Contextual Stochastic Optimization

Learning decision policies from data

Luís Henrique Domingues Bueno
FGV - EMAp

November 4, 2026

Outline

1. Stochastic Optimization: A Recap
2. The Contextual Gap
3. The Three CSO Frameworks
4. Decision Rules (DR)
5. Sequential Learning and Optimization (SLO)
6. Integrated Learning and Optimization (ILO)
7. Contextual Information and Evaluation

The classical stochastic optimization problem

A decision must be made before the uncertain outcome is observed.

Two-stage stochastic optimization

$$Q^* = \min_{z \in \mathcal{Z}} \mathbb{E}[c(z, Y)].$$

- z: here-and-now decision.
- Y: uncertain future outcome.
- $c(z, Y)$: realized cost.
- $\mathcal{Z}$: feasible decision set.

The challenge: choose a good decision without knowing the future realization of Y.

Expected Value, SAA, and Wait-and-See

Expected Value (EV)

Replace uncertainty by its mean:

$$z_{\mathrm{EV}} \in \arg\min_{z \in \mathcal{Z}} c(z, \mathbb{E}[Y]).$$

Evaluate under the true distribution:

$$Q_{\mathrm{EEV}} = \mathbb{E}[c(z_{\mathrm{EV}}, Y)].$$

SAA

Approximate the expected cost using samples:

$$\hat{z}_N \in \arg\min_{z} \frac{1}{N} \sum_{i=1}^{N} c(z, Y_i).$$

SAA approximates the stochastic problem.

Wait-and-See (WS)

Decide after observing Y:

$$Q_{\mathrm{WS}} = \mathbb{E}\left[\min_{z} c(z, Y)\right].$$

This is a perfect-information benchmark.

VSS and EVPI

Let

$$Q^* = \min_{z \in \mathcal{Z}} \mathbb{E}[c(z, Y)].$$

Value of the Stochastic Solution

$$\mathrm{VSS} = Q_{\mathrm{EEV}} - Q^* \geq 0.$$

Measures the benefit of solving the stochastic problem instead of using the expected-value solution.

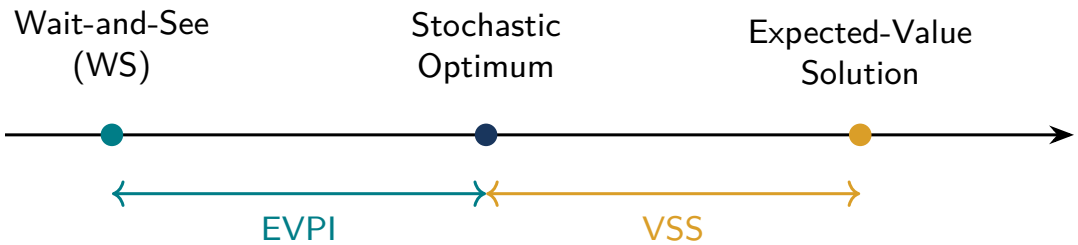
Expected Value of Perfect Information

$$\mathrm{EVPI} = Q^* - Q_{\mathrm{WS}} \geq 0.$$

Measures the benefit of knowing the realization before deciding.

$$\boxed{Q_{\mathrm{WS}} \leq Q^* \leq Q_{\mathrm{EEV}}}$$

Interpreting the classical benchmarks

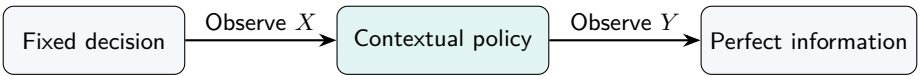


- VSS measures the cost of ignoring uncertainty.
- EVPI measures the value of perfect information.

