Stochastic Optimization

Ricardo Fukasawa

Department of Combinatorics & Optimization
University of Waterloo

July 8, 2022

UNIVERSITY OF WATERLOO
FACULTY OF MATHEMATICS
Department of Combinatorics
and Optimization

Outline

Disclaimers

The only part of my talk that is CERTAIN:

- IP/OR focus
- Incomplete

Disclaimers

The only part of my talk that is CERTAIN:

- IP/OR focus
- Incomplete

My goal:

- Give a very basic overview of the several issues that arise when considering uncertainty

Stochastic IP

$$
\begin{array}{ll}
\min & c(\xi)^T x \\
\text{s.t.} & A(\xi)x \leq b(\xi) \\
 & x \in \mathbb{Z}^p \times \mathbb{R}^{n-p}
\end{array}
$$

where ξ is a random vector.

Stochastic IP

$$\begin{array}{ll} \min & c(\xi)^T x \\ \text{s.t.} & A(\xi)x \leq b(\xi) \\ & x \in \mathbb{Z}^p \times \mathbb{R}^{n-p} \end{array}$$

where ξ is a random vector.

But:

- What does this mean?
- How do we solve it?

Stochastic IP

$$\begin{array}{ll} \min & c(\xi)^T x \\ \text{s.t.} & A(\xi)x \leq b(\xi) \\ & x \in \mathbb{Z}^p \times \mathbb{R}^{n-p} \end{array}$$

where ξ is a random vector.

But:

- What does this mean?
- How do we solve it?

Typical issues:

1. Choose a desired interpretation
2. Formulate a deterministic problem that can (approximately) solve it
3. Solve the deterministic problem

Stochastic IP

$$\begin{array}{ll} \min & c(\xi)^T x \\ \text{s.t.} & A(\xi)x \le b(\xi) \\ & x \in \mathbb{Z}^p \times \mathbb{R}^{n-p} \end{array}$$

where ξ is a random vector.

But:

- What does this mean?
- How do we solve it?

Typical issues:

1. Choose a desired interpretation
2. Formulate a deterministic problem that can (approximately) solve it
3. Solve the deterministic problem

Hidden challenge: dealing with unknown probability distributions. Example:

- Computing $\mathbb{E}[\max\{x^T\xi, b\}]$ where x is an arbitrary vector and each component of ξ is an i.i.d. uniform random variable is #P-hard (Hanasusanto, Kuhn, and Wiesemann, 2016)

Outline

Two-stage stochastic programs

$$\begin{array}{ll} \min & \mathbb{E}[Q(x,\xi)] \\ \text{s.t.} & x \in X \end{array} \tag{SP}$$

where

$$Q(x,\xi) := \begin{array}{ll} \min & q(\xi)^T y \\ \text{s.t.} & W(\xi) y \le h(\xi) - T(\xi) x \end{array} \tag{REC}$$

- x are called first-stage variables
- y are called second-stage variables

Hardness

Graph reliability problem

Given directed graph $G = (V, E)$, $u, v \in V$, count how many subgraphs $H = (V, F)$ of G have a $u - v$ path

Introduce (v, u), $E' = E \cup (v, u)$ $m = |E|$

* y_{ij}, $\forall ij \in E'$

* $g_{ij} = \begin{cases} -2, & \text{w.p. } 1/2 \rightarrow \text{fail} \\ 0, & \text{w.p. } 1/2 \rightarrow \text{exists,} \end{cases}$ $\forall ij \in E$

& $r_{vu} = 1$ w.p. 1

↳ 2^m scenarios

Find max cost circulation in scenario S.

Let's say capacities of arcs are x.

If $\exists$ circul. w/ only edges in $E \Rightarrow$ cost is $\geq x$

Else ≤ 0.

$\max \; \mathbb{E}[Q(x, \xi)]$

s.t. $x = 1$

$Q(x, \xi)$ is cost of circulation.

D has reliability $\geq R$

$\Leftrightarrow z^* \geq \frac{R}{2^m}$

Sample average approximation (Kleywegt, Shapiro, and Mello, 2002)

Addresses issue 2.

