Admissibility of solution estimators to stochastic optimization problems

Amitabh Basu

Joint Work with Tu Nguyen and Ao Sun

23rd Combinatorial Optimization Workshop,
Aussois, France, January 2019

A general stochastic optimization problem

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$

$X \subseteq \mathbb{R}^d$

ξ is a random variable taking values in $\mathbb{R}^m$

$$\min_{x \in X} \ \mathbb{E}_\xi[\ F(x, \xi)\ ]$$

A general stochastic optimization problem

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$

$X \subseteq \mathbb{R}^d$

ξ is a random variable taking values in $\mathbb{R}^m$

$$\min_{x \in X} \ \mathbb{E}_\xi[\ F(x, \xi)\ ]$$

Example:

Supervised Machine Learning: One sees samples $(z, y) \in \mathbb{R}^n \times \mathbb{R}$ of labeled data from some distribution, and one aims to find a function $f \in \mathcal{F}$ in a finitely parameterized hypothesis class $f \in \mathcal{F}$ that minimizes the expected loss $\mathbb{E}_{(z,y)}[\ell(f(z), y)]$, where $\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ is some loss function. Then d is the number of parameters, X is a description of $\mathcal{F}$ via the parameters, $m = n + 1$, $\xi = (z, y)$, and

$$F(f, (z, y)) = \ell(f(z), y).$$

A general stochastic optimization problem

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$

$X \subseteq \mathbb{R}^d$

ξ is a random variable taking values in $\mathbb{R}^m$

$$\min_{x \in X} \ \mathbb{E}_\xi[\, F(x, \xi) \,]$$

Example:

(News) Vendor Problem: (News) Vendor buys some units of a product (newspapers) from supplier at cost of $c > 0$ dollars/unit; at most u units available. Stochastic demand for product. Product sold at price $p > c$ dollars/unit. End of day, vendor can return unsold product to supplier at $r < c$ dollars/unit. Find number of units to buy to maximize (minimize) the expected profit (loss).

A general stochastic optimization problem

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$
$X \subseteq \mathbb{R}^d$
ξ is a random variable taking values in $\mathbb{R}^m$

$$\min_{x \in X} \ \mathbb{E}_\xi[\ F(x,\xi)\ ]$$

Example:

(News) Vendor Problem: (News) Vendor buys some units of a product (newspapers) from supplier at cost of $c > 0$ dollars/unit; at most u units available. Stochastic demand for product. Product sold at price $p > c$ dollars/unit. End of day, vendor can return unsold product to supplier at $r < c$ dollars/unit. Find number of units to buy to maximize (minimize) the expected profit (loss).
$d = m = 1$, $X = [0, u]$,

$$F(x,\xi) = cx - p \min\{x,\xi\} - r \max\{x - \xi, 0\}.$$

Solving the problem

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$

$X \subseteq \mathbb{R}^d$

ξ is a random variable taking values in $\mathbb{R}^m$

$$\min_{x \in X} \; \mathbb{E}_\xi[\, F(x, \xi) \,]$$

Solving the problem

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$

$X \subseteq \mathbb{R}^d$

ξ is a random variable taking values in $\mathbb{R}^m$

$$\min_{x \in X} \; \mathbb{E}_\xi[\, F(x, \xi) \,]$$

Solve the problem only given access to n i.i.d. samples of ξ.

Solving the problem

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$

$X \subseteq \mathbb{R}^d$

ξ is a random variable taking values in $\mathbb{R}^m$

$$\min_{x \in X} \; \mathbb{E}_\xi[\, F(x, \xi) \,]$$

Solve the problem only given access to n i.i.d. samples of ξ.

Natural idea: Given samples $\xi^1, \ldots, \xi^n \in \mathbb{R}^d$, solve the deterministic problem

$$\min_{x \in X} \; \frac{1}{n} \sum_{i=1}^{n} F(x, \xi^i)$$

Solving the problem

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$

$X \subseteq \mathbb{R}^d$

ξ is a random variable taking values in $\mathbb{R}^m$

$$\min_{x \in X} \; \mathbb{E}_\xi[\, F(x, \xi) \,]$$

Solve the problem only given access to n i.i.d. samples of ξ.