From stochastic optimization to contextual optimization

Suppose features X are observed before the decision, and we must take a action $z(X)$ while Y remains uncertain.
The optimal contextual stochastic value is

$$Q^*_{\text{ctx}} = \mathbb{E}_X \left[\min_{z(\cdot)\in\mathcal{Z}} \mathbb{E}[c(z(X), Y) \mid X] \right].$$



Under the same distribution and feasible decisions,

$$\boxed{Q_{\text{WS}} \leq Q^*_{\text{ctx}} \leq Q^*_{\text{fixed}}}$$

where

$$Q^*_{\text{fixed}} = \min_{z\in\mathcal{Z}} \mathbb{E}[c(z, Y)].$$

Why context matters?

Suppose there are two types of days:

$$X \in \{\text{ordinary}, \text{promotion}\}.$$

Demand is typically around 50 units on ordinary days and around 100 units on promotion days.

- A non-contextual policy chooses one capacity level for both types.
- A contextual policy can reserve different capacity for each type.
- The benefit comes from adapting to $P(Y \mid X)$, not from knowing the future realization.

Three frameworks for contextual optimization

All three frameworks aim to learn a policy

$$\pi : \mathcal{X} \to \mathcal{Z}$$

with low expected realized cost:

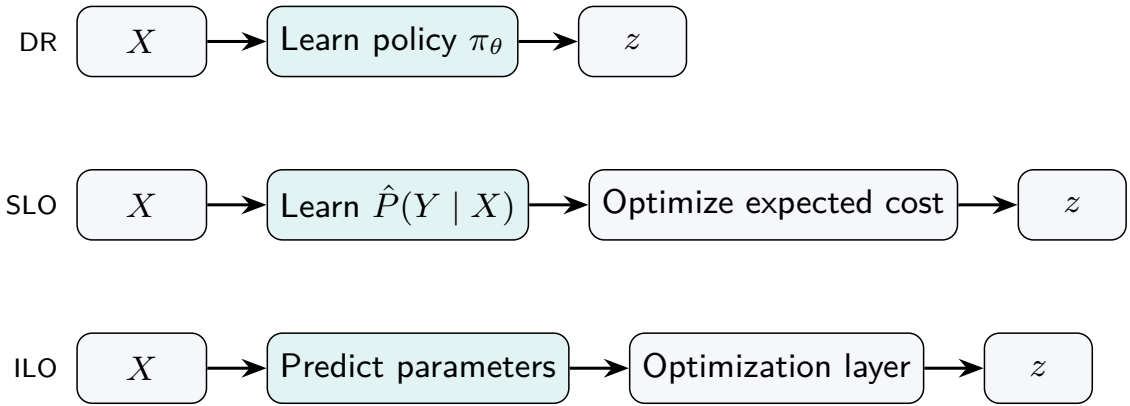
$$\min_{\pi \in \Pi} \mathbb{E}[c(\pi(X), Y)].$$

DR – Decision Rules: learn the action directly as a function of context.

SLO – Sequential Learning and Optimization: estimate the conditional distribution, then optimize.

ILO – Integrated Learning and Optimization: train a predictive component using downstream decision quality.

How the three frameworks work



DR: what is a decision rule?

A decision rule is a parameterized policy:

$$z = \pi_\theta(X).$$

Learn the policy parameters by minimizing empirical decision cost:

$$\hat{\theta} \in \arg\min_{\theta} \left\{ \frac{1}{N} \sum_{i=1}^{N} c(\pi_\theta(x_i), y_i) + \lambda \Omega(\theta) \right\}.$$

- The target is the action, not necessarily a prediction of Y.
- The loss can be aligned directly with operational cost.
- Feasibility must be built into the policy or enforced by constraints.

Introduced to the Operations Research community by Liyanage and Shanthikumar (2005), and popularized by Ban and Rudin (2019).

Why use Decision Rules?

Advantages

- Directly optimizes decisions.
- Simple deployment: evaluate $\pi_\theta(x)$.
- Does not require estimating a full conditional distribution.
- Can incorporate business constraints.

Challenges

- Choosing a suitable policy class.
- Nonconvex training for complex policies.
- Risk of overfitting.
- Feasibility may be difficult to guarantee.