Instead of solving

$$z^* := \begin{array}{ll} \min & \mathbb{E}[Q(x,\xi)] \\ \text{s.t.} & x \in X \end{array} \qquad \text{(SP)}$$

take $\xi^1, \ldots, \xi^N$ i.i.d. samples of ξ and solve

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} Q(x,\xi^i) \\ \text{s.t.} & x \in X \end{array} \qquad \text{(SAA-N)}$$

Let x^N be an optimal solution to (SAA-N)

Sample average approximation (Kleywegt, Shapiro, and Mello, 2002)

Proposition

$\mathbb{E}[z^N] \leq z^*$

Let x^* opt. to (SP)

$z^* = \mathbb{E}[Q(x^*, \xi)]$

x^* feas. for (SAA-N)

$z_N \leq \frac{1}{N} \sum_{i=1}^{N} Q(x^*, \xi^i)$

$\mathbb{E}[z_N] \leq \frac{1}{N} \sum_{i=1}^{N} \underbrace{\mathbb{E}[Q(x^*, \xi^i)]}_{z^*}$

Sample average approximation (Kleywegt, Shapiro, and Mello, 2002)

Proposition

$\mathbb{E}[z^N] \leq z^*$

Proposition

$\mathbb{E}[z^N] \leq \mathbb{E}[z^{N+1}]$

Sample average approximation (Kleywegt, Shapiro, and Mello, 2002)

Proposition

$\mathbb{E}[z^N] \leq z^*$

Proposition

$\mathbb{E}[z^N] \leq \mathbb{E}[z^{N+1}]$

Theorem

Assume X is compact, $Q(x, \xi)$ is real-valued for all $x \in X$ and all ξ. Then $z^N \rightarrow z^*$ and every limit point of $\{x^N\}$ solves (SP) with probability 1.

Sample average approximation (Kleywegt, Shapiro, and Mello, 2002)

Proposition

$\mathbb{E}[z^N] \leq z^*$

Proposition

$\mathbb{E}[z^N] \leq \mathbb{E}[z^{N+1}]$

Theorem

Assume X is compact, $Q(x, \xi)$ is real-valued for all $x \in X$ and all ξ. Then $z^N \rightarrow z^*$ and every limit point of $\{x^N\}$ solves (SP) with probability 1.

Ahmed and Shapiro, 2002 extends it to cases with integral recourse.

Solving (SAA-LP)

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & Ax \le b \end{array} \tag{SAA-LP}$$

where

$$Q(x, \xi) := \begin{array}{ll} \min & q(\xi)^T y \\ \text{s.t.} & W(\xi) y \le h(\xi) - T(\xi) x \end{array} \tag{REC}$$

Solving (SAA-LP)

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & Ax \le b \end{array} \qquad \text{(SAA-LP)}$$

where

$$Q(x, \xi) := \begin{array}{ll} \min & q(\xi)^T y \\ \text{s.t.} & W(\xi) y \le h(\xi) - T(\xi) x \end{array} \qquad \text{(REC)}$$

Given a fixed ξ, different x values give different RHS for (REC).

Value function of an LP:

$$V(d) := \begin{array}{ll} \min & g^T y \\ \text{s.t.} & Dy \le d \end{array}$$

is a PWL convex function

Solving (SAA-LP)

$$\begin{array}{ll} \min & q(\xi)^T y \\ \text{s.t.} & W(\xi) y \leq h(\xi) - T(\xi) x \end{array}$$

$$= \begin{array}{ll} \max & [h(\xi) - T(\xi)]^T w \\ \text{s.t.} & W(\xi)^T w = q(\xi) \\ & w \leq 0 \end{array} = \max_{k=1,\dots,K(\xi)} [h(\xi) - T(\xi)]^T w(\xi)^k$$

Solving (SAA-LP)

$$\begin{array}{ll} \min & q(\xi)^T y \\ \text{s.t.} & W(\xi) y \leq h(\xi) - T(\xi) x \end{array}$$

$$= \begin{array}{ll} \max & [h(\xi) - T(\xi)]^T w \\ \text{s.t.} & W(\xi)^T w = q(\xi) \\ & w \leq 0 \end{array} = \max_{k=1,\dots,K(\xi)} [h(\xi) - T(\xi)]^T w(\xi)^k$$

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & Ax \leq b \end{array} = \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} \theta_i \\ \text{s.t.} & Ax \leq b \\ & \theta_i \geq [h(\xi) - T(\xi)]^T w(\xi^i)^k, \forall k = 1, \dots, K(\xi^i) \end{array}$$

which is solved in a cutting plane fashion.