Natural idea: Given samples $\xi^1, \ldots, \xi^n \in \mathbb{R}^d$, solve the deterministic problem

$$\min_{x \in X} \; \frac{1}{n} \sum_{i=1}^{n} F(x, \xi^i)$$

Stochastic optimizers call this sample average approximation (SAA); machine learners call this empirical risk minimization.

Concrete Problem

$d = m.$ $\quad F(x, \xi) = \xi^T x$

$X \subseteq \mathbb{R}^d$ is a compact set (e.g., polytope, integer points in a polytope).

$\xi \sim N(\mu, \sigma I)$.

$$\min_{x \in X} \ \mathbb{E}_\xi[\, F(x, \xi) \,] = \min_{x \in X} \ \mathbb{E}_\xi[\, \xi^T x \,] = \min_{x \in X} \ \mu^T x$$

Concrete Problem

$d = m. \qquad F(x, \xi) = \xi^T x$

$X \subseteq \mathbb{R}^d$ is a compact set (e.g., polytope, integer points in a polytope).

$\xi \sim N(\mu, \sigma I)$.

$$\min_{x \in X} \ \mathbb{E}_\xi[\, F(x, \xi)\,] \ = \ \min_{x \in X} \ \mathbb{E}_\xi[\, \xi^T x\,] \ = \ \min_{x \in X} \ \mu^T x$$

Solve the problem only given access to n i.i.d. samples of ξ.
Important: μ is unknown.

Concrete Problem

$d = m. \qquad F(x,\xi) = \xi^T x$

$X \subseteq \mathbb{R}^d$ is a compact set (e.g., polytope, integer points in a polytope).

$\xi \sim N(\mu, \sigma I)$.

$$\min_{x \in X} \; \mathbb{E}_\xi[\, F(x,\xi)\,] \;=\; \min_{x \in X} \; \mathbb{E}_\xi[\, \xi^T x\,] \;=\; \min_{x \in X} \; \mu^T x$$

Solve the problem only given access to n i.i.d. samples of ξ.
Important: μ is unknown.

Sample Average Approximation (SAA):

$$\min_{x \in X} \; \frac{1}{n} \sum_{i=1}^{n} F(x, \xi^i) \;=\; \min_{x \in X} \; \overline{\xi}^T x$$

where $\overline{\xi} := \frac{1}{n} \sum_{i=1}^{n} \xi^i$.

A quick tour of Statistical Decision Theory

Set of states of nature, modeled by a set Θ.

Set of possible actions to take, modeled by $\mathcal{A}$.

In a particular state of nature $\theta \in \Theta$, the performance of any action $a \in \mathcal{A}$, is evaluated by a loss function $\mathcal{L}(\theta, a)$. Goal: choose action to minimize loss.

(Partial/Incomplete) Information about θ is obtained through a random variable y taking values in a sample space χ. The distribution of y depends on the particular state of nature θ, denoted by P_θ.

A quick tour of Statistical Decision Theory

Set of states of nature, modeled by a set Θ.

Set of possible actions to take, modeled by $\mathcal{A}$.

In a particular state of nature $\theta \in \Theta$, the performance of any action $a \in \mathcal{A}$, is evaluated by a loss function $\mathcal{L}(\theta, a)$. Goal: choose action to minimize loss.

(Partial/Incomplete) Information about θ is obtained through a random variable y taking values in a sample space χ. The distribution of y depends on the particular state of nature θ, denoted by P_θ.

Decision Rule: Takes $y \in \chi$ as input and reports an action $a \in \mathcal{A}$. Denote by $\delta : \chi \to \mathcal{A}$.

Our problem cast as statistical decision problem

$X \subseteq \mathbb{R}^d$ is a compact set. $\xi \sim N(\mu, I)$.

$$\min_{x \in X} \ \mathbb{E}_\xi[\, F(x, \xi)\,] \ = \ \min_{x \in X} \ \mathbb{E}_\xi[\, \xi^T x \,] \ = \ \min_{x \in X} \ \mu^T x$$

States of Nature: $\Theta = \mathbb{R}^d = \{\text{all possible } \mu \in \mathbb{R}^d\}$.

Set of Actions: $X \subseteq \mathbb{R}^d$.

