Calibration, Scoring, and Blackwell Approachability Advanced Topics in Statistical Learning, Spring 2024 Ryan Tibshirani

(This is just an outline of what we will cover. Notes from last year's lecture on scoring and calibration are available on the course website; they have a slightly different focus/exposition then what we'll pursue this year, but will nonetheless be quite relevant as a reference.)

the many faces of calibration:

- forecasting: sequential and probabilistic notations
- interval coverage generalizes to PIT calibration
- interlude: marginal calibration (critiques)
- conditional calibration (binary and general case)
- do linear ensembles retain calibration?

calibration is not enough—enter scoring rules:

- let's build up a scoring role from the ground up
- irreducible uncertainty, sharpness, calibration
- proper scores, score divergence, entropy
- examples: log score, Brier score, CRPS
- each has different operating characteristics!

... and along came Blackwell, many decades earlier:

- real-valued payoffs: von Neumann's setting
- vector valued payoffs: Blackwell's setting
- response satisfiability, halfspace satisfiability
- Blackwell approachability theorem
- (quick proof? with credit to M. Jordan)
- recalibration via Blackwell approachability