THE OHIO STATE UNIVERSITY

THE OHIO STATE UNIVERSITY

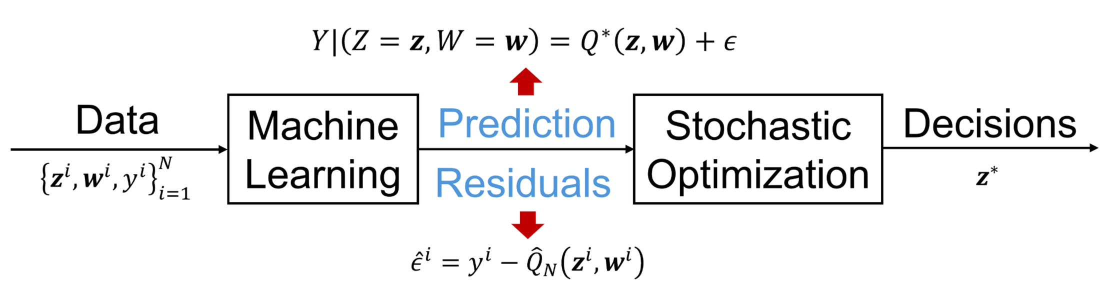
Contextual Stochastic Optimization with Decision-Dependent Uncertainty via Nonparametric Learning

Huangrong Sun (sun.3394@osu.edu), Xian Yu (yu.3610@osu.edu), Department of Integrated System Engineering

Motivation

- Decision-dependent contextual stochastic program:
  (DD-CSP) $v^*(\boldsymbol{w}) = \min_{z\in Z} \mathbb{E}_Y\left[c(\boldsymbol{z},Y) \mid Z=\boldsymbol{z}, W=\boldsymbol{w}\right]$.
- Example: newsvendor problem with pricing:
  $\min_{p,q\in Z} \mathbb{E}_Y\left[fq - pY + h(q-Y)^+ + b(Y-q)^+ \mid P=p, W=\boldsymbol{w}\right]$
- p: price, q: ordering quantity, w: covariate
- Uncertain demand Y depends on both p and $\boldsymbol{w}$:
  - Higher price → lower demand
  - Higher income → higher purchasing power

Framework



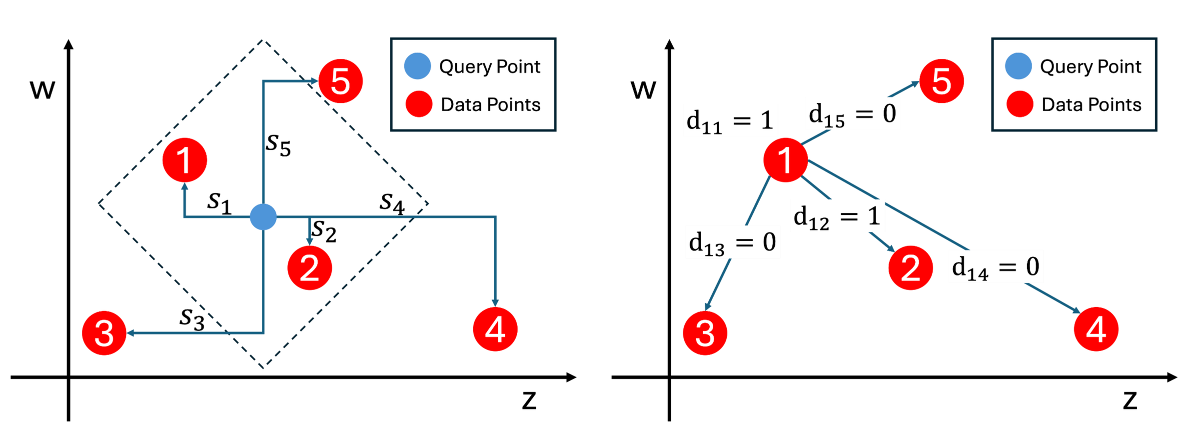
- **Step 1: Machine Learning**. Approximate the true $Q^*(z,w)$ with $\hat{Q}_N(z,w)$ via nonparametric regression.
- **Step 2: Stochastic Optimization**.
  (ER-DD-SAA) $\hat{v}_N(\boldsymbol{w}) = \min_{z\in Z} \frac{1}{N}\sum_{i=1}^N c(\boldsymbol{z}, \hat{Q}_N(\boldsymbol{z},\boldsymbol{w}) + \hat{\epsilon}^i)$
- **Challenge**: Trained nonparametric regression models need to be embedded into the downstream optimization problem to solve for $\boldsymbol{z}$.
- **Main Contributions**:
  - Exact MIP formulations for ER-DD-SAA with k-Nearest Neighbors (kNN), Classification and Regression Tree (CART), ReLU Neural Networks.
  - Tailored reformulations and algorithms for kNN.
  - Asymptotic optimality guarantees under uniform-in-decision assumption on the regression models.