Our problem cast as statistical decision problem

$X \subseteq \mathbb{R}^d$ is a compact set. $\xi \sim N(\mu, I)$.

$$\min_{x \in X} \; \mathbb{E}_\xi[\, F(x,\xi) \,] \;=\; \min_{x \in X} \; \mathbb{E}_\xi[\, \xi^T x \,] \;=\; \min_{x \in X} \; \mu^T x$$

States of Nature: $\Theta = \mathbb{R}^d = \{\text{all possible } \mu \in \mathbb{R}^d\}$.

Set of Actions: $X \subseteq \mathbb{R}^d$.

Loss function: ?

Our problem cast as statistical decision problem

$X \subseteq \mathbb{R}^d$ is a compact set. $\xi \sim N(\mu, I)$.

$$\min_{x \in X} \mathbb{E}_{\xi}[\, F(x, \xi) \,] = \min_{x \in X} \mathbb{E}_{\xi}[\, \xi^T x \,] = \min_{x \in X} \mu^T x$$

States of Nature: $\Theta = \mathbb{R}^d = \{\text{all possible } \mu \in \mathbb{R}^d\}$.

Set of Actions: $X \subseteq \mathbb{R}^d$.

Loss function:

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(\bar{x}, \xi) \,] - \min_{x \in X} \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(x, \xi) \,]$$

Our problem cast as statistical decision problem

$X \subseteq \mathbb{R}^d$ is a compact set. $\xi \sim N(\mu, I)$.

$$\min_{x \in X} \mathbb{E}_\xi[\, F(x, \xi)\,] = \min_{x \in X} \mathbb{E}_\xi[\, \xi^T x\,] = \min_{x \in X} \mu^T x$$

States of Nature: $\Theta = \mathbb{R}^d = \{$all possible $\mu \in \mathbb{R}^d\}$.

Set of Actions: $X \subseteq \mathbb{R}^d$.

Loss function:

$$\begin{aligned} \mathcal{L}(\bar{\mu}, \bar{x}) &= \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(\bar{x}, \xi)\,] - \min_{x \in X} \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(x, \xi)\,] \\ &= \bar{\mu}^T \bar{x} - \bar{\mu}^T x(\bar{\mu}) \end{aligned}$$

Our problem cast as statistical decision problem

$X \subseteq \mathbb{R}^d$ is a compact set. $\xi \sim N(\mu, I)$.

$$\min_{x \in X} \; \mathbb{E}_{\xi}[\, F(x, \xi) \,] \;=\; \min_{x \in X} \; \mathbb{E}_{\xi}[\, \xi^T x \,] \;=\; \min_{x \in X} \; \mu^T x$$

States of Nature: $\Theta = \mathbb{R}^d = \{\text{all possible } \mu \in \mathbb{R}^d\}$.

Set of Actions: $X \subseteq \mathbb{R}^d$.

Loss function:

$$\begin{aligned} \mathcal{L}(\bar{\mu}, \bar{x}) &= \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(\bar{x}, \xi) \,] - \min_{x \in X} \; \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(x, \xi) \,] \\ &= \bar{\mu}^T \bar{x} \;-\; \bar{\mu}^T x(\bar{\mu}) \end{aligned}$$

Sample Space: $\chi = \underbrace{\mathbb{R}^d \times \mathbb{R}^d \times \ldots \times \mathbb{R}^d}_{n \text{ times}}$

Our problem cast as statistical decision problem

$X \subseteq \mathbb{R}^d$ is a compact set. $\xi \sim N(\mu, I)$.

$$\min_{x \in X} \; \mathbb{E}_\xi[\, F(x,\xi)\,] \;=\; \min_{x \in X} \; \mathbb{E}_\xi[\, \xi^T x \,] \;=\; \min_{x \in X} \; \mu^T x$$

States of Nature: $\Theta = \mathbb{R}^d = \{\text{all possible } \mu \in \mathbb{R}^d\}$.

Set of Actions: $X \subseteq \mathbb{R}^d$.

Loss function:

$$\begin{aligned} \mathcal{L}(\bar{\mu}, \bar{x}) &= \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(\bar{x}, \xi)\,] - \min_{x \in X} \; \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(x, \xi)\,] \\ &= \bar{\mu}^T \bar{x} \;-\; \bar{\mu}^T x(\bar{\mu}) \end{aligned}$$

Sample Space: $\chi = \underbrace{\mathbb{R}^d \times \mathbb{R}^d \times \ldots \times \mathbb{R}^d}_{n \text{ times}}$

Decision Rule: $\delta : \chi \to X$.