Linear Decision Rules (LDRs)

For a continuous decision, a linear decision rule can be written as

$$z(x) = \beta_0 + \beta^\top x.$$

With an augmented feature vector

$$\tilde{x} = (1, x^\top)^\top,$$

this becomes

$$z(x) = \beta^\top \tilde{x}.$$

A regularized decision-focused objective is

$$\min_{\beta} \frac{1}{N} \sum_{i=1}^{N} c(\beta^\top \tilde{x}_i, y_i) + \lambda \|\beta\|_2^2.$$

Introduced by Ban and Rudin (2019), at the paper that popularized DR.

RKHS-based decision rules

A Reproducing Kernel Hilbert Space allows flexible nonlinear policies.
A policy can be represented as

$$\pi(x) = \sum_{i=1}^{N} \alpha_i k(x_i, x).$$

For a Gaussian kernel,

$$k(x, x') = \exp\left(-\frac{\|x - x'\|_2^2}{2\sigma^2}\right).$$

A regularized objective is

$$\min_{\pi \in \mathcal{H}_k} \frac{1}{N} \sum_{i=1}^{N} c(\pi(x_i), y_i) + \lambda \|\pi\|_{\mathcal{H}_k}^2.$$

- Nonlinear policies without manually specifying every interaction.
- The representer theorem often reduces the problem to finite coefficients.
- The kernel bandwidth controls locality and smoothness.

Also at Ban and Rudin (2019).

SLO: what is it?

SLO first estimates the conditional distribution and then optimizes:

$$x \longmapsto \widehat{P}_\phi(Y \mid X = x) \longmapsto \pi_\phi(x).$$

The decision solves

$$\pi_\phi(x) \in \arg\min_{z \in \mathcal{Z}} \mathbb{E}_{\widehat{P}_\phi(\cdot|x)}[c(z, Y)].$$

- Preserves distributional information, not only a point prediction.
- Can represent skewness, tails, and multimodal outcomes.
- Can reuse statistical models and stochastic optimization solvers.
- Distributional errors can propagate into decisions.

When is SLO useful?

When the decision depends on distributional characteristics beyond the expectation summary.

It was first formalized at Bertsimas D, McCord C and Sturt B (2023).

Residual-based conditional distributions

Start with a point predictor:

$$\hat{\mu}(x) \approx \mathbb{E}[Y \mid X = x].$$

Compute out-of-fold residuals on training observations, where $\hat{\mu}_{-i}(x_i)$ is trained without observation i:

$$e_i = y_i - \hat{\mu}_{-i}(x_i).$$

For a new context x, approximate the conditional distribution by

$$\widehat{P}(Y \mid x) \approx \frac{1}{N} \sum_{i=1}^{N} \delta_{\hat{\mu}(x)+e_i}.$$

Before the formalization of SLO, it already appears at Sen S and Deng Y (2017).

Kernel-weighted conditional distributions

Give more weight to observations with similar contexts:

$$w_i(x) = \frac{K_h(x - x_i)}{\sum_{j=1}^{N} K_h(x - x_j)}.$$

The conditional empirical distribution becomes

$$\widehat{P}(Y \mid x) = \sum_{i=1}^{N} w_i(x)\delta_{y_i}.$$

For a Gaussian kernel:

$$K_h(x - x_i) = \exp\left(-\frac{\|x - x_i\|_2^2}{2h^2}\right).$$

The corresponding optimization problem is

$$\min_{z \in \mathcal{Z}} \sum_{i=1}^{N} w_i(x)c(z, y_i).$$

First used by Bertsimas D and Kallus N (2020).

ILO: the key idea

Suppose the decision depends on an uncertain quantity Y.
A predictive model may try to estimate the entire conditional distribution

$$P(Y \mid X = x).$$

But the optimizer does not necessarily need to know the distribution accurately everywhere. It only needs the information that changes the optimal decision.

Decision-focused learning

Instead of asking

"How accurately can we predict Y?"

we ask

"How accurately can we predict what matters for the decision?"