Solving (SAA-LP)

$$\begin{array}{ll} \min & q(\xi)^T y \\ \text{s.t.} & W(\xi) y \le h(\xi) - T(\xi) x \end{array}$$

$$= \begin{array}{ll} \max & [h(\xi) - T(\xi)]^T w \\ \text{s.t.} & W(\xi)^T w = q(\xi) \\ & w \le 0 \end{array} = \max_{k=1,\ldots,K(\xi)} [h(\xi) - T(\xi)]^T w(\xi)^k$$

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^N \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & Ax \le b \end{array} = \begin{array}{ll} \min & \sum_{i=1}^N \frac{1}{N} \theta_i \\ \text{s.t.} & Ax \le b \\ & \theta_i \ge [h(\xi) - T(\xi)]^T w(\xi^i)^k, \forall k = 1, \ldots, K(\xi^i) \end{array}$$

which is solved in a cutting plane fashion.

This is called Benders' decomposition.

Solving (SAA-LP)

$$\begin{array}{ll} \min & q(\xi)^T y \\ \text{s.t.} & W(\xi)y \leq h(\xi) - T(\xi)x \end{array}$$

$$= \begin{array}{ll} \max & [h(\xi) - T(\xi)]^T w \\ \text{s.t.} & W(\xi)^T w = q(\xi) \\ & w \leq 0 \end{array} = \max_{k=1,\ldots,K(\xi)} [h(\xi) - T(\xi)]^T w(\xi)^k$$

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^N \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & Ax \leq b \end{array} = \begin{array}{ll} \min & \sum_{i=1}^N \frac{1}{N} \theta_i \\ \text{s.t.} & Ax \leq b \\ & \theta_i \geq [h(\xi) - T(\xi)]^T w(\xi^i)^k, \forall k = 1, \ldots, K(\xi^i) \end{array}$$

which is solved in a cutting plane fashion.

This is called Benders' decomposition.

Assumes second stage problem always has an optimal solution and does not work if second stage variables are integer.

Multistage stochastic programs

Generalizes two-stage to multiple, where uncertainty is revealed in stages and decisions can be made in each stage considering all previous uncertainty realizations as deterministic.

Multistage stochastic programs

Generalizes two-stage to multiple, where uncertainty is revealed in stages and decisions can be made in each stage considering all previous uncertainty realizations as deterministic.

Much more complex, with tons of research questions on their own.

- Stochastic Dual Dynamic Programming: Pereira and Pinto, 1991
  Example performance:
  Matos, Philpott, and Finardi, 2015: 300 power plants, 120 stages, 20 realizations per stage: 2% gap within 2 hours.
- Stochastic Dual Dynamic Integer Programming: Zou, Ahmed, and Sun, 2019
  Example performance:
  9 stages, 2% gap in 3h.

Reducing scenarios (Song and Luedtke, 2015)

Addresses issue 3

Reducing scenarios (Song and Luedtke, 2015)

Addresses issue 3

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & x \in X \end{array}$$

where

$$Q(x, \xi) := \begin{array}{ll} \min & q^T y \\ \text{s.t.} & Wy \leq h(\xi) - T(\xi)x \end{array}$$

Let $\mathcal{N} = \{P_1, \ldots, P_L\}$ be a partition of $1, \ldots, N$.

Reducing scenarios (Song and Luedtke, 2015)

Addresses issue 3

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & x \in X \end{array}$$

where

$$Q(x, \xi) := \begin{array}{ll} \min & q^T y \\ \text{s.t.} & Wy \leq h(\xi) - T(\xi)x \end{array}$$

Let $\mathcal{N} = \{P_1, \ldots, P_L\}$ be a partition of $1, \ldots, N$.

$y^P = \sum_{i \in P} y^i$, $T^P = \sum_{i \in P} T(\xi^i)$, $h^P = \sum_{i \in P} h(\xi^i)$

Reducing scenarios (Song and Luedtke, 2015)

Addresses issue 3

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & x \in X \end{array}$$

where

$$Q(x, \xi) := \begin{array}{ll} \min & q^T y \\ \text{s.t.} & Wy \leq h(\xi) - T(\xi)x \end{array}$$

Let $\mathcal{N} = \{P_1, \ldots, P_L\}$ be a partition of $1, \ldots, N$.

