: 50%

2 Disclaimer

The course syllabus is a general plan for the course; deviations announced in class may be necessary.

3 Anticipated Schedule

Topic	**Weeks**
Overview of key concepts in differential geometry and statistics	3
Statistical models and the Fisher information metric	2
Dual connections, divergences, and the Legendre transformation	3
Geometry of statistical inference	3
Applications and extensions	3