ILO trains the predictive component using the downstream decision objective:

$$x \xrightarrow{g_\theta} \widehat{P}_\theta(Y \mid x) \xrightarrow{\text{optimizer}} z_\theta(x).$$

Why prediction accuracy is not always decision quality

Consider a simple decision with two possible actions:

$$z \in \{A, B\}.$$

Suppose two contexts produce different conditional demand distributions but, despite these differences:

$$\arg\min_{z \in \{A,B\}} \mathbb{E}_{P_1}[c(z, Y)] = \arg\min_{z \in \{A,B\}} \mathbb{E}_{P_2}[c(z, Y)] = A.$$

What ILO tries to learn

The goal is therefore not necessarily to reproduce the true conditional distribution as accurately as possible.
Instead, the predictive model should learn the aspects of uncertainty that are relevant for the optimization problem.

Predictive model:

$$\widehat{P}_\theta(Y \mid x) = g_\theta(x)$$

Decision problem:

$$z^*_\theta(x) \in \arg\min_{z \in \mathcal{Z}} \mathbb{E}_{g_\theta(x)}[c(z, Y)]$$

Learning objective:

$$\min_\theta \frac{1}{N} \sum_{i=1}^{N} c(z^*_\theta(x_i), y_i)$$

What´s the difference between DR and ILO

The key distinction is the object being learned.

Decision Rule Learning

Directly parameterize the decision:

$$z_\theta(x) = f_\theta(x).$$

Learn how decisions should respond to observed information.

Integrated Learning and Optimization

Learn a predictive model:

$$\widehat{P}_\theta(Y \mid x) \to z_\theta(x)$$

Use the learned model to construct and solve an optimization problem, after that it makes a decision.

These approaches are complementary: an ILO framework can incorporate decision rules as its optimization component, but it relies on a predictive model to characterize uncertainty beforehand

First used by Bertsimas D and Kallus N (2020).

ILO method 1: regret minimization

Define regret relative to a hindsight-optimal action:

$$\text{Regret}(z_\theta(x), y) = c(z_\theta(x), y) - \min_{z \in \mathcal{Z}} c(z, y),$$

Train by minimizing average regret:

$$\min_\theta \frac{1}{N} \sum_{i=1}^{N} \text{Regret}(z_\theta(x_i), y_i).$$

Elmachtoub and Grigas (2022) were the first to map a solution to contextual problems from the point of view regret.

ILO method 2: Optimal-action imitation

Generate hindsight-optimal actions from historical observations:

$$z_i^* \in \arg\min_{z \in \mathcal{Z}} c(z, y_i).$$

Use the predicted distribution to obtain decisions:

$$z_\theta^*(x_i) \in \arg\min_{z \in \mathcal{Z}} \mathbb{E}_{g_\theta(x)}[c(z, Y)].$$

Imitation objective

$$\min_\theta \frac{1}{N} \sum_{i=1}^{N} d\left(z_\theta^*(x_i), z_i^*\right).$$

Kong L et al. (2022) used as a way to generalize the regret aprroach.

The Ideal Contextual Stochastic Optimization Problem

Assume an ideal setting where the conditional distribution $P(Y \mid X = x)$ is perfectly known for every context x.

Ideal SLO benchmark

$$Q^*_{\mathrm{CSO}} = \mathbb{E}_X \left[\min_{z \in \mathcal{Z}} \mathbb{E}_{Y|X}[c(z, Y) \mid X] \right].$$

Three levels of information:

- **Stochastic Optimization:** no context is observed.

$$Q^*_{\mathrm{SO}} = \min_{z \in \mathcal{Z}} \mathbb{E}[c(z, Y)].$$

- **Contextual Optimization:** observe X and know $P(Y \mid X)$, but not the realization of Y.
- **Wait-and-See:** observe both X and Y before deciding.

$$Q_{\mathrm{WS}} = \mathbb{E} \left[\min_{z \in \mathcal{Z}} c(z, Y) \right].$$

$$\boxed{Q_{\mathrm{WS}} \leq Q^*_{\mathrm{CSO}} \leq Q^*_{\mathrm{SO}}}$$

Value of Context and Perfect Information

Value of Context (VoC)

$$\boxed{\mathrm{VoC} = Q^*_{\mathrm{SO}} - Q^*_{\mathrm{CSO}}}$$

Benefit of observing context X before deciding.

