COST-WEIGHTED CROSS-ENTROPY FOR DIFFUSION-BASED DECISION-FOCUSED LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Decision-Focused Learning (*DFL*) trains predictive models for downstream decision quality rather than predictive accuracy alone. For stochastic discrete optimization, however, DFL faces two fundamental challenges: downstream decisions are non-differentiable, while nonlinear objectives may depend on the full conditional outcome distribution rather than point predictions. To address these challenges, we propose *Cost-Weighted Cross-Entropy* (*CWCE*), a differentiable decision-alignment objective that uses realized downstream costs to explicitly align the normalized expected-cost profile induced by the learned distribution, without differentiating through the discrete optimizer. We establish two complementary theoretical results. First, arbitrarily small distributional error can induce arbitrarily large decision regret, demonstrating that distributional fidelity alone does not necessarily guarantee downstream decision alignment. Second, when CWCE is incorporated into diffusion training, the excess risk of the resulting joint objective quantitatively controls downstream decision regret. These results motivate combining diffusion training, which promotes fidelity in modeling conditional outcome uncertainty, with CWCE, which provides explicit decision-relevant supervision. Experiments on three stochastic discrete optimization tasks demonstrate consistent improvements in downstream decision quality and regret over deterministic, ranking-based, and generative DFL baselines.

1 INTRODUCTION

Decision-making under uncertainty is central to a wide range of real-world applications, including supply chain management, transportation, energy systems, finance, healthcare, and recommendation systems (Bengio et al., 2021; Davis et al., 2024; Khorshidi et al., 2024; Mandi et al., 2024). In these settings, predictive models estimate uncertain outcomes from observable inputs, and the resulting predictions are subsequently used to solve downstream optimization problems. A canonical approach is the *Predict-then-Optimize* (*PtO*) paradigm, where predictive models are trained using standard statistical losses and their predictions are then passed to an optimizer (Liu & Grigas, 2021; Shah et al., 2024). However, predictive accuracy does not necessarily translate into decision quality: prediction errors can have highly heterogeneous effects on downstream decisions, and statistically accurate predictions may still induce poor decisions (Bertsimas & Kallus, 2020; Sadana et al., 2025).

Decision-Focused Learning (*DFL*) (Donti et al., 2017; Elmachtoub & Grigas, 2022; Mandi et al., 2020; Wilder et al., 2019) addresses this mismatch by incorporating downstream decision quality directly into predictive learning. Classical DFL primarily relies on deterministic point predictions and differentiable surrogates, relaxations, or gradient estimators to propagate decision information through the optimization problem (Mandi et al., 2022; Shah et al., 2022; Mandi et al., 2024). While effective in many settings, deterministic prediction can be insufficient when downstream decisions depend on distributional properties of uncertain outcomes. For objectives of the form $f(y, z) = h(y)^\top z$, the conditional expected cost depends only on $\mathbb{E}[h(y) \mid x]$, so an appropriate point prediction can be sufficient. For more general nonlinear stochastic objectives, however, expected costs may depend on higher-order properties of the conditional distribution. Reducing uncertainty to a single point estimate can therefore discard decision-relevant information.

This observation motivates decision-focused distribution learning, in which a conditional generative model $p_\theta(y \mid x)$ represents outcome uncertainty and downstream decisions are obtained by minimiz-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ing expected cost under the learned distribution. Recent generative DFL methods further incorporate decision-aware objectives into distribution learning. Gen-DFL (Wang et al., 2026) combines generative modeling with contrastive decision-focused surrogates constructed from target and candidate decisions, while Diffusion-DFL (Zhao et al., 2026) uses conditional diffusion models and develops gradient estimators for directly optimizing downstream decision objectives in an end-to-end manner. Diffusion models are particularly appealing in this setting due to their ability to represent complex, multimodal, and high-dimensional conditional distributions (Ho et al., 2020; Song & Ermon, 2019; Dhariwal & Nichol, 2021). These developments highlight the potential of combining expressive conditional distribution learning with explicit downstream decision information.

Despite this progress, incorporating such decision information remains particularly challenging for discrete optimization. Directly differentiating through the downstream optimization problem (Zhao et al., 2026) becomes difficult when the $\arg\min$ operator is discrete and non-differentiable, while constructing surrogate objectives from optimal decisions (Wang et al., 2026) may require sampling from the learned generative model and solving the resulting optimization problem to obtain a target decision that subsequently serves as decision-level supervision. Because this supervision is derived from the learned distribution, distribution approximation errors can propagate to the target decision and, in turn, affect subsequent training.

In this paper, rather than differentiating through the discrete decision itself or relying on optimized target decisions as supervision, we directly target the conditional expected-cost profile, i.e., the expected costs across candidate decisions, whose ordering determines the optimal discrete decision. To this end, we propose *Cost-Weighted Cross-Entropy* (*CWCE*), a differentiable decision-alignment objective for stochastic discrete optimization. Given samples from the learned conditional distribution, CWCE estimates and normalizes the expected-cost profile across candidate decisions and uses realized downstream costs as sample-wise soft target weights for this normalized profile. At the population level, this objective aligns the learned normalized expected-cost profile with its true counterpart. Since normalization preserves expected-cost ordering, this alignment preserves the decision-relevant cost structure and yields the optimal discrete decision, without differentiating through, relaxing, or replacing the downstream $\arg\min$.

We establish two complementary theoretical results that characterize the role of CWCE in diffusion-based DFL. First, we show that arbitrarily small distributional error can induce arbitrarily large downstream decision regret, demonstrating that distributional fidelity alone does not necessarily guarantee downstream decision alignment. Second, we show that CWCE provides explicit decision-relevant supervision: when incorporated into diffusion training, the excess risk of the resulting joint objective quantitatively controls downstream decision regret. Together, these results formalize the complementary roles of diffusion training and CWCE: diffusion training promotes fidelity in modeling conditional outcome uncertainty, while CWCE explicitly aligns the normalized expected-cost profile that determines downstream decisions.

Our contributions are summarized as follows: *(i)* We propose CWCE, a differentiable decision-alignment objective that operates on normalized conditional expected-cost profiles without differentiating through the discrete optimizer. Normalization preserves the expected-cost ordering that determines the downstream decision, while CWCE aligns the learned and true normalized cost profiles, thereby providing explicit decision-relevant supervision. *(ii)* We provide two complementary theoretical results characterizing the role of CWCE in diffusion-based DFL. We first show that arbitrarily small distributional error can induce arbitrarily large decision regret, demonstrating that distributional fidelity alone does not necessarily guarantee decision alignment. We then show that, when CWCE is incorporated into diffusion training, the joint excess risk quantitatively controls downstream decision regret. *(iii)* We conduct extensive experiments on three stochastic discrete optimization tasks and show that CWCE consistently improves downstream decision quality and regret over deterministic, ranking-based, and generative DFL baselines.

2 Diffusion-Based Decision-Focused Learning

We first formulate stochastic discrete decision making under a learned conditional outcome distribution and introduce the diffusion-based DFL framework considered in this work. This formulation makes explicit how the learned distribution induces a conditional expected-cost profile over feasible decisions, which provides the basis for the decision-alignment objective developed in Section 3.

Stochastic decision problem. We consider a stochastic decision problem in which an input $x \in \mathcal{X}$ is associated with an uncertain outcome $y \in \mathcal{Y}$ drawn from an unknown conditional distribution $q(y \mid x)$. Given a feasible decision set $\mathcal{Z}$ and a nonnegative cost function $f : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}_+$, the Bayes-optimal decision minimizes the conditional expected cost:

$$z^*(x) \in \arg\min_{z \in \mathcal{Z}} \mathbb{E}_{y \sim q(\cdot|x)} \big[f(y, z) \big]. \tag{1}$$

For discrete optimization, $\mathcal{Z}$ is typically finite or combinatorial, and the resulting decision mapping is generally discontinuous and non-differentiable.

From decision-focused learning to distribution learning. Standard supervised learning first predicts uncertain outcomes and subsequently solves the downstream optimization problem. In contrast, *Decision-Focused Learning* (*DFL*) explicitly accounts for downstream decision quality during training. Given a deterministic predictor $\widehat{y}_\theta(x)$, the induced decision is

$$z_\theta(x) \in \arg\min_{z \in \mathcal{Z}} f\big(\widehat{y}_\theta(x), z\big),$$

and a natural decision-focused objective evaluates its realized downstream cost:

$$\mathcal{L}_{\mathrm{VA}}(\theta) = \mathbb{E}_{(x,y)\sim q} \left[f\big(y, z_\theta(x)\big) \right]. \tag{2}$$

Although directly aligned with decision quality, optimizing this objective requires differentiating through the optimization mapping, which is generally non-differentiable for discrete decision spaces. Existing DFL methods therefore rely on differentiable surrogates, relaxations, or gradient estimators.

Deterministic prediction can also be insufficient when downstream decisions depend on distributional properties of uncertain outcomes. For objectives of the form $f(y, z) = h(y)^\top z$, $\mathbb{E}_{q(y|x)}[f(y, z)] = \mathbb{E}_{q(y|x)}[h(y)]^\top z$, so the optimal decision depends only on the conditional expectation of $h(y)$. For more general nonlinear stochastic objectives, however, expected costs may depend on higher-order properties of $q(y \mid x)$. Reducing uncertainty to a point prediction can therefore discard decision-relevant information, motivating conditional distribution learning.