Our problem cast as statistical decision problem

$X \subseteq \mathbb{R}^d$ is a compact set. $\xi \sim N(\mu, I)$.

$$\min_{x \in X} \ \mathbb{E}_\xi[\, F(x,\xi)\,] \ = \ \min_{x \in X} \ \mathbb{E}_\xi[\, \xi^T x\,] \ = \ \min_{x \in X} \ \mu^T x$$

States of Nature: $\Theta = \mathbb{R}^d = \{\text{all possible } \mu \in \mathbb{R}^d\}$.

Set of Actions: $X \subseteq \mathbb{R}^d$.

Loss function:

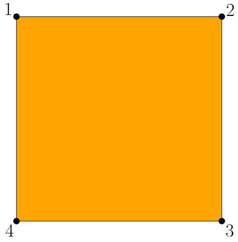
$$\begin{aligned} \mathcal{L}(\bar{\mu}, \bar{x}) &= \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(\bar{x}, \xi)\,] - \min_{x \in X} \ \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(x,\xi)\,] \\ &= \bar{\mu}^T \bar{x} \ - \ \bar{\mu}^T x(\bar{\mu}) \end{aligned}$$

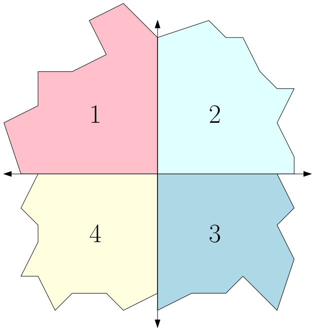
Sample Space: $\chi = \underbrace{\mathbb{R}^d \times \mathbb{R}^d \times \ldots \times \mathbb{R}^d}_{n \text{ times}}$

Decision Rule: $\delta : \chi \to X$.

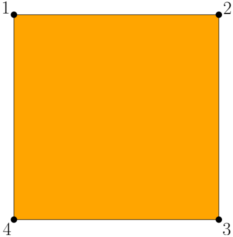
$$\text{SAA:} \quad \delta(\xi^1, \ldots, \xi^n) \in \arg\max\{\bar{\xi}^T x : x \in X\}$$

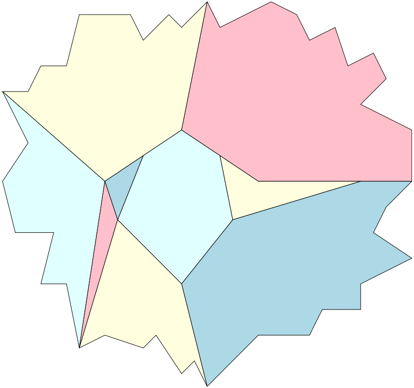
Decision Rules



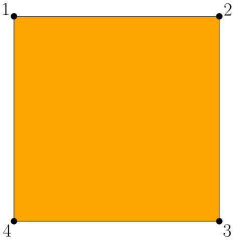


Decision Rules





Decision Rules



How does one decide between decision rules?

How does one decide between decision rules?

States of nature Θ, Actions $\mathcal{A}$, Loss function $\mathcal{L} : \Theta \times \mathcal{A} \to \mathbb{R}$,
Sample space χ with distributions $\{P_\theta : \theta \in \Theta\}$.

How does one decide between decision rules?

States of nature Θ, Actions $\mathcal{A}$, Loss function $\mathcal{L} : \Theta \times \mathcal{A} \to \mathbb{R}$, Sample space χ with distributions $\{P_\theta : \theta \in \Theta\}$.

Given a decision rule $\delta : \chi \to \mathcal{A}$, define the risk function of this decision rule as:

$$R_\delta(\theta) := \mathbb{E}_{y \sim P_\theta}[\, \mathcal{L}(\theta, \delta(y))\,]$$

How does one decide between decision rules?

States of nature Θ, Actions $\mathcal{A}$, Loss function $\mathcal{L} : \Theta \times \mathcal{A} \to \mathbb{R}$, Sample space χ with distributions $\{P_\theta : \theta \in \Theta\}$.