ER-DD-SAA with kNN

- k-nearest neighbors (kNN):

$$\mathcal{N}_k(\boldsymbol{z},\boldsymbol{w}) = \left\{i: \sum_{j=1}^N \mathbb{I}\{\|(\boldsymbol{z},w) - (\boldsymbol{z}^i, w^i)\|_p \geq \|(\boldsymbol{z},w) - (z^j, w^j)\|_p\} \leq k\right\}$$

- Point-estimation: $\hat{Q}_N(\boldsymbol{z},\boldsymbol{w}) = \hat{y}_{kNN}(\boldsymbol{z},\boldsymbol{w}) = \frac{1}{k}\sum_{i\in\mathcal{N}_k(\boldsymbol{z},\boldsymbol{w})} y^i$.



Example: $\hat{y}_{kNN}(\boldsymbol{z},\boldsymbol{w}) = \frac{1}{k}\sum_i t_i y^i = (y^1 + y^2)/2$

- Point 1: $d_{12} = d_{11} = 1, d_{13} = d_{14} = d_{15} = 0 \Rightarrow t_1 = 1$
- Point 2: $d_{22} = 0, d_{21} = d_{23} = d_{24} = d_{25} = 0 \Rightarrow t_2 = 1$

ER-DD-SAA with kNN (Cont'd)

Pairwise Distance Comparison

Decision variables:

- $s_i = \|(\boldsymbol{z},\boldsymbol{w}) - (\boldsymbol{z}^i, \boldsymbol{w}^i)\|_1$: L_1 distance from data point i to query point $(\boldsymbol{z},\boldsymbol{w})$.
- $d_{ij} \in \{0,1\}$: if data i is farther than j, $d_{ij} = 1 \Leftrightarrow s_i \geq s_j$.
- $t_i \in \{0,1\}$: if data i is among the k-nearest neighbors, $t_i = 1 \Leftrightarrow \sum_j d_{ij} \leq k$.

$$\begin{aligned}
&\min_{z\in Z} \frac{1}{N}\sum_{i=1}^N c(\boldsymbol{z}, y_i) && (1a)\\
&s.t.\ M_1(d_{ij}-1) \leq s_i - s_j \leq M_1 d_{ij}, \forall i > j, && (1b)\\
&k - M_2 t_i \leq \sum_{i>j} d_{ij} + \sum_{i<j}(1-d_{ij}) + 1, \forall i \in [N], && (1c)\\
&\sum_{i>j} d_{ij} + \sum_{i<j}(1-d_{ij}) + 1 \leq k + M_2(1-t_i), \forall i \in [N], && (1d)\\
&\sum_{i=1}^N t_i = k, && (1e)\\
&t_i \in \{0,1\}^N, \forall i \in [N], && (1f)\\
&s_i = \|(\boldsymbol{z},\boldsymbol{w}) - (\boldsymbol{z}^i, \boldsymbol{w}^i)\|_1, \forall i \in [N], && (1g)\\
&y_i = \frac{1}{k}\sum_{j=1}^N t_j y^j + \hat{\epsilon}^i_{kNN}. && (1h)
\end{aligned}$$

Decomposition Algorithm

- We first solve a restricted master problem (RMP), i.e., problem (1) without (1b) − (1d), to obtain a candidate decision $\boldsymbol{z}$, a selected kNN set F, and a lower bound (LB).
- Given $\boldsymbol{z}$, identify the true kNN set S, fix $\boldsymbol{z}$ in problem (1) to obtain an upper bound (UB).
- If $F \not\subseteq S$ and $\left|\frac{UB-LB}{LB}\right| \leq$ tolerance, then $I \leftarrow I \cup S$, $J \leftarrow J \cup F$, and add the following constraints to the RMP:

$$\begin{aligned}
&M_1(d_{ij}-1) \leq s_i - s_j \leq M_1 d_{ij}, \forall i \in I, j \in J,\\
&k - M_2 t_i \leq \sum_{j\in I\cup J} d_{ij} \leq k + M_2(1-t_i), \forall i \in I \cup J,
\end{aligned}$$

Equivalent Bilevel Reformulation

- Lower-level problem for kNN selection:

$$\boldsymbol{t} \in \operatorname*{argmin}_{t}\left\{\sum_{j=1}^N s_j t_j : \sum_{j=1}^N t_j = k,\ t_j \in [0,1], \forall j \in [N]\right\}.$$