Generative decision-making. Following recent generative DFL formulations, we model the conditional outcome distribution $p_\theta(y \mid x)$ rather than a deterministic point prediction. The conditional expected cost induced by p_θ is $\mathcal{E}_\theta(z; x) := \mathbb{E}_{y \sim p_\theta(\cdot|x)} \big[f(y, z) \big]$, with the corresponding decision

$$z_\theta(x) \in \arg\min_{z \in \mathcal{Z}} \mathcal{E}_\theta(z; x). \tag{3}$$

When $p_\theta(\cdot|x) = q(\cdot|x)$, this rule recovers the Bayes-optimal decision in Eq. (1). In practice, $\mathcal{E}_\theta(z; x)$ can be estimated using Monte Carlo samples $\{\widehat{y}_s\}_{s=1}^S \sim p_\theta(\cdot \mid x)$ by $\widehat{\mathcal{E}}_\theta(z; x) = \frac{1}{S} \sum_{s=1}^S f(\widehat{y}_s, z)$. By Eq. (3), a conditional generative model $p_\theta(\cdot \mid x)$ determines downstream decisions through the induced expected-cost profile $\{\mathcal{E}_\theta(z; x)\}_{z \in \mathcal{Z}}$.

Diffusion-based DFL. We instantiate $p_\theta(y \mid x)$ using a conditional diffusion model. Given a clean outcome y, the forward process is $y_t = \sqrt{\bar{\alpha}_t}\, y + \sqrt{1 - \bar{\alpha}_t}\, \epsilon,\ \epsilon \sim \mathcal{N}(0, I)$, where $\bar{\alpha}_t$ is determined by a predefined noise schedule. A noise predictor $\epsilon_\theta(y_t, x, t)$ is trained using the standard diffusion model objective

$$\mathcal{L}_{\mathrm{DM}}(\theta) = \mathbb{E}_{(x,y)\sim q, t, \epsilon \sim \mathcal{N}(0,I)} \left[\|\epsilon - \epsilon_\theta(y_t, x, t)\|_2^2 \right].$$

This objective provides a standard distribution-matching signal for learning $p_\theta(y \mid x)$ and thereby promotes distributional fidelity. It does not, however, explicitly encode the downstream cost function $f(y, z)$ or the expected-cost ordering of $\{\mathcal{E}_\theta(z; x)\}_{z \in \mathcal{Z}}$ that determines discrete decisions.

Recent generative DFL methods complement distribution matching $\mathcal{L}_{\mathrm{DM}}$ with explicit decision-aware objectives $\mathcal{L}_{\mathrm{DA}}$. Diffusion-DFL (Zhao et al., 2026) directly optimizes the downstream decision objective $\mathcal{L}_{\mathrm{VA}}(\theta) = \mathbb{E}_q \big[f(y, z_\theta(x)) \big]$ as $\mathcal{L}_{\mathrm{DA}}$ using gradient estimators. Gen-DFL (Wang et al., 2026) instantiates $\mathcal{L}_{\mathrm{DA}}$ as contrastive surrogates over candidate decisions. Together, these works establish the paradigm of jointly optimizing distributional and decision-aware objectives $\mathcal{L}_{\mathrm{DM}} + \mathcal{L}_{\mathrm{DA}}$.

For discrete optimization, however, both approaches face challenges: contrastive surrogates may require sampling from the learned distribution and solving the downstream optimization problem to obtain target decisions as supervision, through which distribution learning errors can propagate into the resulting decision-level supervision. In contrast, directly optimizing $\mathcal{L}_{\mathrm{VA}}$ is hindered by the discrete, non-differentiable $\arg\min$ mapping from the expected-cost profile $\{\mathcal{E}_\theta(z; x)\}_{z \in \mathcal{Z}}$ to the induced decision $z_\theta(x)$.

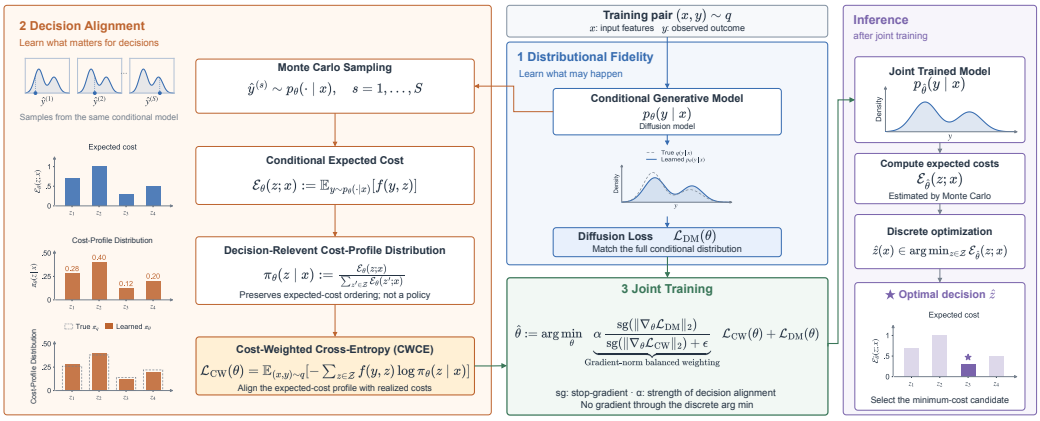

Figure 1: **Overview of Cost-Weighted Cross-Entropy (CWCE) for decision-focused distribution learning.** Given training pairs $(x, y) \sim q$, a conditional diffusion model $p_\theta(y|x)$ learns the outcome distribution $q(y|x)$ through the diffusion loss L_{DM}, providing *distributional fidelity*. Samples from the same model are used to estimate the expected-cost profile $\mathcal{E}_\theta(z; x)$, which is normalized into $\pi_\theta(z \mid x)$ while preserving the ordering that determines the optimal discrete decision. CWCE uses realized costs $f(y, z)$ to align this decision-relevant cost profile, providing *decision alignment* without differentiating through the discrete $\arg\min$. Joint training combines these complementary objectives, while inference solves the downstream optimization: expected costs are estimated under the jointly trained model and the minimum-cost candidate is selected.

3 Cost-Weighted Cross-Entropy for Decision Alignment

The previous section identifies the central challenge in decision-focused distribution learning for discrete optimization: distribution matching alone does not explicitly preserve the expected-cost ordering that determines downstream decisions, whereas directly optimizing decision quality requires handling a non-differentiable discrete $\arg\min$. We address this challenge with *Cost-Weighted Cross-Entropy* (*CWCE*), a differentiable decision-alignment loss that operates directly on the conditional expected-cost profile induced by the learned distribution. Figure 1 summarizes the overall framework: diffusion learning provides distributional fidelity, CWCE aligns the decision-relevant expected-cost profile, and the jointly trained model is used with the original discrete optimizer at inference.

Decision-Relevant Cost-Profile Distribution. Given a learned conditional distribution $p_\theta(y \mid x)$, recall $\mathcal{E}_\theta(z; x) = \mathbb{E}_{y \sim p_\theta(\cdot|x)}[f(y, z)]$. Since the downstream decision depends only on the ordering of the cost profile $\{\mathcal{E}_\theta(z; x)\}_{z \in \mathcal{Z}}$, we normalize these expected costs into the *cost-profile distribution*

$$\pi_\theta(z \mid x) := \frac{\mathcal{E}_\theta(z; x)}{\sum_{z' \in \mathcal{Z}} \mathcal{E}_\theta(z'; x)}. \tag{4}$$

For nonnegative costs with a positive denominator, $\pi_\theta(\cdot \mid x)$ defines a probability distribution over candidate decisions. Importantly, π_θ is *not* a decision policy: larger probability corresponds to larger expected cost. Rather, it is a normalized representation of the expected-cost profile induced by $p_\theta(y \mid x)$.

Because all expected costs share the same positive normalization, $\mathcal{E}_\theta(z_i; x) < \mathcal{E}_\theta(z_j; x)$ holds if and only if $\pi_\theta(z_i \mid x) < \pi_\theta(z_j \mid x)$. Thus, π_θ preserves precisely the ordering that determines the downstream discrete decision.

