Report 1

Lu Sun

1 Introduction

1. **History**

   (a)stochastic optimization(SP):

   $$\max_{\mathbf{x}\in X} \mathbf{E}_{F_\epsilon}[h(\mathbf{x},\epsilon)]$$

   -F_ϵ (distribution function)need to be exact, otherwise result in sub-optimal solutions

   -no robustness in the error of distribution;need to much computation

   (b)(static) robust optimization:

   $$\max_{\mathbf{x}\in X} \min_{\epsilon\in support} h(\mathbf{x},\epsilon)$$

   -intractable: include too much unrealistic distribution(single point) on support set(the solution may not be realistic to worst case)

   (c)distributional robust optimization(DRO):

   -also named as DRSP(distributional robust stochastic programming)

   -an intermediate approach between SP and robust optimization

2. **Model**

   (a)prerequisite:

   -known certain observed samples

   eg: sum-of-squares polynomial density functions of known degrees/ a fixed distance, which measured by KL divergence, from a nominal distribution

   -known certain training data

   eg: data-driven sets with statistical hypothesis tests

   -other statistical information

   eg: higher-order moments;conditional probabilities; conditional moments; marginal distributions

   (b)properties:

   -uncertain(exact distribution is unknown), fictitious, decision-making problems, data-driven optimization, apply to worst case

   $$\begin{cases} pros: & \text{uncertain,fictitious} \\ cons: & \text{pathological discrete distribution} \end{cases}$$

   (c)classes:

   #1 : continuous action space; learn optimal strategy(make decision)

#2 : under Wasserstein metric, change to finite convex programs(eg tractable linear programs/ conic program)

#3 :data-driven optimization

(d)composition:

$$\max_{\mathbf{x}\in X}\min_{F_\epsilon\in D}\mathbf{E}_{F_\epsilon}[h(\mathbf{x},\epsilon)]$$

Here, D is a set of distributions and the purpose is to give best $\mathbf{x}\in X$. In order to choose a proper D, following things need to be taken into account:

-tractability

-practical(statistical) meanings

-performance(eg: the potential loss comparing to the benchmark cases)

3. **Software**

   ROME

4. **Comparison with other methods**

   By the use of several carefully designed data structures, DRO provides at little extra cost compared to empirical risk minimization & stochastic gradient methods

2 Application

1. **Finance**

   (a)data source:

   portfolio optimization problem(insurance company);

   risk aggregation problem(insurance company);

   price-post learning

   (b)origin:

   DRO method comes from risk-managing decision-make optimization with portfolio problem

2. **Engineering**

   (a)purpose:

   give out power generation solution by minimizing cost

Main point

1. **2018_Klerk_SIAM**

   (a)introduce a sets(sum-of-squares polynomial density functions of known degrees) instead of pathological distribution

   (b)apply to portfolio optimization problem and risk aggregation problem

2. **2018_Gao_book**

   (a)give method to choose distributional function

   (b)give simple algorithms for local optima

   (c)extend to optimization problems: orthogonal constraints and coupled constraints over simplex set and polytopes

   (d)explain by convergence rate and examples

3. **2018_Esfahani_MP**

   (a)under Wasserstein metric and ball, change DOR to finite convex programs(eg: tractable linear programs)

   (b)global optimization techniques

   (c)apply to mean-risk portfolio optimization

4. **2018_Bertsimas_MP**

   (a)data-driven distributionally function

   (b)apply to portfolio management

5. **2017_Ye_ppt**

   (a)stochastic optimization

   $$\max_{\mathbf{x}\in X} \mathbf{E}_{F_\epsilon}[h(\mathbf{x},\epsilon)]$$

   F_ϵ (distribution function)need to be exact, otherwise result in sub-optimal solutions

   (b)(static)robust optimization

   $$\max_{\mathbf{x}\in X} \min_{\epsilon\in support} h(\mathbf{x},\epsilon)$$

   Robust to any distribution, but may make bad decision in worst-case

   (c)DRO

   intermediate approach: robustness in the error of distribution; provide realistic non-single-point distribution on the support

6. **2017_Chen_IEEE**

   (a)distance-based distributionally robust unit commitment(DB-DRUC)

   (b)apply to wind power generation

   (c)minimize the expected cost under the worst case wind distributions(fixed distance from a nominal distribution)

7. **2017_Wang_arXiv**

   (a)optimal power flow

   (b)risk term(objective function): penalties for load shedding, wind generation curtailment and line overload

   (b)ambiguity set: second-order moment and Wasserstein distance(change to conic program)

   (c)test systems: 5-bus/IEEE 118-bus/ Polish 2736-bus

8. **2016_Namkoong_AINIP**

   (a)give calibrated confidence intervals on performance

   (b)weight more on observations (loss)

   (c)stochastic gradient method is a little better than DRO