Contextual EVPI

$$\boxed{\mathrm{EVPI}_{\mathrm{ctx}} = Q^*_{\mathrm{CSO}} - Q_{\mathrm{WS}}}$$

Additional benefit of knowing the realization of Y.

Information decomposition

$$\boxed{\mathrm{EVPI} = \mathrm{VoC} + \mathrm{EVPI}_{\mathrm{ctx}}}$$

Classical EVPI is decomposed into two information gains.

Contextual VSS

Replace the conditional distribution by its conditional mean:

$$z_{\mathrm{EV}}(x) \in \arg\min_z c(z, \mathbb{E}[Y \mid X = x]).$$

$$\boxed{\mathrm{VSS}_{\mathrm{ctx}} = Q^{\mathrm{ctx}}_{\mathrm{EEV}} - Q^*_{\mathrm{CSO}}}$$

Benefit of modeling the full conditional uncertainty instead of using its mean.

Evaluating Contextual Optimization Models

The ideal SLO solution is a theoretical benchmark: it assumes perfect knowledge of $P(Y \mid X = x)$.

Performance of a learned policy

For any feasible learned policy $\hat{\pi}$:

$$J(\hat{\pi}) = \mathbb{E}[c(\hat{\pi}(X), Y)].$$

Its gap to the ideal contextual solution is:

$$\boxed{\text{Gap}(\hat{\pi}) = J(\hat{\pi}) - Q^*_{\text{CSO}} \geq 0}$$

- **DR:** how close can a directly learned policy get? ($EVPI_{ctx}$ and Gap)
- **SLO:** how much does estimating $P(Y \mid X)$ affect decision quality? (VSS_{ctx} and Gap)
- **ILO:** how effectively can decision-focused learning approach the ideal performance? ($EVPI_{ctx}$, Gap and SLO comparative metrics)

Take-home messages

1. CSO uses observed context to adapt decisions before the uncertain outcome is known.
2. VSS and EVPI can be adapted, provided information assumptions and evaluation distributions are consistent.
3. DR learns the policy; SLO learns the conditional distribution; ILO learns predictions through decision quality.
4. No framework is universally best.

Questions?

Selected References I


G. Y. Ban and C. Rudin.
The Big Data Newsvendor: Practical Insights from Machine Learning.
Operations Research, 67(1):90–108, 2019.


D. Bertsimas and N. Kallus.
From Predictive to Prescriptive Analytics.
Management Science, 66(3):1025–1044, 2020.


D. Bertsimas and N. Koduri.
Data-driven Optimization: A Reproducing Kernel Hilbert Space Approach.
Operations Research, 70(1):454–471, 2022.


D. Bertsimas, C. McCord, and B. Sturt.
Dynamic Optimization with Side Information.
European Journal of Operational Research, 304(2):634–651, 2023.


A. N. Elmachtoub and P. Grigas.
Smart "Predict, then Optimize".
Management Science, 68(1):9–26, 2022.

Selected References II



L. Kong, J. Cui, Y. Zhuang, R. Feng, B. A. Prakash, and C. Zhang.
End-to-end Stochastic Optimization with Energy-based Model.
Advances in Neural Information Processing Systems, 35:11341–11354, 2022.



L. H. Liyanage and J. G. Shanthikumar.
A Practical Inventory Control Policy Using Operational Statistics.
Operations Research Letters, 33(4):341–348, 2005.



U. Sadana, A. Chenreddy, E. Delage, A. Forel, E. Frejinger, and T. Vidal.
A Survey of Contextual Optimization Methods for Decision-Making under Uncertainty.
European Journal of Operational Research, 320(2):271–289, 2025.
https://doi.org/10.1016/j.ejor.2024.03.020



S. Sen and Y. Deng.
Learning Enabled Optimization: Towards a Fusion of Statistical Learning and Stochastic Programming.
Available at: https://optimization-online.org/?p=14456, 2017.