$y^P = \sum_{i \in P} y^i$, $T^P = \sum_{i \in P} T(\xi^i)$, $h^P = \sum_{i \in P} h(\xi^i)$

$$\begin{array}{ll} \min & \sum_{P \in \mathcal{N}} \frac{1}{N} q^T y^P \\ \text{s.t.} & x \in X \\ & Wy^P \leq h^P - T^P x, \forall P \in \mathcal{N} \end{array}$$

Reducing scenarios (Song and Luedtke, 2015)

Addresses issue 3

$$z^N := \begin{array}{ll} \min & \sum_{i=1}^{N} \frac{1}{N} Q(x, \xi^i) \\ \text{s.t.} & x \in X \end{array}$$

where

$$Q(x, \xi) := \begin{array}{ll} \min & q^T y \\ \text{s.t.} & Wy \leq h(\xi) - T(\xi)x \end{array}$$

Let $\mathcal{N} = \{P_1, \ldots, P_L\}$ be a partition of $1, \ldots, N$.

$y^P = \sum_{i \in P} y^i$, $T^P = \sum_{i \in P} T(\xi^i)$, $h^P = \sum_{i \in P} h(\xi^i)$

$$\begin{array}{ll} \min & \sum_{P \in \mathcal{N}} \frac{1}{N} q^T y^P \\ \text{s.t.} & x \in X \\ & Wy^P \leq h^P - T^P x, \forall P \in \mathcal{N} \end{array}$$

Song and Luedtke, 2015: Propose a way to iteratively refine $\mathcal{N}$ until we reach optimal solution to original problem.

Stochastic cutting planes (Bertsimas and Li, 2022)

Addresses issue 3

Idea: Adapt cutting plane approach to Stochastic MINLP problems giving high probability of getting good solution with much lower number of cuts.

Other measures of risk

Adresses issue 1

Expected value is risk neutral:

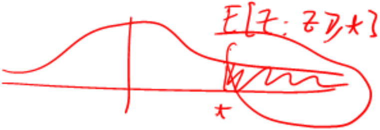
Which is better:

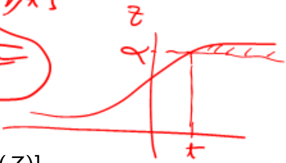
- A solution that 90% of the time has cost 1,000 and 10% of the time has cost 0
- A solution that 100% of the time has cost 901

depends on how risk-averse you are.

Ruszczyński and Shapiro, 2006: Investigate optimization of risk functions (not recommend for reading group).

Other measures of risk





Adresses issue 1

- Value at Risk: $V@R_\alpha(Z) := \inf\{t : \mathbb{P}[Z \leq t] \geq \alpha\}$
- Conditional Value-at-Risk: $CV@R_\alpha(Z) := \mathbb{E}[Z | Z \geq V@R_\alpha(Z)]$

Lemma (Rockafellar and Uryasev, 2002)

If $\mathbb{P}[Z = V@R(Z)] = 0$ then $CV@R_\alpha(Z) = \min_{\gamma \in \mathbb{R}} \left\{ \gamma + \frac{1}{1-\alpha} \mathbb{E}[Z - \gamma]_+ \right\}$

Other measures of risk

Adresses issue 1

- Value at Risk: $V@R_\alpha(Z) := \inf\{t : \mathbb{P}[Z \leq t] \geq \alpha\}$
- Conditional Value-at-Risk: $CV@R_\alpha(Z) := \mathbb{E}[Z | Z \geq V@R_\alpha(Z)]$

Lemma (Rockafellar and Uryasev, 2002)

If $\mathbb{P}[Z = V@R(Z)] = 0$ then $CV@R_\alpha(Z) = \min_{\gamma \in \mathbb{R}} \left\{ \gamma + \frac{1}{1-\alpha} \mathbb{E}[Z - \gamma]_+ \right\}$

If $Z = Q(x, \xi)$ with ξ having finite support distribution $\mathbb{P}(\xi = \xi^i) = p_i, i = 1, \ldots, N$ then:

$$\begin{array}{ll} \min & CV@R_\alpha(Q(x,\xi)) \\ \text{s.t.} & x \in X \end{array} = \begin{array}{ll} \min & \gamma + \frac{1}{1-\alpha} \sum_{i=1}^{N} p_i w_i \\ \text{s.t.} & x \in X \\ & w_i \geq Q(x, \xi^i) - \gamma, i = 1, \ldots, N \\ & \gamma \in \mathbb{R}, w \geq 0 \end{array} \tag{1}$$

Outline

Chance constraint

(Joint) chance-constraint:

$$z_\epsilon^* := \begin{array}{ll} \min & c^T x \\ \text{s.t.} & x \in X \\ & \mathbb{P}\{f_i(x,\xi) \le 0, \forall i = 1,\dots,m\} \ge 1-\epsilon \end{array} \tag{CC}$$

Individual chance-constraint: $m = 1$ (our focus next)