Cost-Weighted Cross-Entropy. For an observed training pair (x, y), the realized costs $\{f(y, z)\}_{z \in \mathcal{Z}}$ quantify the downstream consequences of candidate decisions. We use these costs as weights and define

$$\mathcal{L}_{\mathrm{CW}}(\theta; x, y) = -\sum_{z \in \mathcal{Z}} f(y, z) \log \pi_\theta(z \mid x), \tag{5}$$

with population objective $\mathcal{L}_{\mathrm{CW}}(\theta) = \mathbb{E}_{(x,y)\sim q}\big[\mathcal{L}_{\mathrm{CW}}(\theta; x, y)\big]$. CWCE can therefore be viewed as a soft-target cross-entropy over the candidate decision space, where $f(y, z)$ provides an unnormalized

soft target weight for each decision z. Here, $\pi_\theta(z \mid x)$ represents the normalized expected cost associated with decision z under the learned outcome distribution, rather than the probability of selecting z. Thus, realized downstream costs provide sample-wise supervision for learning the decision-relevant expected-cost profile. To see how this supervision operates at the population level, conditioning on x gives

$$\mathbb{E}_{y\sim q(\cdot|x)}[\mathcal{L}_{\mathrm{CW}}(\theta; x, y)] = -\sum_{z\in\mathcal{Z}} \mathcal{E}_q(z; x) \log \pi_\theta(z \mid x), \tag{6}$$

where $\mathcal{E}_q(z; x) := \mathbb{E}_{y\sim q(\cdot|x)}[f(y, z)]$ is the true conditional expected cost of decision z. Defining the corresponding cost-profile distribution as $\pi_q(z \mid x) := \mathcal{E}_q(z; x)/\sum_{z'\in\mathcal{Z}} \mathcal{E}_q(z'; x)$, we can rewrite Eq. (6) as

$$\mathbb{E}_{y\sim q(\cdot|x)}\left[\mathcal{L}_{\mathrm{CW}}(\theta; x, y)\right] = S_q(x)\left[H(\pi_q(\cdot \mid x)) + D_{\mathrm{KL}}\left(\pi_q(\cdot \mid x) \,\|\, \pi_\theta(\cdot \mid x)\right)\right],$$

where $S_q(x) := \sum_{z\in\mathcal{Z}} \mathcal{E}_q(z; x)$. Since $S_q(x) > 0$ and $H(\pi_q(\cdot \mid x))$ is independent of θ, minimizing the population CWCE is equivalent to aligning the learned and true cost-profile distributions. Under the realizability conditions, the minimum is attained when $\pi_\theta(z \mid x) = \pi_q(z \mid x)$. Because normalization in $\pi(z|x)$ rescales all expected costs by a common positive factor, this alignment preserves their ordering across candidate decisions, i.e., $\mathcal{E}_\theta(z_i; x) < \mathcal{E}_\theta(z_j; x)$ holds if and only if $\mathcal{E}_q(z_i; x) < \mathcal{E}_q(z_j; x)$ for any z_i, z_j. In particular, the decision minimizing the learned expected cost also minimizes the true expected cost, i.e., $\arg\min_{z\in\mathcal{Z}} \mathcal{E}_\theta(z; x) = \arg\min_{z\in\mathcal{Z}} \mathcal{E}_q(z; x)$. CWCE therefore provides decision-relevant supervision by aligning the expected-cost structure that determines the downstream optimum, without differentiating through the discrete $\arg\min$ in Eq. (3).

Integrating CWCE with Diffusion Training. CWCE complements the diffusion objective. The diffusion loss $\mathcal{L}_{\mathrm{DM}}$ promotes distributional fidelity, whereas $\mathcal{L}_{\mathrm{CW}}$ explicitly aligns the learned distribution with downstream expected-cost structure. We combine them during diffusion-model training:

$$\widehat{\theta} = \arg\min_{\theta} \left[\alpha \cdot \frac{\mathrm{sg}\left(\|\nabla_\theta \mathcal{L}_{\mathrm{DM}}\|_2\right)}{\mathrm{sg}\left(\|\nabla_\theta \mathcal{L}_{\mathrm{CW}}\|_2\right) + \varepsilon} \cdot \mathcal{L}_{\mathrm{CW}}(\theta) + \mathcal{L}_{\mathrm{DM}}(\theta)\right], \tag{7}$$

where $\alpha > 0$ controls the strength of decision alignment, $\mathrm{sg}(\cdot)$ denotes stop-gradient, and $\varepsilon = 10^{-8}$ ensures numerical stability. The batch-wise gradient-norm ratio compensates for scale differences between the two objectives without introducing additional gradients through the balancing coefficient.

This joint training retains the complementary roles of the two objectives: diffusion learning models the conditional outcome distribution, whereas CWCE steers the model toward preserving the decision-relevant expected-cost ordering. Importantly, CWCE provides this alignment without backpropagating through the discrete downstream optimizer.

At *inference* time, we solve the original stochastic optimization problem under the learned distribution: $\widehat{z}(x) \in \arg\min_{z\in\mathcal{Z}} \mathbb{E}_{y\sim p_{\hat{\theta}}(\cdot|x)} f(y, z)$, where the expectation is estimated by Monte Carlo sampling. Thus, CWCE changes how the predictive distribution is trained for decision alignment while leaving the downstream optimization problem itself unchanged.

Practical Optimization. For combinatorial problems, computing CWCE over the entire decision space $\mathcal{Z}$ can be computationally expensive. Following the *candidate-set* strategies commonly adopted in DFL (Mandi et al., 2022; Wang et al., 2026), we instead compute CWCE over a feasible candidate subset $\mathcal{S} \subseteq \mathcal{Z}$ by replacing $\mathcal{Z}$ with $\mathcal{S}$ in Eq. (4) and Eq. (5). We initialize $\mathcal{S} = \{z^*(y_i)\}_{i=1}^n$, where $z^*(y_i) \in \arg\min_{z\in\mathcal{Z}} f(y_i, z)$. During training, we additionally draw $\{y_i^{(m)}\}_{m=1}^M \sim p_\theta(\cdot \mid x_i)$ and, with probability $p_{\mathrm{solve}} = 0.1$, compute $\widehat{z}(x_i) \in \arg\min_{z\in\mathcal{Z}} \frac{1}{M}\sum_{m=1}^M f(y_i^{(m)}, z)$. We then update $\mathcal{S} \leftarrow \mathcal{S} \cup \{\widehat{z}(x_i)\}$, allowing the candidate set to track decisions that become relevant under the evolving learned distribution without exhaustively enumerating $\mathcal{Z}$.

Using the reparameterization $y = R(\varepsilon, x; \theta)$, $\varepsilon \sim p(\varepsilon)$, we obtain the unbiased estimator

$$\nabla_\theta \mathbb{E}_{y\sim p_\theta(\cdot|x)} f(y, z) = \mathbb{E}_{\varepsilon\sim p(\varepsilon)}\left[\nabla_\theta f(R(\varepsilon, x; \theta), z)\right].$$

Therefore, the proposed approach avoids computing gradients via the discrete decision variable. Instead, it propagates gradients through the reparameterizable conditional distribution $p_\theta(y \mid x)$ over the continuous outcome y. Algorithm 1 in the appendix summarizes the complete CWCE training procedure, including diffusion pre-training, dynamic candidate-set construction, and CWCE-augmented fine-tuning.

4 THEORETICAL ANALYSIS

We provide two complementary theoretical results that characterize the role of CWCE in diffusion-based DFL. First, we show that distributional fidelity alone does not guarantee decision alignment: an arbitrarily small distributional discrepancy can still induce arbitrarily large downstream decision regret. Second, we show that once CWCE is incorporated into the training objective, the excess risk of the resulting joint objective quantitatively controls downstream decision regret. Together, these results formalize the complementary roles of distribution matching and CWCE: distribution matching promotes fidelity in modeling conditional outcome uncertainty, whereas CWCE explicitly controls the normalized expected-cost profile that determines downstream decisions.

4.1 DISTRIBUTIONAL FIDELITY DOES NOT GUARANTEE DECISION ALIGNMENT

We first examine whether accurately approximating the outcome distribution is sufficient for downstream decision making. In discrete optimization, the decision is determined by the ordering of expected costs across candidate decisions. Distributional proximity alone, however, does not necessarily preserve this ordering. The following result makes this separation explicit.

Theorem 4.1 (Small Distributional Error Can Induce Arbitrarily Large Regret)**.** *For any $\delta > 0$ and any $R > 0$, there exist a true distribution q, a learned distribution p, a finite decision set $\mathcal{Z}$, and a nonnegative cost function f such that $D_{\mathrm{KL}}(q \,\|\, p) < \delta$, while the regret of the decision induced by p, evaluated under q, satisfies*

$$\mathrm{Regret}_q(p) := \mathbb{E}_x \left[\mathcal{E}_q(z_p(x); x) - \mathcal{E}_q(z^*(x); x)\right] > R,$$

where $z_p(x) \in \arg\min_{z\in\mathcal{Z}} \mathcal{E}_p(z; x)$ and $z^*(x) \in \arg\min_{z\in\mathcal{Z}} \mathcal{E}_q(z; x)$.

The construction exploits a rare but decision-critical region of the outcome space. Its probability can be perturbed by an arbitrarily small amount, keeping the KL divergence arbitrarily small, while sufficiently large costs in this region can reverse the expected-cost ordering and induce arbitrarily large decision regret. The complete construction and proof are provided in Appendix B.1.

Theorem 4.1 establishes a fundamental separation between *distributional fidelity* and *decision alignment*. A learned distribution can be arbitrarily close to the true distribution while still inducing an incorrect expected-cost ordering and hence a poor downstream decision. Distributional approximation error alone therefore cannot, in general, control downstream decision regret.

This result does not diminish the importance of distribution learning. Rather, the two objectives capture complementary aspects of stochastic decision making: distributional fidelity concerns faithful representation of outcome uncertainty, whereas decision alignment concerns preservation of the expected-cost structure that determines downstream decisions. This motivates augmenting diffusion-based distribution learning with an explicit decision-alignment objective such as CWCE.

4.2 JOINT EXCESS RISK CONTROLS DECISION REGRET

We next show that incorporating CWCE provides precisely such decision-level control. For population analysis, we abstract the product of α and the gradient-balancing factor in Eq. (7) into a generic positive coefficient λ and consider

$$\mathcal{L}_{\mathrm{joint}}(p) = \mathcal{L}_{\mathrm{DM}}(p) + \lambda \cdot \mathcal{L}_{\mathrm{CW}}(p), \qquad \lambda > 0. \tag{8}$$

Here, $\mathcal{L}_{\mathrm{DM}}$ promotes distributional fidelity, while $\mathcal{L}_{\mathrm{CW}}$ explicitly aligns the normalized expected-cost profile. The following theorem connects optimization of this joint objective directly to downstream decision quality.

