Robust Data-Driven Decisions Under Model Uncertainty

XIAOYU CHENG, Northwestern University

In data-driven decisions, while the extrapolation from sample data is often based on observable similarities within a population, one also needs to take into account possible unobserved heterogeneity across individuals in the population. Because of the unobservability, a decision-maker (DM) can be uncertain about not only how the individuals might vary but also how different individuals are going to be sampled from the population. As a result, she might worry about the possibility that the sample data are more from one type of individual, whereas future draws that determine the payoff of decisions may be more of a different type. This paper captures this concern by considering a decision environment where the underlying data-generating process (DGP) is a sequence of independent but possibly non-identical distributions. Specifically, the DM observes sample data given by realizations of marginal distributions of a DGP and then makes a decision whose payoff depends only on future realizations of the same DGP. In addition, the DM faces model uncertainty in the form of ambiguity, i.e., she only knows there is a set of possible DGPs but cannot form any probabilistic assessment over them.

For making decisions, I suppose the DM applies the maxmin expected-utility (MEU) criterion [1] to cope with ambiguity. That is, she makes an optimal decision under the worst possible DGP that she contemplates. The DM can either make a data-free decision based on her initial belief, or a data-driven decision based on her updated belief taking into account the sample data. Given these two types of decisions, I study updating rules in terms of how to guarantee the data-driven decisions to be better than the data-free decisions according to objective payoff, i.e., the expected utility under the true DGP that governs the future uncertainty. In other words, while the DM makes decisions considering the worst case, the quality of her decisions will be evaluated against the ground truth. When an updating rule can guarantee improvement for all possible DGPs in the initial belief, the data-driven decisions are said to robustly improve upon the data-free decisions.

In this paper, I formalize two achievable notions of how data-driven decisions can robustly improve upon data-free decisions across decision problems. I show that these two notions are both equivalent to the intuitive requirement that the updated set of DGPs should accommodate (i.e., contain with some technical generalization) the true DGP that generates the data (Theorem 4.2, 4.6, Corollary 4.7).

Based on this equivalence, I further study updating rules in terms of this property. In Section 2 of the paper, I make a critical observation that in the presence of independent but non-identical distributions, common updating rules such as maximum likelihood and Bayesian updating, can almost surely rule out the true DGP. Thus, implied by the previous result, they can almost surely lead to strictly worse decisions than simply ignoring the data.

To achieve robustly better decisions in such a decision environment, I develop two new updating rules in this paper. When the sample size can increase unboundedly to infinity, I propose the average-then-update rule that guarantees to accommodate the true DGP asymptotically almost surely (Theorem 5.4). When the sample size is finite, I propose the robust i.i.d. statistical tests that guarantee to accommodate the true DGP with a pre-specified probability (Theorem 5.7). Under the robust i.i.d. statistical tests, one effectively obtains a confidence region of the true DGP. While constructing confidence regions for non-identically distributed DGP can be computationally challenging, the robust i.i.d. statistical tests require constructing confidence regions for only the independent and identically distributed (i.i.d.) DGPs. Therefore, they are very tractable and thus easy to implement.
Finally, the decision framework studied in this paper is often applied to model economic problems such as dynamic portfolio choice, asset pricing, and social learning under model uncertainty and under ambiguity. For those problems, the existing literature obtains conclusions primarily by assuming that the DM applies full Bayesian updating (some may refer to it as prior-by-prior updating). However, the asymptotic result under full Bayesian updating is often hard to solve. In a commonly studied model of learning from Gaussian signals with ambiguous variances, I show that applying the average-then-update rule reduces to a simple and intuitive step. More importantly, applying the average-then-update rule also implies that learning is significantly more effective than applying full Bayesian updating (Proposition 6.2). This learning outcome proves to be more intuitive in such a model. In addition, I provide a more concrete illustration of both proposed updating rules in Section 6 by studying a Bernoulli model with ambiguous nuisance parameters. There, I show that the updated sets of DGPs have tractable expressions and intuitive interpretations.

Full paper: https://arxiv.org/abs/2205.04573

CCS Concepts:
• Applied computing → Economics; Multi-criterion optimization and decision-making;
• Mathematics of computing → Hypothesis testing and confidence interval computation.

Additional Key Words and Phrases: statistical decisions, robustness, model uncertainty, ambiguity, updating

ACM Reference Format:
Xiaoyu Cheng. 2022. Robust Data-Driven Decisions Under Model Uncertainty. In Proceedings of the 23rd ACM Conference on Economics and Computation (EC ’22), July 11–15, 2022, Boulder, CO, USA. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3490486.3538356

REFERENCES
[1] Itzhak Gilboa and David Schmeidler. 1989. Maxmin expected utility with non-unique prior. Journal of Mathematical Economics 18, 2 (1989), 141–153. https://doi.org/10.1016/0304-4068(89)90018-9