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)

-classes:

(1)without recourse, the static settings;

(2)with recourse, the dynamic settings;

(3)connection with stochastic optimization and risk theory, distributionally robust optimization

-applications: inventory and logistics, finance, revenue mangement, queueing networks, machine learning, energy systems, public good

-static: one stage, decide at once

(c)distributional robust optimization(DRO):

-also named as DRSP(distributional robust stochastic programming) or DRSO(distributional robust stochastic optimization)

-an intermediate approach between SP and robust optimization, i.e.,"robustifying" stochastic optimization

-SP with uncertainty, a chosen set of distributions

-semidefinite optimization model with known 1st and 2nd moments, proof tractable approximation; directly on distributions, "well structured" with high probability to contain the distribution

2. **Model**

   (a)composition:

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

Here, D is uncertain set,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)

(b)ways to choose D

-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

(1)robust optimization directly on the probability distributions

eg: data-driven sets with statistical hypothesis tests

(2)robust optimization using moment information

eg: other statistical information: higher-order moments;conditional probabilities; conditional moments; marginal distributions

(c)properties:

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

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

(d)classes:

#1 : decision making problems: likelihood robust optimization(LRO);continuous action space; learn optimal strategy(make decision)

#2 :Find distributional function: data-driven optimization, under Wasserstein metric, change to finite convex programs(eg tractable linear programs/ conic program)

#3 :Classify problems: DRO be used in Logistic Regression

3. **Software**

   ROME

4. **Comparison with other methods**

   (a) exist method

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

   (b) exist function

   data-driven function is resonable and tractable than exist popular function

5. **Theorem related**

   (a) existence for worst-case

   Conditions for worst-case to exist

   (b) tractability for data-driven DRSO

   Conditions for tractable; approximate for violate conditions

   (c) duality problem

   By using Wasserstein Distance, duality problem can be translated

   (d) problem related

   eg: infinite dimensional process, worst-case in risk managment

   (e)asymptotic behavior

2 Application

management science, system design, optimal control

1. **Finance**

   (a)data source:

   portfolio optimization problem(insurance company);

   risk aggregation problem(insurance company);

   price-post learning

   Nash equilibrium problem

   public good

   (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

   -using wind power (partial information) to give out wind power probability distribution

   (b)energy systems:

3. **Traffic**

   (a) on-line signal control:

4. **Classify**

   (a)Logistic Regression:

   using DRO to define upper and lower confidence bounds and giving classifier

5. **newsvendor**

3 Basic definition

from(WIKI)

1. **Wasserstein metric**

   The Wasserstein or Kantorovich-Rubinstein metric or distance is a distance function defined between probability distributions on a given metric space M.

   $$W_p(\mu,\nu) := \left(\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{M\times M} d(x,y)^p \, \mathrm{d}\gamma(x,y) \right)^{1/p},$$

   or

   $$W_p(\mu,\nu)^p = \inf \mathbf{E}\big[d(X,Y)^p\big]$$

   -measurement

   -distance

   -ball

2. **worst-case analysis**

   a solution is evaluated using the realization of the uncertainty that is most unfavorable

   The following linear programming problem

   $$\max_{x,y} \{3x+2y\} \ \text{ subject to } \ x,y\geq 0; cx+dy \leq 10, \forall (c,d)\in P$$

   where P is a given subset of $\mathbb{R}^2$.

   Satisfy the worst$(c,d)\in P$ pertaining to (x,y) , namely the pair $(c,d)\in P$ that maximizes the value of $cx+dy$ for the given value of (x,y) .

   If P is not a finite set, then this problem is a linear semi-infinite programming problem, namely a linear programming problem with finitely many decision variables and infinitely many constraints.

3. **tractable**

   Namely, Computational complexity theory

4. **asymptotic**

5. **portfolio problem**

   The term portfolio refers to any combination of financial assets such as stocks, bonds and cash.

4 Problems last time

1. examples and difference between RO and DRO

   classic RO: minimax or maximini optimization problem(wiki), infinite constrains

   DRO: data-driven, finite constrains

2. way to construct data-driven distributionally function sets

   By moments.

3. risk managment area(finance/security/health/supply chain)

   undo

4. game theory: difference between definition of continuity

   undo

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

9. **2016_Liu_IEEE**

   (a)SP model with uncertainty

   (b)two-stage power system

   (c)unit commitment: effectiveness

10. **2016_Gao_arXiv**

    (a) Wasserstein Distance: an empirical distribution resulting from available data;advantage of it:one–reasonable than functions at hand;two–tractability

    (b)the way to choose distributional function set;popular functions but not appropriate

    (c)proof:

    -conditions for the existence of worst-case distribution

    -worst-case can have concise and clear structure

    -data-driven DRSO tractable with tools in RO

    -strong duality

    -can be used to infinite dimensional process and worst-case

    (d)note: a hedge is a risk management technique used to reduce any substantial losses or gains suffered by an individual or an organization

11. **2015_Sun_MOOR**

    (a) construct data-driven distribution functions

    (b) apply to Nash equilibrium problem (one-stage problem)

12. **2015_Liu_TRP**

    (a) on-line signal control (two-stage problem) traffic problem

    (b) linear decision rule(LDR)

    (c) DRO: guarantee performance by off-line computation and improve traffic

13. **2015_Bian_IEEE**

    (a) using partial information of wind power to find wind power probability distribution

    (b) SP change into BMI(bilinear matrix inequality problem)

14. **2015_Abadeh_AINIP**

    (a) apply distributionally robust method to logistic regression problem

    (b) operate within Wasserstein ball under worst-case

15. **2014_Wiesemann_OR**

    (a) solve DRO problems

    (b) domain: standardized ambiguity sets

    (c) conditions for computationally tractable; approximations for violate conditions

16. **2014_Wang_arXiv**

    (a)propose LRO(likelihood robust optimization) for optimal decision-making problems

    (b)data-driven distributional function

    (c)analyses model by Bayesian statistics and empirical likelihood theory

    (d)proof for asymptotic behavior

    (e)apply to newsvendor problem and portfolio selection problem

17. **2014_Gabrel_EJOOR**

    (a)overview of RO in theory and practice

    (b)classes:

    -without recourse, the static settings;

    -with recourse, the dynamic settings;

    -connection with stochastic optimization and risk theory, distributionally robust optimization

(c)applications: inventory and logistics, finance, revenue mangement, queueing networks, machine learning, energy systems, public good

(d)DRO: linking uncertainty sets to risk theory; application in public good and energy systems