Theorem 4.2 (Joint Excess Risk Controls Decision Regret)**.** *Assume that the CWCE cost profiles are well defined and that $\sum_{z\in\mathcal{Z}} \mathbb{E}_{x\sim q}\mathcal{E}_q(z; x) \le B$. For any $\epsilon > 0$, if $\mathcal{L}_{\mathrm{joint}}(p) - \mathcal{L}_{\mathrm{joint}}(q) \le \epsilon$, then*

$$\mathrm{Regret}_q(p) \le \sqrt{2B/\lambda} \cdot \sqrt{\epsilon}.$$

Theorem 4.2 establishes a direct calibration guarantee between the learning objective and downstream decision quality: as the joint excess risk vanishes, the induced decision regret also vanishes at the

rate $O(\sqrt{\epsilon})$. The result follows because the CWCE excess risk controls the divergence between the true and learned normalized expected-cost profiles, which in turn controls the regret of the induced discrete decision.

Together, Theorems 4.1 and 4.2 establish the theoretical motivation for augmenting diffusion training with CWCE. The first result shows that distributional proximity alone cannot generally control downstream decision regret, whereas the second establishes that the CWCE-augmented objective provides explicit quantitative control of such regret. Diffusion matching and CWCE therefore play complementary roles: the former promotes fidelity in modeling outcome uncertainty, while the latter explicitly controls the normalized expected-cost profile that determines downstream decisions.

5 EXPERIMENTS

We evaluate CWCE on stochastic discrete optimization tasks from three complementary perspectives: *(i)* its downstream decision quality relative to deterministic and generative DFL baselines; *(ii)* the benefit of explicit expected-cost alignment beyond distribution-only diffusion training; and *(iii)* its computational efficiency relative to generative DFL methods.

5.1 EXPERIMENTAL SETUP

Tasks and datasets. We consider three stochastic combinatorial optimization tasks with distinct decision structures: `Web Advertising`, `Portfolio Optimization`, and `Diverse Recommendation`. These tasks cover diverse combinatorial decision structures and downstream objectives, where conditional outcome uncertainty can affect the quality of the resulting decisions. Further details are provided in Appendix C.1.

`Web Advertising`. Following (Wilder et al., 2019), each instance contains $m = 100$ websites, $K = 500$ users, and a budget $b = 5$. The click-through-rate vector of website i is $y_i \in [0, 1]^K$, obtained from the Yahoo! Webscope dataset (Yahoo! Webscope, 2007), with features generated by $x_i = Ay_i$. The downstream problem is

$$\min_{z\in\{0,1\}^m} \quad \sum_{j=1}^{K}\prod_{i=1}^{m}(1 - z_i y_{ij}) \qquad \text{s.t. } \sum_{i=1}^{m} z_i = b.$$

`Portfolio Optimization`. Following stochastic portfolio selection (Donti et al., 2017; Wang et al., 2020), each instance contains $m = 50$ assets with budget $b = 3$. Features are derived from historical indicators, while $y_i \in \mathbb{R}$ denotes the future return of asset i. We solve

$$\min_{z\in\{0,1\}^m} \quad -y^\top z + \beta z^\top \Sigma z \qquad \text{s.t. } \sum_{i=1}^{m} z_i = b,$$

where Σ is the covariance matrix and $\beta = 0.1$. To ensure positive expected costs, we subtract from all costs the minimum realized cost over the training set and candidate decision space, $\min_{(x,y)\in D,\, z\in\mathcal{Z}} f(y, z)$, and add a small constant 10^{-5}, where D denotes the training set; this constant shift does not affect the ordering of candidate decisions.

`Diverse Recommendation`. Following (Wilder et al., 2019; Ashkan et al., 2015), each instance contains $m = 100$ items, $K = 500$ topics, and recommendation budget $b = 20$. The objective minimizes the number of uncovered topics:

$$\min_{z\in\{0,1\}^m} \quad \sum_{j=1}^{K}\mathbf{1}\left\{\sum_{i=1}^{m} z_i y_{ij} = 0\right\} \qquad \text{s.t. } \sum_{i=1}^{m} z_i = b.$$

Since the indicator-based objective is non-differentiable with respect to y, following Wilder et al. (2019), we use its continuous relaxation as the training cost for the `Web Advertising` task, while retaining the original number of uncovered topics as the evaluation cost.

Compared methods. We compare against three groups of baselines. First, *PtO* predicts the outcome using only the least squared loss and subsequently solves the optimization problem under

the predicted outcome. Second, deterministic DFL methods include Quadratic (Shah et al., 2022), Pointwise Ranking (Mandi et al., 2022), and Listwise Ranking (Mandi et al., 2022). Third, Diffusion-DFL (Zhao et al., 2026) incorporates downstream decision information into diffusion training through optimization-aware gradients. Because Diffusion-DFL is formulated for continuous decision spaces, we apply it using a continuous relaxation followed by top-b projection to obtain feasible discrete decisions. This comparison distinguishes CWCE from both distribution-only diffusion learning and existing decision-focused diffusion training. Additional baseline details are provided in Appendix C.2.

Implementation. All diffusion-based methods use the same architecture, sampling procedure, data splits, and optimization settings whenever applicable. We first pre-train the diffusion model with $\mathcal{L}_{\mathrm{DM}}$ and then fine-tune it using the CWCE-augmented objective in Eq. (7). Conditional expected costs are estimated by Monte Carlo sampling. Unless otherwise stated, all methods use the same downstream solver and evaluation protocol. Full implementation details are provided in Appendix C.3.

Evaluation metrics. We evaluate downstream performance using normalized decision quality (nDQ) and percentage regret. We split each dataset into training, validation, and test sets. Let $D_t = \{(x_i, y_i)\}_{i=1}^{n_t}$ denote the test set, where the outcomes y_i are unobserved when making decisions. Following the standard definition of nDQ in (Shah et al., 2022), we first compute the random decision quality (RDQ) and optimal decision quality (ODQ): $\mathrm{RDQ} = n_t^{-1} \sum_{i=1}^{n_t} \min_{k=1,\dots,10} f(y_i, z_k)$, where $z_k \in \mathcal{Z}$ are 10 randomly generated feasible decisions, and $\mathrm{ODQ} = n_t^{-1} \sum_{i=1}^{n_t} f(y_i, z^*(x_i))$. Similarly, the decision quality of the learned model is $\mathrm{DQ} = n_t^{-1} \sum_{i=1}^{n_t} f(y_i, z_{\widehat{\theta}}(x_i))$. The normalized decision quality is then defined as $\mathrm{nDQ} = (\mathrm{RDQ} - \mathrm{DQ})/(\mathrm{RDQ} - \mathrm{ODQ})$. Following (Mandi et al., 2022), percentage regret is computed as $\mathrm{Regret} = n_t^{-1} \sum_{i=1}^{n_t} [f(y_i, \widehat{z}(x_i)) - f(y_i, z^*(x_i))]/f(y, z^*(x_i))$. Higher nDQ and lower regret indicate better decision quality. Results are averaged over 10 independent runs and reported with standard deviations.

5.2 Main Results

Table 1 reports downstream performance on all three tasks. CWCE achieves the best nDQ and regret across all three tasks.

Table 1: Downstream decision performance on three stochastic discrete optimization tasks.

Loss Function	`Web Advertising`		`Portfolio Optimization`		`Diverse Recommendation`	
	nDQ ↑	Regret ↓	nDQ ↑	Regret ↓	nDQ ↑	Regret ↓
Quadratic	3.22 ± 1.99	80.03 ± 1.76	26.11 ± 4.91	129.72 ± 7.20	8.48 ± 1.32	59.31 ± 0.39
Pointwise	12.44 ± 11.47	71.15 ± 9.31	23.45 ± 0.00	133.23 ± 0.00	12.06 ± 5.09	57.44 ± 2.80
Listwise	24.30 ± 4.46	61.41 ± 3.90	35.96 ± 0.00	116.61 ± 0.00	24.70 ± 8.40	49.51 ± 5.16
PtO	$\underline{33.82 \pm 0.89}$	$\underline{53.88 \pm 0.36}$	14.12 ± 4.24	145.60 ± 4.26	28.34 ± 1.60	$\underline{46.87 \pm 1.05}$
Diffusion-DFL	10.08 ± 0.49	70.94 ± 0.68	43.87 ± 0.48	$\underline{101.81 \pm 0.35}$	$\underline{28.39 \pm 3.27}$	47.88 ± 2.10
Gen-DFL	31.91 ± 4.62	56.27 ± 3.71	$\underline{46.28 \pm 5.15}$	105.05 ± 8.65	28.11 ± 4.54	49.31 ± 5.31
CWCE (Ours)	$\mathbf{51.63 \pm 5.83}$	$\mathbf{39.66 \pm 4.68}$	$\mathbf{56.77 \pm 0.67}$	$\mathbf{83.28 \pm 1.36}$	$\mathbf{36.38 \pm 2.72}$	$\mathbf{42.31 \pm 1.85}$

The best and second-best results are marked in **bold** and underline, respectively.

CWCE also consistently outperforms PtO and deterministic DFL surrogates. The gains are especially pronounced on `Web Advertising` and `Diverse Recommendation`, where the combination of stochastic uncertainty and combinatorial decision structure makes distribution-aware decision learning particularly relevant.