Tractable approximation (Nemirovskii and Shapiro, 2006)

Suppose $\psi : \mathbb{R} \to \mathbb{R}$ is:

- Nonnegative
- nondecreasing
- convex
- $\psi(a) > \psi(0) = 1, \forall z > 0.$

Tractable approximation (Nemirovskii and Shapiro, 2006)

Suppose $\psi : \mathbb{R} \to \mathbb{R}$ is:

- Nonnegative
- nondecreasing
- convex
- $\psi(a) > \psi(0) = 1, \forall z > 0.$

If Z is a random variable and $t > 0$

$$\mathbb{E}[\psi(tZ)] \geq \mathbb{E}[\mathbb{I}_{[0,+\infty]}(tZ)] = \mathbb{P}[tZ \geq 0] = \mathbb{P}[Z \geq 0] \geq \mathbb{P}[Z > 0]$$

Tractable approximation (Nemirovskii and Shapiro, 2006)

Suppose $\psi : \mathbb{R} \to \mathbb{R}$ is:

- Nonnegative
- nondecreasing
- convex
- $\psi(a) > \psi(0) = 1, \forall z > 0.$

If Z is a random variable and $t > 0$

$$\mathbb{E}[\psi(tZ)] \geq \mathbb{E}[\mathbb{I}_{[0,+\infty]}(tZ)] = \mathbb{P}[tZ \geq 0] = \mathbb{P}[Z \geq 0] \geq \mathbb{P}[Z > 0]$$

Consider

$$S = \{x : \mathbb{P}\{f(x,\xi) \leq 0\} \geq 1 - \epsilon\} = \{x : \mathbb{P}\{f(x,\xi) > 0\} \leq \epsilon\}$$

Tractable approximation (Nemirovskii and Shapiro, 2006)

Suppose $\psi : \mathbb{R} \to \mathbb{R}$ is:

- Nonnegative
- nondecreasing
- convex
- $\psi(a) > \psi(0) = 1, \forall z > 0$.

If Z is a random variable and $t > 0$

$$\mathbb{E}[\psi(tZ)] \geq \mathbb{E}[\mathbb{I}_{[0,+\infty]}(tZ)] = \mathbb{P}[tZ \geq 0] = \mathbb{P}[Z \geq 0] \geq \mathbb{P}[Z > 0]$$

Consider

$$S = \{x : \mathbb{P}\{f(x,\xi) \leq 0\} \geq 1 - \epsilon\} = \{x : \mathbb{P}\{f(x,\xi) > 0\} \leq \epsilon\}$$

Therefore, if we can guarantee:

$$\mathbb{E}[\psi(tf(x,\xi))] \leq \epsilon$$

then chance constraint is satisfied.

Tractable approximation (Nemirovskii and Shapiro, 2006)

If f is convex on x, leads to conservative convex constraint approximation of chance-constraint:

$$\inf_{t>0}\left[t\mathbb{E}[\psi(\frac{1}{t}f(x,\xi))] - t\epsilon\right] \leq 0$$

Nemirovskii and Shapiro, 2006 study choices of ψ

Tractable approximation (Nemirovskii and Shapiro, 2006)

If f is convex on x, leads to conservative convex constraint approximation of chance-constraint:

$$\inf_{t>0}\left[t\mathbb{E}[\psi(\frac{1}{t}f(x,\xi))] - t\epsilon\right] \leq 0$$

Nemirovskii and Shapiro, 2006 study choices of ψ

Ahmed et al., 2017 give a different way to approximate chance-constraints.
Xie and Ahmed, 2020 give a bi-criteria approximation for covering programs

Scenario approximation (Calafiore and Campi, 2005)

Take samples $\xi^1, \ldots, \xi^N$ and solve:

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & x \in X \\ & f(x, \xi^i) \leq 0, \forall i = 1, \ldots, N \end{array}$$

Theorem

Assume X convex, $f(x, \xi)$ convex on x for every ξ. Let $\delta > 0$.
If $N \geq \frac{2}{\epsilon} \log\left(\frac{1}{\delta}\right) + 2n + \frac{2n}{\epsilon} \log(\frac{2}{\epsilon})$ then, with confidence $1 - \delta$, the optimal solution x^* to the scenario approximation problem will satisfy chance-constraint.