- Reduce to single-level using strong duality for LP:

$$\begin{aligned}
&\min_{z\in Z} \frac{1}{N}\sum_{i=1}^N c(\boldsymbol{z}, y_i) && (2a)\\
&s.t.\ (1e) - (1h), && (2b)\\
&\beta - \pi_j \leq s_j, \forall j \in [N], && (2c)\\
&\sum_{j=1}^N s_j t_j = -\sum_{j=1}^N \pi_j + k\beta, && (2d)\\
&\boldsymbol{\pi} \geq \boldsymbol{0}. && (2e)
\end{aligned}$$

ER-DD-SAA with CART

- Classifiation and Regression Tree (CART):

$$\hat{y}_{CART}(\boldsymbol{z},\boldsymbol{w}) = \sum_{r=1}^{N_2} \hat{y}^r \cdot \mathbb{I}\{(\boldsymbol{z},\boldsymbol{w}) \in \mathcal{B}_r\},$$

- $\mathcal{B}_r = \{(\boldsymbol{z},\boldsymbol{w}) \in \mathbb{R}^{d_z}\times\mathbb{R}^{d_w}: \boldsymbol{a}^r < \boldsymbol{z} \leq \boldsymbol{b}^r, \boldsymbol{a}^r_w < \boldsymbol{w} \leq \boldsymbol{b}^r_w\}$.
- $\hat{y}^r = \frac{1}{|\mathcal{B}_r|}\sum_{i\in\mathcal{B}_r} y^i$: average response over samples assigned to region $\mathcal{B}_r$.
- $l_r, u_r \in \{0,1\}$: $l_r = 1$/ $u_r = 1$ if $\boldsymbol{z}$ satisfies the lower/upper bounds of region r:
  - $l_r = 1 \Rightarrow \boldsymbol{z} \geq \boldsymbol{a}^r + \delta\boldsymbol{1}$
  - $u_r = 1 \Rightarrow \boldsymbol{z} \leq \boldsymbol{b}^r$
- $R_r \in \{0,1\}$: $R_r = 1$ if query point $\boldsymbol{z}$ lies within region r:
  $R_r = 1 \Rightarrow \boldsymbol{a}^r + \delta\boldsymbol{1} \leq \boldsymbol{z} \leq \boldsymbol{b}^r$

$$\begin{aligned}
&\min_{z\in Z} \frac{1}{N}\sum_{i=1}^N c(\boldsymbol{z}, y_i) && (3a)\\
&s.t.\ M(l_r - 1) + \delta\boldsymbol{1} \leq \boldsymbol{z} - \boldsymbol{a}^r \leq M l_r, \forall r \in [N_2], && (3b)\\
&M(u_r - 1) \leq \boldsymbol{b}^r - \boldsymbol{z} \leq M u_r - \delta\boldsymbol{1}, \forall r \in [N_2], && (3c)\\
&l_r + u_r \geq 2R_r, \forall r \in [N_2], && (3d)\\
&\sum_{r=1}^{N_2} R_r = 1, && (3e)\\
&y_i = \sum_{r=1}^{N_2} R_r \hat{y}^r + \hat{\epsilon}^i_{CART}. && (3f)
\end{aligned}$$

ER-DD-SAA with ReLU Neural Networks

- Neural Networks prediction model:
  $\boldsymbol{h}^0 = [\boldsymbol{z},\boldsymbol{w}]$,
  $h^{(l)} = \sigma^{(l)}(\boldsymbol{v}^{(l)}\boldsymbol{h}^{(l-1)} + \boldsymbol{\alpha}^{(l)}), \forall l \in [L]$.
- ReLU activation function: $\sigma(x) = \max\{0,x\}$.
- Output layer: $\hat{y}_{NN} = \boldsymbol{v}^{(L+1)}\boldsymbol{h}^{(L)} + \boldsymbol{\alpha}^{(L+1)}$.
- $\{\hat{v}^{(l)}, \hat{\alpha}^{(l)}\}_{l=1}^{L+1}$ are learned by minimizing empirical MSE.

$$\begin{aligned}
&\min_{z\in Z} \frac{1}{N}\sum_{i=1}^N c(\boldsymbol{z}, y_i) && (4a)\\
&s.t.\ \boldsymbol{h}^0 = [\boldsymbol{z},\boldsymbol{w}], && (4b)\\
&h^{(l)} = \sigma^{(l)}(\boldsymbol{v}^{(l)}\boldsymbol{h}^{(l-1)} + \boldsymbol{\alpha}^{(l)}), \forall l \in [L], && (4c)\\
&\hat{y}_{NN} = \boldsymbol{v}^{(L+1)}\boldsymbol{h}^{(L)} + \boldsymbol{\alpha}^{(L+1)}, && (4d)\\
&y_i = \hat{y}_{NN} + \hat{\epsilon}^i_{NN}. && (4e)
\end{aligned}$$