Diffusion-DFL provides a stronger decision-aware generative baseline, yet requires continuous relaxation and projection on these discrete tasks. CWCE instead aligns the expected-cost profile directly over discrete candidate decisions, without backpropagating through or continuously relaxing the discrete $\arg\min$. Its consistent improvement over Diffusion-DFL suggests that explicitly designing the decision-alignment loss around discrete expected-cost ordering is beneficial.

Gen-DFL employs a contrastive loss that requires the optimal decision as supervision, which must be inferred from expected costs estimated using samples from the learned model. Particularly early in training, these estimates may be inaccurate, and in discrete decision spaces even small errors can alter the inferred optimal decision and yield unreliable supervision. In contrast, CWCE avoids optimal-decision supervision by directly targeting the ordering of expected costs across candidate decisions.

5.3 Decision Alignment with CWCE

To isolate the contribution of CWCE, we compare CWCE-augmented diffusion training (i.e., Eq. (7) with $\alpha > 0$) against standard diffusion training using only $\mathcal{L}_{\mathrm{DM}}$ (i.e., $\alpha = 0$), while keeping the diffusion architecture, distribution-matching component, sampling procedure, and downstream solver unchanged. Figure 2 reports the effect of α across the three tasks. Compared with DM-only training, incorporating CWCE generally improves decision quality over a range of $\alpha > 0$, as reflected by higher nDQ and lower regret. While this overall benefit is observed across tasks, the relative performance of different α values varies across datasets. We therefore select α for each dataset from $\{0.1, 0.5, 1, 5, 10\}$ based solely on validation regret.

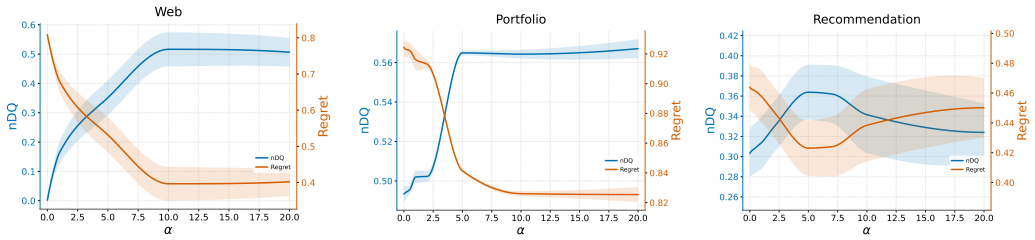

Figure 2: Effect of α on performance across the three tasks.

This comparison also provides empirical evidence consistent with Theorems 4.1 and 4.2. Distribution-only diffusion training optimizes distributional fidelity but does not explicitly account for downstream decision quality. CWCE complements this objective by incorporating downstream expected-cost information into training. The observed improvements from incorporating CWCE are consistent with our theoretical characterization that distributional fidelity alone need not ensure decision alignment, while jointly accounting for both can improve downstream decision quality.

5.4 Computational Efficiency

We report running times together with nDQ on three tasks in Figure 3 to evaluate the computational efficiency of CWCE alongside its downstream decision quality. Compared with the two generative baselines, Diffusion-DFL and Gen-DFL, CWCE achieves better decision quality with shorter running time, indicating a more favorable efficiency–performance trade-off among the generative approaches. Although deterministic methods are generally faster than all three generative approaches, including ours, they do not explicitly model the conditional outcome distribution and yield lower decision quality than CWCE. Overall, these results demonstrate a favorable trade-off between computational efficiency and downstream decision quality for CWCE.

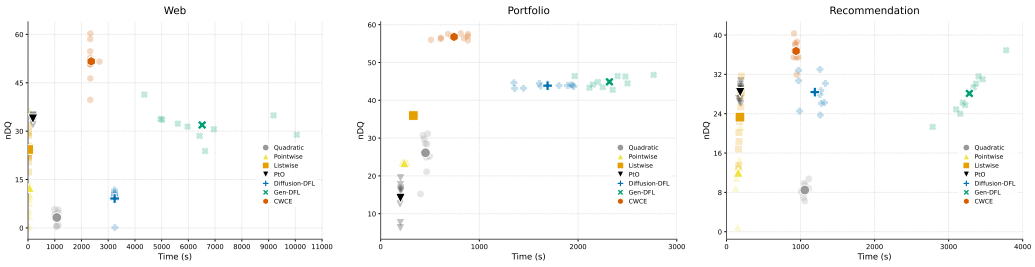

Figure 3: Efficiency Analysis on three tasks.

6 Conclusion

We study diffusion-based decision-focused learning for stochastic discrete optimization, where distributional fidelity alone does not necessarily guarantee downstream decision alignment. To address this limitation, we propose *Cost-Weighted Cross-Entropy* (*CWCE*), a differentiable objective that uses realized downstream costs to align the normalized expected-cost profile induced by the learned distribution, without differentiating through the discrete optimizer. We show that arbitrarily small distributional error can induce arbitrarily large decision regret, while incorporating CWCE into diffusion training yields a joint objective whose excess risk quantitatively controls downstream regret. Across three stochastic discrete optimization tasks, CWCE consistently improves decision quality over deterministic, ranking-based, and generative DFL baselines. Overall, CWCE complements generative distribution learning with explicit decision-relevant supervision.

AI USE STATEMENT

In this work, we used generative AI tools to assist with intermediate mathematical derivations and provide ingredients for proving mathematical claims, assist with the writing and verification of proofs, provide feedback on research methodology and experimental design, and assist with the interpretation and presentation of experimental results. We did not use generative AI tools to generate synthetic datasets, develop the theoretical models or conceptual framework of the proposed method, formulate mathematical claims, propose or refine research hypotheses, implement the proposed method, assist with translation, or clean or reformat datasets. Qualitative and thematic data analysis was not applicable to this work.

Additionally, we used generative AI tools for brainstorming, summarizing and discussing relevant literature, suggesting experimental parameters, creating or editing auxiliary software code, drafting and restructuring parts of the manuscript, improving the clarity and readability of the writing, suggesting paper structure and presentation, and assisting with reference formatting.

All AI-assisted content was critically reviewed and revised by the authors. In particular, all AI-assisted mathematical derivations and proofs were independently checked by the authors, experimental choices and interpretations were validated against the actual experimental results, and AI-assisted code was reviewed and tested before use. The research questions, theoretical and conceptual framework, mathematical claims, final methodological choices, theoretical results, and experimental conclusions were determined and verified by the authors. We take responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work develops a methodological framework for decision-focused learning in stochastic discrete optimization and does not involve the collection of new personal data. Our experiments use existing benchmark datasets for methodological evaluation. While the proposed approach may improve decision quality under uncertainty, biases or inaccuracies in the data and learned distribution may propagate to downstream decisions, and the specified decision objective may not capture fairness, safety, or other application-specific requirements. Real-world deployment, particularly in high-stakes settings, should therefore include appropriate validation, safeguards, and human oversight.

REPRODUCIBILITY STATEMENT

We have taken several steps to facilitate the reproducibility of this work. The complete CWCE training procedure, including diffusion pre-training, dynamic candidate-set construction, and CWCE-augmented fine-tuning, is provided in Appendix A and Algorithm 1. For the theoretical results, all assumptions and statements are given in Section 4, with complete proofs and supporting derivations provided in Appendix B. Detailed experimental information is provided in Appendix C, including the datasets and task formulations (Appendix C.1), compared methods (Appendix C.2), and implementation details, hyperparameters, and training configurations (Appendix C.3). Finally, the supplementary materials include anonymized source code and experiment configurations for reproducing the reported results.

REFERENCES

Azin Ashkan, Branislav Kveton, Shlomo Berkovsky, and Zheng Wen. Optimal greedy diversity for recommendation. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI'15, pp. 1742–1748. AAAI Press, 2015.

Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: A methodological tour d'horizon. *European Journal of Operational Research*, 290(2):405–421, 2021.

Dimitris Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044, March 2020.

Andrew M. Davis, Shawn Mankad, Charles J. Corbett, and Elena Katok. OM Forum—The best of both worlds: Machine learning and behavioral science in operations management. *Manufacturing & Service Operations Management*, 26(5):1605–1621, September 2024.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Proceedings of the 35th International Conference on Neural Information Processing Systems*, volume 34 of *NeurIPS '21*, pp. 8780–8794, Red Hook, NY, USA, 2021. Curran Associates, Inc.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Proceedings of the 31st International Conference on Neural Information Processing Systems*, volume 30 of *NeurIPS'17*, pp. 5490–5500, Red Hook, NY, USA, 2017. Curran Associates, Inc.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1): 9–26, January 2022.

F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TIIS)*, 5(4):1–19, December 2015.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Proceedings of the 34th International Conference on Neural Information Processing Systems*, volume 33 of *NeurIPS '20*, pp. 6840–6851, Red Hook, NY, USA, 2020. Curran Associates, Inc.

Hadi A. Khorshidi, Marzieh Soltanolkottabi, Richard Allmendinger, and Uwe Aickelin. Editorial: Special issue on operations research and machine learning. *INFOR: Information Systems and Operational Research*, 62(2):183–185, 2024.

Heyuan Liu and Paul Grigas. Risk bounds and calibration for a smart predict-then-optimize method. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Proceedings of the 35th International Conference on Neural Information Processing Systems*, volume 34 of *NeurIPS '21*, pp. 22083–22094, Red Hook, NY, USA, 2021. Curran Associates, Inc.