Given a decision rule $\delta : \chi \to \mathcal{A}$, define the risk function of this decision rule as:

$$R_\delta(\theta) := \mathbb{E}_{y \sim P_\theta}[\, \mathcal{L}(\theta, \delta(y)) \,]$$

We say that a decision rule δ' dominates a decision rule δ if $R_{\delta'}(\theta) \leq R_\delta(\theta)$ for all $\theta \in \Theta$, and $R_{\delta'}(\theta^*) < R_\delta(\theta^*)$ for some $\theta^* \in \Theta$.

How does one decide between decision rules?

States of nature Θ, Actions $\mathcal{A}$, Loss function $\mathcal{L} : \Theta \times \mathcal{A} \to \mathbb{R}$, Sample space χ with distributions $\{P_\theta : \theta \in \Theta\}$.

Given a decision rule $\delta : \chi \to \mathcal{A}$, define the risk function of this decision rule as:

$$R_\delta(\theta) := \mathbb{E}_{y \sim P_\theta}[\, \mathcal{L}(\theta, \delta(y))\,]$$

We say that a decision rule δ' dominates a decision rule δ if $R_{\delta'}(\theta) \leq R_\delta(\theta)$ for all $\theta \in \Theta$, and $R_{\delta'}(\theta^*) < R_\delta(\theta^*)$ for some $\theta^* \in \Theta$.

If a decision rule δ is not dominated by any other decision rule, we say that δ is admissible. Otherwise, it is inadmissible.

Is the Sample Average Approximation (SAA) rule admissible?

Admissibility in stochastic optimization

Stochastic optimization setup:

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$, $X \subseteq \mathbb{R}^d$, ξ is a R.V. in $\mathbb{R}^m$

$$\min_{x \in X} \; \mathbb{E}_{\xi}[\, F(x, \xi) \,]$$

Want to solve with access to n i.i.d. samples of ξ.

Admissibility in stochastic optimization

Stochastic optimization setup:

$F : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$, $X \subseteq \mathbb{R}^d$, ξ is a R.V. in $\mathbb{R}^m$

$$\min_{x \in X} \; \mathbb{E}_\xi[\, F(x, \xi) \,]$$

Want to solve with access to n i.i.d. samples of ξ.

Statistical decision theory view:

$\xi \sim N(\mu, I)$; states of nature $\Theta = \mathbb{R}^m = \{\text{all possible } \mu \in \mathbb{R}^m\}$.

Set of actions $\mathcal{A} = X$, Sample space $\chi = \underbrace{\mathbb{R}^d \times \mathbb{R}^d \times \ldots \times \mathbb{R}^d}_{n \text{ times}}$

Loss function

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(\bar{x}, \xi) \,] - \min_{x \in X} \; \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(x, \xi) \,]$$

Given a decision rule $\delta : \chi \to X$, the risk of δ

$$R_\delta(\mu) := \mathbb{E}_{\xi^1, \ldots, \xi^n}[\, \mathcal{L}(\mu, \delta(\xi^1, \ldots, \xi^n)) \,]$$

Admissibility in stochastic optimization

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(\bar{x}, \xi)\,] - \min_{x \in X} \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\, F(x, \xi)\,]$$

Given a decision rule $\delta : \chi \to X$, the risk of δ

$$R_\delta(\mu) := \mathbb{E}_{\xi^1, \ldots, \xi^n}[\, \mathcal{L}(\mu, \delta(\xi^1, \ldots, \xi^n))\,]$$

Sample Average Approximation (SAA):

$$\min_{x \in X} \; \frac{1}{n} \sum_{i=1}^{n} F(x, \xi^i)$$

Admissibility in stochastic optimization

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(\bar{x}, \xi)\ ] - \min_{x \in X}\ \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(x, \xi)\ ]$$

Given a decision rule $\delta : \chi \to X$, the risk of δ

$$R_\delta(\mu) := \mathbb{E}_{\xi^1, \ldots, \xi^n}[\ \mathcal{L}(\mu, \delta(\xi^1, \ldots, \xi^n))\ ]$$

Sample Average Approximation (SAA):

$$\min_{x \in X}\ \frac{1}{n} \sum_{i=1}^{n} F(x, \xi^i)$$

Sample Average Approximation (SAA) can be inadmissible!!

