Syllabus for STAT 547S: Topics on Symmetry in Statistics and Machine Learning - 2022WT2

Benjamin Bloem-Reddy

January 9, 2023

Overview

Symmetry plays a central role in many branches of science. In this graduate-level course, we will investigate some ways in which symmetry appears in the analysis of data. The first half of the course will be devoted to a body of work from the statistics literature on parameter estimation in the presence of symmetries in a statistical model. Much of that work is classical, and worth another look in the context of modern problems. In the second half, we will study some relatively recent work that incorporates symmetry into machine learning methods. If we're successful, connections between the two bodies of work will come into focus.

Our technical tools will based on group theory (the primary mathematical language of symmetry) and probability (the primary mathematical language of statistics). Background in group theory is helpful but not required; background in probability theory (and at least a familiarity with measure theory) is required.

Topics (tentative)

A tentative list of topics is below. The main constraint will be class time.

1. Equivariant estimation theory for location families (i.e., random variables in $\mathbb{R}$ with density $f(x-\theta)$). Notions of decision-theoretic optimality, connections to Bayesian estimation. (2 classes)
2. Mathematical background: a bit of group theory and measure theory for groups and spaces on which they act. (1-2 classes)
3. Equivariant estimation theory for group families. (1-2 classes)
4. Equivariant prediction. (1-2 classes)
5. Kernel methods with symmetry. (1-2 classes)
6. Symmetry in deep learning: data augmentation, convolutional neural networks. (1-2 classes)
7. Statistical learning in the presence of symmetry. (1 class)

Logistical information

Class meetings: T/Th, 1-2:30 pm, ESB 4192. Jan. 10 - Feb. 16 (12 class meetings).
Course instructor: Benjamin Bloem-Reddy

- Email / website: `benbr@stat.ubc.ca` / https://www.stat.ubc.ca/~benbr/
- Office: ESB 3168
- Office hour: Immediately after class on Thursdays (location TBD)

Pre-requisites: At least one graduate course in mathematical statistics (at the level of STAT 560) *and* one graduate course in probability with some measure theory (at the level of STAT 547C). I will try my best to make things understandable even to those who have not studied measure theoretic probability, but I will not shy away from it.

This course is intended primarily for graduate students in Statistics and related fields. Auditors are welcome, provided there is space. *Please see me if you intend to audit (and register as an auditor).*

Learning activities

You will learn through the following activities:

- Attending and actively participating in class meetings.
- Reading lecture notes and supplemental references (textbooks and journal/conference papers). The notes will be largely self-contained, though consulting the supplemental references is encouraged. In the first half of the course, I will draw primarily on the textbooks listed below; the second half will be based primarily on published research papers.
- Completing two written assignments.
- Completing a final project.

Assessment

Your overall grade will be based on:

- Two written assignments (40%). These will consist of questions of varying levels of difficulty; basic exercises can be completed for a grade between 80-84; questions of intermediate difficulty can be completed for a grade between 85-89; harder/more involved questions can be completed for a grade between 90-100. Students may work together to solve problems, and each student must write up their own solutions independently. Solutions must be typeset in LaTeX. (I will provide a template.)
- A final project (60%). The project will focus on a topic related to the course and consist of a 10-15 minute presentation followed by 10 minutes of Q&A. Presentations will be scheduled for the week after midterm break (week of Feb. 27); attendance at presentations other than your own is optional (but encouraged). Students may work alone or in pairs; topic selection will be made in consultation with the instructor. More details forthcoming.

Reference textbooks

Electronic versions of most of these are available freely online or through the UBC library.

- M. L. Eaton. *Group invariance in applications in statistics*. Vol. 1. Regional Conference Series in Probability and Statistics. Haywood, CA and Alexandria, VA: Institute of Mathematical Statistics and American Statistical Association, 1989
- O. Kallenberg. *Random Measures, Theory and Applications*. Springer International Publishing, 2017
- E. L. Lehmann and G. Casella. *Theory of Point Estimation*. 3rd ed. New York, NY: Springer–Verlag, 1998
- C. P. Robert. *The Bayesian Choice*. 2nd ed. Springer New York, 2007
- M. J. Schervish. *Theory of Statistics*. Springer-Verlag New York, 1995
- R. A. Wijsman. *Invariant measures on groups and their use in statistics*. Vol. 14. Lecture Notes–Monograph Series. Hayward, CA: Institute of Mathematical Statistics, 1990

University policies

UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise and so there are additional resources to access including those for survivors of sexual violence. UBC values respect for the person and ideas of all members of the academic community. Harassment and discrimination are not tolerated nor is suppression of academic freedom. UBC provides appropriate accommodation for students with disabilities and for religious observances. UBC values academic honesty and students are expected to acknowledge the ideas generated by others and to uphold the highest academic standards in all of their actions. Details of the policies and how to access support are available on https://senate.ubc.ca/policies-resources-support-student-success.