Jayanta Mandi, Emir Demirović, Peter J. Stuckey, and Tias Guns. Smart predict-and-optimize for hard combinatorial optimization problems. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, volume 34, pp. 1603–1610. AAAI Press, 2020.

Jayanta Mandi, Víctor Bucarey, Maxime Mulamba Ke Tchomba, and Tias Guns. Decision-focused learning: Through the lens of learning to rank. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 14935–14947. PMLR, 17–23 Jul 2022.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Víctor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 81:1623–1701, September 2024.

Quandl. WIKI various end-of-day data. `https://www.quandl.com/data/WIKI`, 2022.

Utsav Sadana, Abhilash Chenreddy, Erick Delage, Alexandre Forel, Emma Frejinger, and Thibaut Vidal. A survey of contextual optimization methods for decision-making under uncertainty. *European Journal of Operational Research*, 320(2):271–289, 2025.

Sanket Shah, Kai Wang, Bryan Wilder, Andrew Perrault, and Milind Tambe. Decision-focused learning without differentiable optimization: learning locally optimized decision losses. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Proceedings of the 36th International Conference on Neural Information Processing Systems*, volume 35 of *NeurIPS '22*, pp. 1320–1332, Red Hook, NY, USA, 2022. Curran Associates, Inc.

Sanket Shah, Bryan Wilder, Andrew Perrault, and Milind Tambe. Leaving the nest: going beyond local loss functions for predict-then-optimize. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, volume 38 of *AAAI '24*, pp. 14903–14909. AAAI Press, 2024.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, volume 32 of *NeurIPS '19*, pp. 11918–11930, Red Hook, NY, USA, 2019. Curran Associates, Inc.

Kai Wang, Bryan Wilder, Andrew Perrault, and Milind Tambe. Automatically learning compact quality-aware surrogates for optimization problems. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Proceedings of the 34th International Conference on Neural Information Processing Systems*, volume 33 of *NeurIPS '20*, pp. 9586–9596, Red Hook, NY, USA, 2020. Curran Associates, Inc.

Prince Zizhuang Wang, Shuyi Chen, Jinhao Liang, Ferdinando Fioretto, and Shixiang Zhu. Gen-DFL: Decision-focused generative learning for robust decision making. In *The Fourteenth International Conference on Learning Representations*, 2026.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI)*, volume 33, pp. 1658–1666. AAAI Press, 2019.

Yahoo! Webscope. Yahoo! Webscope dataset: ydata-ysm-advertiser-bids-v1.0. `https://webscope.sandbox.yahoo.com/`, 2007.

Zihao Zhao, Christopher Yeh, Lingkai Kong, and Kai Wang. Diffusion-DFL: Decision-focused diffusion models for stochastic optimization. In *The Fourteenth International Conference on Learning Representations*, 2026.

APPENDIX

This appendix provides the complete CWCE training procedure, theoretical proofs, and additional experimental details supporting the main results of the paper. Appendix A presents the practical CWCE training procedure, including diffusion pre-training, dynamic candidate-set construction, and CWCE-augmented fine-tuning, as summarized in Algorithm 1. Appendix B provides the complete proofs for the theoretical results in Section 4. Specifically, Appendix B.1 constructs a rare-event counterexample showing that arbitrarily small distributional error can induce an incorrect discrete decision with arbitrarily large regret, formally demonstrating that distributional fidelity alone does not guarantee decision alignment. Appendix B.2 proves that the excess risk of CWCE-augmented diffusion training quantitatively controls downstream decision regret, establishing that near-optimal joint training leads to near-optimal downstream decisions. Finally, Appendix C provides additional experimental details and analyses: Appendix C.1 describes the datasets and task formulations, Appendix C.2 details the compared methods, and Appendix C.3 reports implementation details and training configurations.

A CWCE TRAINING PROCEDURE

Algorithm 1 summarizes the practical optimization procedure. Following Section 3, we first pre-train the diffusion model for distribution learning and subsequently fine-tune it with CWCE. Because enumerating the full combinatorial decision space is intractable, CWCE is evaluated over a dynamically maintained candidate set $\mathcal{S} \subset \mathcal{Z}$.

Algorithm 1 CWCE-Augmented Diffusion Training

Pre-train $p_\theta(y \mid x)$ using $\mathcal{L}_{\mathrm{DM}}$.
Initialize the candidate set $\mathcal{S} \leftarrow \{z^*(y_i) : i = 1, \dots, n\}$, $z^*(y_i) \in \arg\min_{z \in \mathcal{Z}} f(y_i, z)$.
while not converged **do**
  for each mini-batch $\{(x_i, y_i)\}_{i=1}^{n_{\mathrm{batch}}}$ **do**
    for each (x_i, y_i) in the mini-batch **do**
      Draw M samples $\{y_i^{(m)}\}_{m=1}^{M} \sim p_\theta(\cdot \mid x_i)$.
      Estimate $\widehat{\mathcal{E}}_\theta(z; x_i) = \frac{1}{M} \sum_{m=1}^{M} f(y_i^{(m)}, z)$, $z \in \mathcal{S}$.
      Compute the normalized cost profile

$$\widehat{\pi}_\theta(z \mid x_i) = \frac{\widehat{\mathcal{E}}_\theta(z; x_i)}{\sum_{z' \in \mathcal{S}} \widehat{\mathcal{E}}_\theta(z'; x_i)}.$$

      Compute $\mathcal{L}_{\mathrm{CW}} = -\sum_{z \in \mathcal{S}} f(y_i, z) \log \widehat{\pi}_\theta(z \mid x_i)$.
      With probability $p_{\mathrm{solve}} = 0.1$, compute

$$\widehat{z}(x_i) \in \arg\min_{z \in \mathcal{Z}} \frac{1}{M} \sum_{m=1}^{M} f(y_i^{(m)}, z).$$

      Update $\mathcal{S} \leftarrow \mathcal{S} \cup \{\widehat{z}(x_i)\}$.
    end for
    Compute the CWCE-augmented training objective

$$\mathcal{L} = \alpha \cdot \frac{\mathrm{sg}\left(\|\nabla_\theta \mathcal{L}_{\mathrm{DM}}\|_2\right)}{\mathrm{sg}\left(\|\nabla_\theta \mathcal{L}_{\mathrm{CW}}\|_2\right) + \epsilon} \cdot \mathcal{L}_{\mathrm{CW}} + \mathcal{L}_{\mathrm{DM}}.$$

    Update $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$.
  end for
end while

The gradient-normalization coefficient is detached from the computational graph and serves only to balance the scales of the diffusion and CWCE training signals.

B PROOFS RELATED TO SECTION 4

This section provides the complete proofs for the theoretical results in Section 4. We first construct a rare-event counterexample showing that arbitrarily small distributional error can induce arbitrarily large downstream decision regret. We then prove that the excess risk of the CWCE-augmented joint objective quantitatively controls downstream decision regret.

B.1 PROOFS RELATED TO SECTION 4.1

We prove Theorem 4.1 through a binary decision problem in which an arbitrarily small perturbation of a rare but decision-critical event reverses the conditional expected-cost ordering. The construction makes explicit why statistical proximity between outcome distributions does not, in general, control downstream discrete decisions.

Construction. Consider a binary decision set $\mathcal{Z} = \{z_1, z_2\}$ and a measurable event $A \subseteq \mathcal{Y}$. For constants $m > 0$ and $\varepsilon > 0$, define the nonnegative cost function

$$f(y, z_1) = m, \tag{A.1}$$

$$f(y, z_2) = \begin{cases} 0, & y \in A, \\ m + \varepsilon, & y \notin A. \end{cases} \tag{A.2}$$

Thus, z_2 is favorable when the rare event A occurs, whereas its cost is slightly larger than that of z_1 outside A.

Let $r > 0$ denote a prescribed regret level. Choose

$$\pi := \frac{r + \varepsilon}{m + \varepsilon}, \qquad \Delta := \frac{r + \eta}{m + \varepsilon}, \qquad 0 < \eta < \varepsilon, \tag{A.3}$$

and define

$$q(A \mid x) = \pi, \qquad p(A \mid x) = \pi - \Delta. \tag{A.4}$$

For sufficiently large m, $0 < \pi < 1$, and $\pi - \Delta = (\varepsilon - \eta)/(m + \varepsilon) > 0$, so both conditional distributions are well defined.

Under the true distribution q, the two conditional expected costs are

$$\mathcal{E}_q(z_1; x) = m, \qquad \mathcal{E}_q(z_2; x) = (1 - \pi)(m + \varepsilon) = m - r. \tag{A.5}$$

Hence $\mathcal{E}_q(z_2; x) < \mathcal{E}_q(z_1; x)$, and the Bayes-optimal decision is $z_q^*(x) = z_2$.

Under the learned distribution p,

$$\mathcal{E}_p(z_1; x) = m, \qquad \mathcal{E}_p(z_2; x) = (1 - \pi + \Delta)(m + \varepsilon) = m + \eta. \tag{A.6}$$

Therefore, $\mathcal{E}_p(z_1; x) < \mathcal{E}_p(z_2; x)$, so the learned distribution induces $z_p^*(x) = z_1 \neq z_2 = z_q^*(x)$. The perturbation has therefore reversed the expected-cost ordering and changed the optimal discrete decision.