Inadmissibility of SAA: Stein's Paradox

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(\bar{x}, \xi)\ ] - \min_{x \in X}\ \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(x, \xi)\ ]$$

Sample Average Approximation (SAA) can be inadmissible!!

$$d = m, \qquad F(x, \xi) = \|x - \xi\|^2, \qquad X = \mathbb{R}^d, \qquad \xi \sim N(\mu, I).$$

$$\min_{x \in \mathbb{R}^d}\ \mathbb{E}_{\xi}[\ F(x, \xi)\ ]\ =\ \min_{x \in \mathbb{R}^d}\ \mathbb{E}_{\xi}[\ \|x - \xi\|^2\ ]$$

Inadmissibility of SAA: Stein's Paradox

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(\bar{x}, \xi)\ ] - \min_{x \in X}\ \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(x, \xi)\ ]$$

Sample Average Approximation (SAA) can be inadmissible!!

$$\min_{x \in \mathbb{R}^d}\ \ \mathbb{E}_{\xi \sim N(\mu, I)}[\ \|x - \xi\|^2\ ]$$

Optimal solution: $x(\bar{\mu}) = \bar{\mu}$, Optimal value: d. Simple calculation:

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \|\bar{x} - \bar{\mu}\|^2.$$

Inadmissibility of SAA: Stein's Paradox

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(\bar{x}, \xi)\ ] - \min_{x \in X}\ \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(x, \xi)\ ]$$

Sample Average Approximation (SAA) can be inadmissible!!

$$\min_{x \in \mathbb{R}^d}\ \ \mathbb{E}_{\xi \sim N(\mu, I)}[\ \|x - \xi\|^2\ ]$$

Optimal solution: $x(\bar{\mu}) = \bar{\mu}$, Optimal value: d. Simple calculation:

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \|\bar{x} - \bar{\mu}\|^2.$$

Sample Average Approximation (SAA):

$$\min_{x \in \mathbb{R}^d}\ \ \frac{1}{n} \sum_{i=1}^{n} \|x - \xi^i\|^2$$

Inadmissibility of SAA: Stein's Paradox

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(\bar{x}, \xi)\ ] - \min_{x \in X}\ \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(x, \xi)\ ]$$

Sample Average Approximation (SAA) can be inadmissible!!

$$\min_{x \in \mathbb{R}^d}\ \ \mathbb{E}_{\xi \sim N(\mu, I)}[\ \|x - \xi\|^2\ ]$$

Optimal solution: $x(\bar{\mu}) = \bar{\mu}$, Optimal value: d. Simple calculation:

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \|\bar{x} - \bar{\mu}\|^2.$$

Sample Average Approximation (SAA):

$$\min_{x \in \mathbb{R}^d}\ \ \frac{1}{n} \sum_{i=1}^{n} \|x - \xi^i\|^2$$

$$\delta_{SAA}(\xi^1, \ldots, \xi^n)\ \ =\ \ \bar{\xi}\ \ :=\ \ \sum_{i=1}^{n} \xi^i.$$

Inadmissibility of SAA: Stein's Paradox

$$\mathcal{L}(\bar{\mu}, \bar{x}) = \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(\bar{x}, \xi)\ ] - \min_{x \in X}\ \mathbb{E}_{\xi \sim N(\bar{\mu}, I)}[\ F(x, \xi)\ ]$$

Sample Average Approximation (SAA) can be inadmissible!!

$$\min_{x \in \mathbb{R}^d}\ \ \mathbb{E}_{\xi \sim N(\mu, I)}[\ \|x - \xi\|^2\ ]$$

Generalized to arbitrary convex quadratic function with uncertain linear term in Davarnia and Cornuéjols 2018. Follow-up work from a Bayesian perspective in Davarnia, Kocuk and Cornuéjols 2018.

A class of problems with no Stein's paradox

THEOREM Basu-Nguyen-Sun 2018

Consider the problem of optimizing an uncertain linear objective $\xi \sim N(\mu, I)$ over a fixed compact set $X \subseteq \mathbb{R}^d$:

$$\min_{x \in X} \; \mathbb{E}_{\xi \sim N(\mu, I)}[\, \xi^T x \,]$$

The Sample Average Approximation (SAA) rule is admissible.

Main technical ideas/tools

Main technical ideas/tools

Sufficient Statistic: $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ family of distributions for r.v. y in sample space χ. Sufficient statistic for this family is a function $T : \chi \to \tau$ such that the conditional probability $P(y|T = t)$ does not depend on θ.