Sample average approximation (Luedtke and Ahmed, 2008)

Take samples $\xi^1, \dots, \xi^N$ and solve the chance-constrained problem with confidence α over finite distribution.

$$z^*_{N,\alpha} := \begin{array}{ll} \min & c^T x \\ \text{s.t.} & x \in X \\ & \sum_{i=1}^{N} \frac{1}{N} \mathbb{I}(f(x, \xi^i) \le 0) \ge 1 - \alpha \end{array} \quad \text{(SAA-CC)}$$

Assume that (CC) has an optimal solution.

Sample average approximation (Luedtke and Ahmed, 2008)

Take samples $\xi^1, \ldots, \xi^N$ and solve the chance-constrained problem with confidence α over finite distribution.

$$z^*_{N,\alpha} := \begin{array}{ll} \min & c^T x \\ \text{s.t.} & x \in X \\ & \sum_{i=1}^{N} \frac{1}{N} \mathbb{I}(f(x, \xi^i) \leq 0) \geq 1 - \alpha \end{array} \qquad \text{(SAA-CC)}$$

Assume that (CC) has an optimal solution.

Theorem

If $\alpha > \epsilon$ then

$$\mathbb{P}[z^*_{N,\alpha} \leq z^*_\epsilon] \geq 1 - exp\{-2N(\alpha - \epsilon)^2\}$$

Sample average approximation (Luedtke and Ahmed, 2008)

Take samples $\xi^1, \ldots, \xi^N$ and solve the chance-constrained problem with confidence α over finite distribution.

$$z^*_{N,\alpha} := \begin{array}{ll} \min & c^T x \\ \text{s.t.} & x \in X \\ & \sum_{i=1}^{N} \frac{1}{N} \mathbb{I}(f(x, \xi^i) \leq 0) \geq 1 - \alpha \end{array} \qquad \text{(SAA-CC)}$$

Assume that (CC) has an optimal solution.

Theorem

If $\alpha > \epsilon$ then

$$\mathbb{P}[z^*_{N,\alpha} \leq z^*_\epsilon] \geq 1 - exp\{-2N(\alpha - \epsilon)^2\}$$

Theorem

Suppose X is finite and $\alpha < \epsilon$. Then, all feasible solutions to the sample problem at level α are feasible to the nominal problem at level ϵ with probabaility at least

$$1 - |X| exp\{-2N(\epsilon - \alpha)^2\}$$

Outline

Distributional robustness

Main idea:

- Often times, exact probability distribution is not really known.
- Want to have some guarantees of feasibility/optimality for ALL probability distributions that share certain characteristics

Distributional robustness

Main idea:

- Often times, exact probability distribution is not really known.
- Want to have some guarantees of feasibility/optimality for ALL probability distributions that share certain characteristics

Example: We want to solve

$$\min\{\max_{\mathbb{P}\in B_W(\mathbb{Q})} \mathbb{E}^{\mathbb{P}}[Q(x,\xi)] : x \in X\}$$

where $\mathbb{Q}$ is a given probability distribution, and $B_W(\mathbb{Q})$ is a set of probability distributions that are "like" $\mathbb{Q}$.

Distributional robustness (Esfahani and Kuhn, 2018)

Distance between two probability distributions:

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf \left\{ \mathbb{E}^{\mathbb{Q}} ||\xi^1 - \xi^2|| : \begin{array}{l} \mathbb{Q} \text{ is a joint distribution of } \xi^1 \text{ and } \xi^2 \\ \text{with marginals } \mathbb{Q}_1 \text{ and } \mathbb{Q}_2 \end{array} \right\}$$

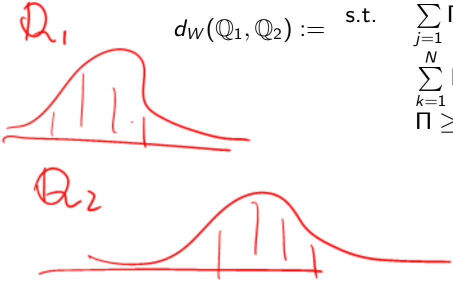
Distributional robustness (Esfahani and Kuhn, 2018)

Distance between two probability distributions:

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf \left\{ \mathbb{E}^{\mathbb{Q}} ||\xi^1 - \xi^2|| : \begin{array}{l} \mathbb{Q} \text{ is a joint distribution of } \xi^1 \text{ and } \xi^2 \\ \text{with marginals } \mathbb{Q}_1 \text{ and } \mathbb{Q}_2 \end{array} \right\}$$

If $\mathbb{Q}_i$ $(i = 1, 2)$ are finite discrete distributions with N possible realizations $\xi^i_1, \ldots, \xi^i_k$ and probabilities $q^i_1, \ldots, q^i_N$, then