The regret of the decision induced by p, evaluated under the true distribution, is

$$\mathrm{Regret}_q(p) = \mathcal{E}_q(z_p^*(x); x) - \mathcal{E}_q(z_q^*(x); x) = m - (m - r) = r. \tag{A.7}$$

It remains to show that the corresponding distributional error can be made arbitrarily small. Because q and p differ only in the probability assigned to A, their KL divergence is the Bernoulli KL divergence

$$D_{\mathrm{KL}}(q \,\|\, p) = \pi \cdot \log\left(\frac{\pi}{\pi - \Delta}\right) + (1 - \pi) \cdot \log\left(\frac{1 - \pi}{1 - \pi + \Delta}\right).$$

From Eq. (A.3), $\pi = (r + \varepsilon)/(m + \varepsilon)$, $\pi - \Delta = (\varepsilon - \eta)/(m + \varepsilon)$. Hence the first term satisfies

$$\pi \log\left(\frac{\pi}{\pi - \Delta}\right) = \frac{r + \varepsilon}{m + \varepsilon} \cdot \log\left(\frac{r + \varepsilon}{\varepsilon - \eta}\right) = O(m^{-1}).$$

Moreover, $\pi = O(m^{-1})$ and $\Delta = O(m^{-1})$, and therefore

$$(1-\pi)\cdot\log\left(\frac{1-\pi}{1-\pi+\Delta}\right) = O(m^{-1}).$$

Consequently,

$$D_{\mathrm{KL}}(q\,\|\,p) = O(m^{-1}). \tag{A.8}$$

This implies that for any $\delta > 0$, we can choose an $m > 0$ such that $D_{\mathrm{KL}}(q\,\|\,p) < \delta$. The construction therefore exhibits a rare but decision-critical event whose probability perturbation becomes arbitrarily small while its downstream impact remains fixed. The key mechanism is not large global distributional error, but a change in the expected-cost ordering that determines the discrete decision.

Proof of Theorem 4.1. Fix arbitrary $\delta > 0$ and $R > 0$. Choose any $r > R$, and fix constants $\varepsilon > 0$ and $\eta \in (0, \varepsilon)$. Consider the construction in Eq. (A.1)–Eq. (A.4).

By Eq. (A.5) and Eq. (A.6), the true and learned expected-cost orderings satisfy

$$\mathcal{E}_q(z_2; x) < \mathcal{E}_q(z_1; x), \qquad \mathcal{E}_p(z_1; x) < \mathcal{E}_p(z_2; x).$$

Thus,

$$\arg\min_{z\in\mathcal{Z}} \mathbb{E}_{y\sim p}[f(y,z)] \neq \arg\min_{z\in\mathcal{Z}} \mathbb{E}_{y\sim q}[f(y,z)].$$

By Eq. (A.7), $\mathrm{Regret}_q(p) = r > R$. Finally, Eq. (A.8) implies that there exists a sufficiently large m such that $D_{\mathrm{KL}}(q\,\|\,p) < \delta$. Therefore, for arbitrary $\delta > 0$ and $R > 0$, there exist q, p, $\mathcal{Z}$, and a nonnegative cost function f such that the learned distribution is arbitrarily close to the true distribution in KL divergence, yet induces an incorrect discrete decision with regret exceeding R. □

B.2 Proofs Related to Section 4.2

Proof of Theorem 4.2. For each input x, define the normalized expected-cost profiles

$$\pi_q(z \mid x) := \frac{\mathcal{E}_q(z;x)}{S_q(x)}, \qquad \pi_p(z \mid x) := \frac{\mathcal{E}_p(z;x)}{S_p(x)},$$

where

$$\mathcal{E}_p(z;x) := \mathbb{E}_{Y\sim p(\cdot|x)}[f(Y,z)], \qquad S_p(x) := \sum_{z\in\mathcal{Z}} \mathcal{E}_p(z;x).$$

Step 1: Joint excess risk controls normalized cost-profile divergence. For a fixed input x, the conditional population CWCE risk satisfies

$$\mathcal{L}_{\mathrm{CW}}(p;x) = -\sum_{z\in\mathcal{Z}} \mathcal{E}_q(z;x)\log\pi_p(z\mid x) = S_q(x)\left[H(\pi_q(\cdot\mid x)) + D_{\mathrm{KL}}\left(\pi_q(\cdot\mid x)\,\|\,\pi_p(\cdot\mid x)\right)\right].$$

At $p = q$, $\mathcal{L}_{\mathrm{CW}}(q;x) = S_q(x)H(\pi_q(\cdot\mid x))$. Therefore,

$$\mathcal{L}_{\mathrm{CW}}(p;x) - \mathcal{L}_{\mathrm{CW}}(q;x) = S_q(x) D_{\mathrm{KL}}\left(\pi_q(\cdot\mid x)\,\|\,\pi_p(\cdot\mid x)\right).$$

Taking expectation over x gives

$$\mathcal{L}_{\mathrm{CW}}(p) - \mathcal{L}_{\mathrm{CW}}(q) = \mathbb{E}_x\left[S_q(x) D_{\mathrm{KL}}\left(\pi_q(\cdot\mid x)\,\|\,\pi_p(\cdot\mid x)\right)\right].$$

By the conditional-mean decomposition of the population diffusion denoising objective,

$$\mathcal{L}_{\mathrm{DM}}(p) - \mathcal{L}_{\mathrm{DM}}(q) = \mathbb{E}\left[\|\epsilon_p(Y_t,t,X) - \epsilon_q(Y_t,t,X)\|_2^2\right] \geq 0,$$

where $\epsilon_p(Y_t,t,X)$ and $\epsilon_q(Y_t,t,X)$ denote the population-optimal denoisers induced by p and q, respectively. Hence,

$$\mathcal{L}_{\mathrm{joint}}(p) - \mathcal{L}_{\mathrm{joint}}(q) = \mathcal{L}_{\mathrm{DM}}(p) - \mathcal{L}_{\mathrm{DM}}(q) + \lambda\left[\mathcal{L}_{\mathrm{CW}}(p) - \mathcal{L}_{\mathrm{CW}}(q)\right] \geq \lambda\left[\mathcal{L}_{\mathrm{CW}}(p) - \mathcal{L}_{\mathrm{CW}}(q)\right].$$

Combining this inequality with the assumed joint excess risk bound yields

$$\mathbb{E}_x \left[S_q(x) D_{\mathrm{KL}} \left(\pi_q(\cdot \mid x) \,\|\, \pi_p(\cdot \mid x) \right) \right] \leq \epsilon / \lambda. \tag{A.9}$$

Step 2: Cost-profile divergence controls conditional decision regret. Let

$$z^*(x) \in \arg\min_{z \in \mathcal{Z}} \mathcal{E}_q(z; x), \qquad z_p(x) \in \arg\min_{z \in \mathcal{Z}} \mathcal{E}_p(z; x).$$

Since normalization is by a common positive scalar, $z^*(x)$ and $z_p(x)$ equivalently minimize $\pi_q(\cdot \mid x)$ and $\pi_p(\cdot \mid x)$, respectively. Hence, $\pi_p(z_p(x) \mid x) \leq \pi_p(z^*(x) \mid x)$. It follows that

$$\begin{aligned}
&\pi_q(z_p(x) \mid x) - \pi_q(z^*(x) \mid x) \\
&= \pi_q(z_p \mid x) - \pi_p(z_p \mid x) + \pi_p(z_p \mid x) - \pi_p(z^* \mid x) + \pi_p(z^* \mid x) - \pi_q(z^* \mid x) \\
&\leq |\pi_q(z_p \mid x) - \pi_p(z_p \mid x)| + |\pi_p(z^* \mid x) - \pi_q(z^* \mid x)| \\
&\leq \|\pi_q(\cdot \mid x) - \pi_p(\cdot \mid x)\|_1 .
\end{aligned}$$

Since $\mathcal{E}_q(z; x) = S_q(x) \pi_q(z \mid x)$, the conditional decision regret satisfies

$$\mathrm{Regret}_q(p; x) := \mathcal{E}_q(z_p(x); x) - \mathcal{E}_q(z^*(x); x) \leq S_q(x) \|\pi_q(\cdot \mid x) - \pi_p(\cdot \mid x)\|_1 .$$

By Pinsker's inequality,

$$\|\pi_q(\cdot \mid x) - \pi_p(\cdot \mid x)\|_1 \leq \sqrt{2 D_{\mathrm{KL}} \left(\pi_q(\cdot \mid x) \,\|\, \pi_p(\cdot \mid x) \right)}.$$

Therefore,

$$\mathrm{Regret}_q(p; x) \leq S_q(x) \sqrt{2 D_{\mathrm{KL}} \left(\pi_q(\cdot \mid x) \,\|\, \pi_p(\cdot \mid x) \right)}. \tag{A.10}$$

Step 3: Population regret is controlled by joint excess risk. Taking expectation over x in Eq. (A.10) and applying Cauchy–Schwarz gives

$$\mathrm{Regret}_q(p) := \mathbb{E}_x \left[\mathrm{Regret}_q(p; x) \right] \leq \sqrt{2 \, \mathbb{E}_x[S_q(x)] \, \mathbb{E}_x \left[S_q(x) D_{\mathrm{KL}} \left(\pi_q(\cdot \mid x) \,\|\, \pi_p(\cdot \mid x) \right) \right]}.$$

The assumption $\mathbb{E}_x[S_q(x)] \leq B$ together with Eq. (A.9) yields

$$\mathrm{Regret}_q(p) \leq \sqrt{(2B/\lambda) \cdot \epsilon}.$$

This completes the proof. □

C Additional Experimental Details and Analysis

This section provides additional experimental details and analyses supporting the results in Section 5. We first describe the three stochastic discrete optimization tasks and their datasets, then provide details of the compared methods and implementation settings, and finally present additional computational efficiency analysis across all three tasks.