Main technical ideas/tools

Sufficient Statistic: $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ family of distributions for r.v. y in sample space χ. Sufficient statistic for this family is a function $T : \chi \to \tau$ such that the conditional probability $P(y|T = t)$ does not depend on θ.

FACT:

$$\chi = \underbrace{\mathbb{R}^d \times \ldots \times \mathbb{R}^d}_{n \text{ times}}, \ \mathcal{P} = \{\underbrace{N(\mu, I) \times \ldots \times N(\mu, I)}_{n \text{ times}} : \mu \in \mathbb{R}^d\},$$

i.e., $(\xi^1, \ldots, \xi^n) \in \chi$ are i.i.d samples from the normal distribution $N(\mu, I)$. Then $T(\xi^1 \ldots, \xi^n) = \overline{\xi} := \frac{1}{n}\sum_{i=1}^n \xi^i$ is a sufficient statistic for $\mathcal{P}$.

Main technical ideas/tools

For any decision rule δ, define the function

$$F(\mu) = R_{\delta}(\mu) - R_{\delta_{SAA}}(\mu).$$

Suffices to show that there exists $\hat{\mu} \in \mathbb{R}^d$ such that $F(\hat{\mu}) > 0$.

Main technical ideas/tools

For any decision rule δ, define the function

$$F(\mu) = R_\delta(\mu) - R_{\delta_{SAA}}(\mu).$$

Suffices to show that there exists $\hat{\mu} \in \mathbb{R}^d$ such that $F(\hat{\mu}) > 0$.

First observe: $F(0) = 0$.

Main technical ideas/tools

For any decision rule δ, define the function

$$F(\mu) = R_{\delta}(\mu) - R_{\delta_{SAA}}(\mu).$$

Suffices to show that there exists $\hat{\mu} \in \mathbb{R}^d$ such that $F(\hat{\mu}) > 0$.

First observe: $F(0) = 0$.

Then compute $\nabla^2 F(0)$; show it has a strictly positive eigenvalue.

Main technical ideas/tools

For any decision rule δ, define the function

$$F(\mu) = R_{\delta}(\mu) - R_{\delta_{SAA}}(\mu).$$

Suffices to show that there exists $\hat{\mu} \in \mathbb{R}^d$ such that $F(\hat{\mu}) > 0$.

First observe: $F(0) = 0$.

Then compute $\nabla^2 F(0)$; show it has a strictly positive eigenvalue.

Use a fact from probability theory that for any Lebesgue integrable function $f : \mathbb{R}^n \to \mathbb{R}^d$, the map

$$\mu \mapsto \mathbb{E}_{y \in N(\mu,\Sigma)} \left[f(y) \right] := \int_{\mathbb{R}^d} f(y) \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu) \right) dy$$

has derivatives of all orders and these can be computed by taking the derivative under the integral sign.

Open Questions

- What about nonlinear objectives over compact feasible regions? For example, what if $F(x, \xi) = x^T Q x + \xi^T x$ for some fixed PSD matrix Q, and X is a compact (convex) set?

Open Questions

- What about nonlinear objectives over compact feasible regions? For example, what if $F(x, \xi) = x^T Qx + \xi^T x$ for some fixed PSD matrix Q, and X is a compact (convex) set?
- What about piecewise linear objectives $F(x, \xi)$? Recall News Vendor Problem.

Open Questions

- What about nonlinear objectives over compact feasible regions? For example, what if $F(x, \xi) = x^T Qx + \xi^T x$ for some fixed PSD matrix Q, and X is a compact (convex) set?
- What about piecewise linear objectives $F(x, \xi)$? Recall News Vendor Problem.
- Objectives coming from machine learning problems, such as neural network training with squared or logistic loss (admissibility of "empirical risk minimization"). Maybe this depends on the hypothesis class that is being learnt?

Open Questions

- What about nonlinear objectives over compact feasible regions? For example, what if $F(x, \xi) = x^T Qx + \xi^T x$ for some fixed PSD matrix Q, and X is a compact (convex) set?
- What about piecewise linear objectives $F(x, \xi)$? Recall News Vendor Problem.
- Objectives coming from machine learning problems, such as neural network training with squared or logistic loss (admissibility of "empirical risk minimization"). Maybe this depends on the hypothesis class that is being learnt?
- METATHEOREM (from Gérard Cornuéjols): Admissible if and only if feasible region is bounded !?

THANK YOU !

Questions/Comments ?