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \begin{array}{ll} \min & \sum_{j=1}^{N} \sum_{k=1}^{N} ||\xi^1_j - \xi^2_k|| \Pi_{jk} \\ \text{s.t.} & \sum_{j=1}^{N} \Pi_{jk} = q^2_k, \forall k = 1, \ldots, N \\ & \sum_{k=1}^{N} \Pi_{jk} = q^1_j, \forall j = 1, \ldots, N \\ & \Pi \geq 0 \end{array}$$



Distributional robustness (Esfahani and Kuhn, 2018)

Distance between two probability distributions:

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf \left\{ \mathbb{E}^{\mathbb{Q}} ||\xi^1 - \xi^2|| : \begin{array}{c} \mathbb{Q} \text{ is a joint distribution of } \xi^1 \text{ and } \xi^2 \\ \text{with marginals } \mathbb{Q}_1 \text{ and } \mathbb{Q}_2 \end{array} \right\}$$

If $\mathbb{Q}_i$ $(i = 1, 2)$ are finite discrete distributions with N possible realizations $\xi_1^i, \ldots, \xi_k^i$ and probabilities $q_1^i, \ldots, q_N^i$, then

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \begin{array}{ll} \min & \sum_{j=1}^N \sum_{k=1}^N ||\xi_j^1 - \xi_k^2|| \Pi_{jk} \\ \text{s.t.} & \sum_{j=1}^N \Pi_{jk} = q_k^2, \forall k = 1, \ldots, N \\ & \sum_{k=1}^N \Pi_{jk} = q_j^1, \forall j = 1, \ldots, N \\ & \Pi \geq 0 \end{array}$$

We can then define: $B_W(\mathbb{Q}) := \{\mathbb{P} : d_W(\mathbb{P}, \mathbb{Q}) \leq \epsilon\}$
and rewrite our problem as a tractable optimization problem.

Distributional robustness (Esfahani and Kuhn, 2018)

Distance between two probability distributions:

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf \left\{ \mathbb{E}^{\mathbb{Q}} ||\xi^1 - \xi^2|| : \begin{array}{l} \mathbb{Q} \text{ is a joint distribution of } \xi^1 \text{ and } \xi^2 \\ \text{with marginals } \mathbb{Q}_1 \text{ and } \mathbb{Q}_2 \end{array} \right\}$$

If $\mathbb{Q}_i$ $(i = 1, 2)$ are finite discrete distributions with N possible realizations $\xi^i_1, \ldots, \xi^i_k$ and probabilities $q^i_1, \ldots, q^i_N$, then

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \begin{array}{ll} \min & \sum_{j=1}^{N} \sum_{k=1}^{N} ||\xi^1_j - \xi^2_k|| \Pi_{jk} \\ \text{s.t.} & \sum_{j=1}^{N} \Pi_{jk} = q^2_k, \forall k = 1, \ldots, N \\ & \sum_{k=1}^{N} \Pi_{jk} = q^1_j, \forall j = 1, \ldots, N \\ & \Pi \geq 0 \end{array}$$

We can then define: $B_W(\mathbb{Q}) := \{\mathbb{P} : d_W(\mathbb{P}, \mathbb{Q}) \leq \epsilon\}$
and rewrite our problem as a tractable optimization problem.

Chen, Kuhn, and Wiesemann, 2018 deals with distributionally robust chance-constraints
Rahimian and Mehrotra, 2019 provides a review of DRO.

Outline

Conclusion

Takeaways:

- Tons of possible ways to interpret an optimization problem with uncertainty
- Theoretical and computational challenges are plentiful
- Mix of statistics, optimization, engineering, economics

Conclusion

Takeaways:

- Tons of possible ways to interpret an optimization problem with uncertainty
- Theoretical and computational challenges are plentiful
- Mix of statistics, optimization, engineering, economics

Missing:

- Tons of details
- Lots more references
- Specific ideas to particular combinatorial optimization problems
- Combinatorial/Approximation algorithms

Papers suggested

- Ahmed and Shapiro, 2002: SAA two-stage with integer recourse
- Song and Luedtke, 2015: Aggregate scenarios for two-stage
- Bertsimas and Li, 2022: Stochastic cutting planes
- Zou, Ahmed, and Sun, 2019: Multistage stochastic programs with integer variables
- Luedtke and Ahmed, 2008: SAA chance-constraint
- Ahmed et al., 2017, Xie and Ahmed, 2020 and Ahmed and Xie, 2018: Approximations of chance-constraint
- Chen, Kuhn, and Wiesemann, 2018, Esfahani and Kuhn, 2018: DRO
- Barrera et al., 2016: SAA chance-constraint and sampling rare events
- Jiang and Guan, 2018: SAA with DRO and risk-averse

References I

Ahmed, Shabbir and Alexander Shapiro (2002). "The Sample Average Approximation Method for Stochastic Programs with Integer Recourse". In: *SIAM Journal of Optimization* 12, pp. 479–502.