C.1 Datasets

Throughout this section, x denotes the input features, y the ground-truth uncertain outcome, and $z \in \{0, 1\}^m$ a binary decision vector satisfying the cardinality constraint $\sum_{i=1}^m z_i = b$. We consider the same three stochastic combinatorial optimization tasks as in Section 5: `Web Advertising`, `Portfolio Optimization`, and `Diverse Recommendation`.

Web Advertising. Following (Wilder et al., 2019), this benchmark considers a cardinality-constrained advertising allocation problem with $m = 100$ candidate websites, $K = 500$ users, and budget $b = 5$. For website i and user j, $y_{ij} \in [0, 1]$ denotes the ground-truth click-through rate (CTR), and $y_i = (y_{ij})_{j=1}^K \in [0, 1]^K$ denotes the corresponding CTR vector. The outcomes are obtained from the Yahoo! Webscope advertising dataset (Yahoo! Webscope, 2007). Following the benchmark protocol, the input features are generated from the underlying CTRs through a random linear transformation, $x_i = A y_i$, where $A \in \mathbb{R}^{K \times K}$ is a random matrix.

The downstream optimization problem selects b websites to maximize unique user reach:

$$\min_{z\in\{0,1\}^m} \quad \sum_{j=1}^{K}\left(\prod_{i=1}^{m}(1-z_i y_{ij})\right) \qquad \text{s.t. } \sum_{i=1}^{m} z_i = b. \tag{A.11}$$

The multiplicative coverage term makes the objective nonlinear in the uncertain CTRs, providing a natural setting for evaluating distribution-aware decision learning.

Portfolio Optimization. Following stochastic portfolio selection (Donti et al., 2017; Wang et al., 2020), this benchmark contains $m = 50$ candidate assets with budget $b = 3$. For asset i, $x_i \in \mathbb{R}^F$ denotes a feature vector constructed from historical market indicators, and $y_i \in \mathbb{R}$ denotes its future return. The outcome vector is $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$. Features and returns are constructed from the Quandl WIKI dataset (Quandl, 2022).

Let $\Sigma \in \mathbb{R}^{m\times m}$ denote the covariance matrix estimated from the training returns. The downstream optimization problem selects b assets by maximizing a mean–variance objective:

$$\min_{z\in\{0,1\}^m} \quad -y^\top z + \beta z^\top \Sigma z. \qquad \text{s.t. } \sum_{i=1}^{m} z_i = b, \tag{A.12}$$

where $\beta = 0.1$ is the risk-aversion parameter. To ensure positive expected costs, we subtract from all costs the minimum realized cost over the training set and candidate decision space, $\min_{(x,y)\in D,\, z\in\mathcal{Z}} f(y, z)$, and add a small constant 10^{-5}, where D denotes the training set; this constant shift does not affect the ordering of candidate decisions.

The data are indexed chronologically by trading day, with each day defining one decision instance consisting of the vectorized features x and future-return vector y. The covariance matrix is estimated using only the training data and remains fixed during validation and testing.

Diverse Recommendation. Following (Ashkan et al., 2015), this benchmark considers a topic-coverage problem with $m = 100$ candidate movies, $K = 500$ topics, and recommendation budget $b = 20$. Topics correspond to actors, and $y_{ij} \in \{0, 1\}$ indicates whether movie i is associated with topic j. The ground-truth outcome for movie i is therefore $y_i = (y_{ij})_{j=1}^{K} \in \{0, 1\}^K$. The data are constructed from the MovieLens benchmark (Harper & Konstan, 2015). Each movie is represented by a 2113-dimensional user-rating feature vector $x_i \in \mathbb{R}^{2113}$, with missing entries filled with zeros.

The downstream problem selects b movies to maximize the number of distinct topics covered:

$$\min_{z\in\{0,1\}^m} \quad \sum_{j=1}^{K}\mathbf{1}\left(\sum_{i=1}^{m} z_i y_{ij} = 0\right) \qquad \text{s.t. } \sum_{i=1}^{m} z_i = b. \tag{A.13}$$

The benchmark contains multiple independent decision instances, each consisting of exactly m movies and an $m \times K$ binary outcome matrix. Training, validation, and test partitions are defined at the instance level. The predictive model operates on per-movie features, while the downstream optimizer operates jointly on the complete instance-level outcome matrix.

C.2 Compared Methods

We provide additional details on the deterministic and diffusion-based baselines used in Section 5.

Pointwise Ranking. The pointwise ranking surrogate (Mandi et al., 2022) regresses predicted decision costs toward their realized counterparts over a candidate decision set. It is applied to deterministic predictions and provides a differentiable alternative to directly differentiating through the downstream discrete optimizer.

Listwise Ranking. The listwise ranking surrogate (Mandi et al., 2022) constructs a soft ranking distribution over candidate decisions from their objective values and minimizes a cross-entropy loss between the predicted and target rankings. Unlike CWCE, this surrogate operates on the ranking induced by a deterministic point prediction rather than the conditional expected-cost profile induced by a predictive distribution.

Quadratic. Following the locally optimized decision-loss construction of (Shah et al., 2022), we use a convex quadratic surrogate in prediction space: $L_{\mathrm{quad}}(\widehat{y}, y) = (\widehat{y} - y)^\top H(\widehat{y} - y)$, where $H \succeq 0$ is parameterized as $H = L^\top L$ for tractability. Given local perturbations $\widehat{y}^{(1)}, \ldots, \widehat{y}^{(K)}$ and their corresponding downstream decision losses $\mathrm{DL}\big(y, \widehat{y}^{(k)}\big) := f\big(y, z^*(\widehat{y}^{(k)})\big)$, the quadratic form is fitted to approximate the local decision-loss landscape and is subsequently used as a differentiable surrogate during predictive model training.

Predict-then-optimize (PtO). We first train a neural network for outcome prediction using a least-squares loss, and then infer the decision by minimizing the downstream cost under the predicted outcome.

Diffusion-DFL. Diffusion-DFL (Zhao et al., 2026) is formulated for continuous decision spaces. For our discrete tasks, we therefore relax $\mathcal{Z} = \{z \in \{0,1\}^m : \sum_{i=1}^m z_i = b\}$ to $\widetilde{\mathcal{Z}} = \{z \in [0,1]^m : \sum_{i=1}^m z_i = b\}$. Diffusion-DFL is applied to the resulting continuous problem to obtain a solution $\widetilde{z}(x)$. We then project it back to the discrete feasible set using a top-b operation: the b largest components of $\widetilde{z}(x)$ are set to 1, and all remaining components are set to 0. This yields the feasible discrete decision $\widehat{z}(x)$ used for evaluation.

Gen-DFL. Gen-DFL (Wang et al., 2026) is a robust generative DFL baseline originally developed for minimizing conditional value-at-risk (CVaR), which reduces to the conditional expected cost when the tail probability is set to one. Its training objective combines a distribution-matching loss with a decision-focused contrastive loss. However, the true decision labels required by the contrastive loss are unavailable from standard supervised data and must therefore be reconstructed from the learned conditional distribution.

C.3 Implementation Details

All experiments are implemented in PyTorch and run on a server equipped with dual Intel Xeon Platinum 8358 CPUs (128 logical cores in total) and one NVIDIA A100 GPU with 80GB memory.

Across all tasks, we use a conditional diffusion model with 1000 diffusion steps and a linear beta schedule. The denoising network is a multilayer perceptron with three hidden layers of width 1024, SiLU activations, and LayerNorm. The timestep embedding dimension is 16, followed by two linear projection layers.

We first pre-train the diffusion model using $\mathcal{L}_{\mathrm{DM}}$ for 1500 epochs using Adam optimizer. The pre-trained model is subsequently fine-tuned using the CWCE-augmented objective in Eq. (7). Task-specific configurations are summarized in Table 2.

Table 2: Dataset statistics and task-specific training configurations.

Configuration	`Web Advertising`	`Portfolio Optimization`	`Diverse Recommendation`
Per-instance data	100 websites × 500 users	50 assets	100 movies × 500 topics
Feature dimension	500	28	2113
Budget b	5	3	20
Fine-tuning epochs	60	200	200
Learning rate	10^{-3}	10^{-3}	5×10^{-4}
Weight decay	10^{-4}	10^{-4}	10^{-4}
Batch size	8	32	4
Monte Carlo samples	20	80	80
DDIM sampling steps	50	100	10

All compared methods use identical data partitions and preprocessing whenever applicable. For `Web Advertising` and `Portfolio Optimization`, we follow the benchmark protocols of (Shah et al., 2022). For `Diverse Recommendation`, we follow (Wilder et al., 2019). Specifically, `Web Advertising` contains $240/60/500$ training/validation/test instances, while `Portfolio Optimization` contains $200/200/400$ instances. For MovieLens recommendation, we use the official training/test partition of (Wilder et al., 2019), split $80\%/20\%$ into training and validation, with the official test partition reserved exclusively for testing. We report results over 10 independent runs with different training seeds, using the same data split across methods within each run.