Minimax Optimal Estimation of Stability Under Distribution Shift

Yuanzhe Ma, Columbia IEOR
Joint work with Hongseok Namkoong (Columbia DRO) and Peter Glynn (Stanford MS&E)

2023

Introduction

- Distribution shift (training $\neq$ test distributions) often happens $\rightarrow$ model performance drops
- Evaluation of model robustness is important
- Question: how to do it in an *interpretable* way?
- Contributions:
  1. Developed an intuitive stability measure (for cost r.v.)
  2. Constructed an estimator that's minimax optimal: minimizes the worst-case risk
  3. Empirical results showing the utility of the stability measure

Formulation

- Setting: $R \sim P$ is cost, n i.i.d. data $R_i \stackrel{\text{iid}}{\sim} P$, y is a given threshold in the *cost scale* (e.g. $R_1 = 100, R_2 = 200, R_3 = 150, y = 400$)
- Stability measure (larger means more stable):

$$I_y(P) := \inf_{Q} \left\{ D_{\text{kl}}\left(Q \| P\right) : \mathbb{E}_Q[R] \geq y \right\}$$

- Duality result (Donsker and Varadhan, 1976):

$$I_y(P) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda y - \log \mathbb{E}_P[e^{\lambda R}] \right\}$$

- Estimator using dual formulation (replace P with $\widehat{P}_n$):

$$\widehat{I}_n := \sup_{\lambda \in \mathbb{R}} \left\{ \lambda y - \log \mathbb{E}_{\widehat{P}_n}[e^{\lambda R}] \right\},$$

where $\widehat{P}_n$ is the empirical distribution over the data $R_1, \ldots, R_n$

Theoretical Results

- Consider $\mathcal{Q} = \{\text{RVs similar to Gamma}(\alpha, \sigma)\}$ with with $\alpha \in (\frac{1}{2}, 1)$, $\sigma = \inf\{\lambda : \mathbb{E}_P[e^{\lambda R}] = \infty\}$ for $P \in \mathcal{Q}$

- Minimax rates of convergence achieved by our estimator $\widehat{I}_n$

$$\inf_{I_n} \sup_{P \in \mathcal{Q}} \mathbb{E}_P \left| I_n - I_y(P) \right| \gtrsim n^{-\left(\frac{1}{2} \wedge \frac{\alpha}{\sigma y}\right)}$$

$$\sup_{P \in \mathcal{Q}} \mathbb{E}_P \left| \widehat{I}_n - I_y(P) \right| \lesssim n^{-\left(\frac{1}{2} \wedge \frac{\alpha}{\sigma y}\right)}$$

- Whether $e^{\lambda^\star R}$ has a second moment: easy/ hard case, where $\lambda^\star$ is the optimal dual variable that grows with y

- Higher σ (lighter-tailed RV) or threshold y: harder, since extreme events are less likely to be observed, which relates to our I

Proof Idea

$$\inf_{I_n} \sup_{P \in \mathcal{Q}} \mathbb{E}_P \left| I_n - I_y(P) \right| \gtrapprox n^{-\left(\frac{1}{2} \wedge \frac{\alpha}{\sigma y}\right)}$$

$$\sup_{P \in \mathcal{Q}} \mathbb{E}_P \left| \widehat{I}_n - I_y(P) \right| \lessapprox n^{-\left(\frac{1}{2} \wedge \frac{\alpha}{\sigma y}\right)}$$

- Upper bound: Dudley's integral entropy bound
- Lower bound: Le Cam's method: constructed $P_1, P_2 \in \mathcal{Q}$ with small $\|P_1 - P_2\|_{\mathrm{TV}}$ but large $|I_y(P_1) - I_y(P_2)|$

Empirical Results

- Experiments on sequential decision-making and supervised learning frameworks
- Can differentiate between brittle vs. robust models, in contrast to typical average-case performance metrics

Conclusion and Future Directions

- Proposed an interpretable stability measure with a minimax estimator based on dual formulation
- Future directions:
    - Asymptotic results
    - From one-dim to multi-dim
    - More general notions to quantify distribution shifts: Wasserstein distance/ likelihood ratio bound/ alignment of covariance matrices
    - More structured distribution shifts, e.g. subpopulation shifts
    - Connections to large deviations theory/ estimation of rare event probabilities

Thank you!

Contact info: *ym2865@columbia.edu*

Paper link: *https://arxiv.org/pdf/2212.06338.pdf*