Consistency of ER-DD-SAA

Theorem (Consistency and Asymptotic Optimality).
If the regression estimator is uniform-in-decision consistent, i.e., $\min_{z\in Z}|\hat{Q}_N(\boldsymbol{z},\boldsymbol{w}) - Q^*(\boldsymbol{z},\boldsymbol{w})| \to 0$, as $N \to \infty$, for a.e. $\boldsymbol{w} \in W$, then, we have for a.e. $\boldsymbol{w} \in W$

$$(i)\ \hat{v}_N(\boldsymbol{w}) \xrightarrow{P} v^*(\boldsymbol{w})$$
$$(ii)\ \mathbb{D}(\hat{S}_N(\boldsymbol{w}), S^*(\boldsymbol{w})) \xrightarrow{P} 0$$
$$(iii)\ \sup_{z\in\hat{S}_N(\boldsymbol{w})} g(\boldsymbol{z},\boldsymbol{w}) \xrightarrow{P} v^*(\boldsymbol{w})$$

$\hat{S}_N(\boldsymbol{w})/\hat{v}_N(\boldsymbol{w})$: optimal solution set/objective of ER-DD-SAA.
$S^*(\boldsymbol{w})/v^*(\boldsymbol{w})$: optimal solution set/objective of DD-CSP.
$g(\boldsymbol{z},\boldsymbol{w})$: true expected cost.

Numerical Experiments

- Ground-truth demand model in newsvendor problem:

$$y = \frac{500}{1 + exp(p-10)} \cdot \max\{2w - 110, 0\} + \epsilon.$$

Computational Time of kNN

- As a benchmark, [2] proposed the following alternative formulation for (1b) − (1d):
  $s_i - s_j \leq M(t_j - t_i + 1), \forall i,j \in [N]$. (5)

Table 1. Computational Time Comparison for kNN

N	k	(5)	Bilevel (2)	Algorithm+(5)	Algorithm+(1)
500	1	107.99	**51.96**	125.24	136.41
1000	1	1,013.45	741.33	878.53	**511.00**
1500	1	3,144.03	1557.29	2,617.81	**1,332.20**
500	2	252.25	**131.99**	281.16	942.64
1000	2	2,379.66	**1489.06**	2,426.54	8,444.56
1500	2	7,049.64	**2948.52**	6,731.62	>10,800(204%)

Computational Time of CART

Table 2. Computational Time Comparison for CART

N	MILP (3)			Gurobi ML		
	IS	OOS	Time(s)	IS	OOS	Time(s)
1000	-57,019.52	-47,744.39	**0.94**	-57,019.52	-47,743.59	22.81
3000	-52,818.05	-49,381.72	**3.14**	-52,818.05	-49,383.92	105.18
5000	-50,477.90	-50,841.05	**5.60**	-50,477.90	-50,841.05	371.81
7000	-51,437.71	-50,053.88	**8.75**	-51,437.71	-50,053.88	453.47

Computational Time of NN

Table 3. Computational Time Comparison for NN

N	IS	OOS	Training Time(s)	Opt. Time(s)
1000	-54,797.54	-54,813.98	38.60	19.72
3000	-54,814.90	-54,818.87	110.64	110.72
5000	-54,815.16	-54,819.11	201.18	506.29
7000	-54,816.37	-54,819.11	279.86	471.94

Comparison of Different Regression Models

Table 4. Comparison of OOS Cost across regression models

N	Linear	CART	kNN	NN
500	-48,824.34	-47,667.35	-54,427.11	**-54,809.19**
1000	-48,871.54	-47,843.39	-54,649.79	**-54,813.98**
1500	-48,793.94	-49,090.62	-54,723.50	**-54,818.67**
2000	-48,919.54	-50,472.93	-54,780.53	**-54,818.02**

- ➢ The proposed MIP formulations achieve substantial reductions in computational time.
- ➢ Across different parametric and nonparametric regression models, NN obtains the best OOS cost.

References

1. Kannan, R., Bayraksan, G., & Luedtke, J. R. (2025). Data-driven sample average approximation with covariate information. *Operations Research*, *73*(6), 3245-3259.
2. Liu, W., & Zhang, Z. (2023). Solving data-driven newsvendor pricing problems with decision-dependent effect. *arXiv preprint arXiv:2304.13924*.