Ahmed, Shabbir and Weijun Xie (2018). "Relaxations and approximations of chance constraints under finite distributions". In: *Math. Program.* 170, pp. 43–65.

Ahmed, Shabbir et al. (2017). "Nonanticipative duality, relaxations, and formulations for chance-constrained stochastic programs". In: *Mathematical Programming* 162.1–2, pp. 51–81.

Barrera, Javiera et al. (2016). "Chance-constrained problems and rare events: an importance sampling approach". In: *Mathematical Programming* 157, pp. 153–189.

Bertsimas, Dimitris and Michael Lingzhi Li (2022). "Stochastic Cutting Planes for Data-Driven Optimization". In: *INFORMS Journal on Computing* . Ahead of print.

Calafiore, G. C. and M. C. Campi (2005). "Uncertain convex programs: Randomized solutions andconfidence levels". In: *Math. Program.* 102, 25—46.

Chen, Zhi, Daniel Kuhn, and Wolfram Wiesemann (2018). *Data-Driven Chance Constrained Programs over Wasserstein Balls*. Tech. rep. https://arxiv.org/pdf/1809.00210.pdf.

References II

Esfahani, Peyman Mohajerin and Daniel Kuhn (2018). "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations". In: *Mathematical Programming* 171, pp. 115–166.

Hanasusanto, G. A., D. Kuhn, and W. Wiesemann (2016). "A comment on Computational complexity of stochastic programming problems". In: *Mathematical Programming*, 557—569.

Jiang, Ruiwei and Yongpei Guan (2018). "Risk-Averse Two-Stage Stochastic Program with Distributional Ambiguity". In: *Operations Research* 66.5, pp. 1390–1405.

Kleywegt, Anton J., Alexander Shapiro, and Tito Homem-de Mello (2002). "The Sample Average Approximation Method for Stochastic Discrete Optimization". In: *SIAM Journal on Optimization* 12.2, pp. 479–502.

Luedtke, James and Shabbir Ahmed (2008). "A Sample Approximation Approach for Optimization with Probabilistic Constraints". In: *SIAM Journal on Optimization* 19.2, pp. 674–699.

Matos, Vitor L. de, Andy B. Philpott, and Erlon C. Finardi (2015). "Improving the performance of Stochastic Dual Dynamic Programming". In: *Journal of Computational and Applied Mathematics* 290, pp. 196–208.

References III

Nemirovskii, Arkadi and Alexander Shapiro (2006). “Convex approximations of chance constrained programs”. In: *SIAM Journal on optimization* 17.4, pp. 969–996.

Pereira, M.V.F. and L.M.V.G. Pinto (1991). “Multi-stage stochastic optimization applied to energy planning”. In: *Mathematical Programming* 52, pp. 359–375.

Rahimian, Hamed and Sanjay Mehrotra (2019). *Distributionally robust optimization: A review*. Tech. rep. https://arxiv.org/pdf/1908.05659.pdf.

Rockafellar, R. and S. Uryasev (2002). “Conditional value-at-risk for general loss distributions”. In: *Journal of Banking & Finance* 26.7, pp. 1443–1471.

Ruszczyński, Andrzej and Alexander Shapiro (2006). “Optimization of Convex Risk Functions”. In: *Mathematics of Operations Research* 31.3, pp. 433–452.

Song, Yongjia and James Luedtke (2015). “An Adaptive Partition-Based Approach for Solving Two-Stage Stochastic Programs with Fixed Recourse”. In: *SIAM Journal on Optimization* 25.3, pp. 1344–1367.

Xie, Weijun and Shabbir Ahmed (2020). “Bicriteria Approximation of Chance-Constrained Covering Problems”. In: *Operations Research* 68.2, pp. 516–533.

Zou, Jikai, Shabbir Ahmed, and Xu Andy Sun (2019). “Stochastic dual dynamic integer programming”. In: *Mathematical Programming* 175, pp. 461–502.