

BAYESIAN STOCHASTIC GRADIENT DESCENT FOR
STOCHASTIC OPTIMIZATION WITH STREAMING INPUT DATA*TIANYI LIU[†], YIFAN LIN[†], AND ENLU ZHOU[†]

Abstract. We consider stochastic optimization under distributional uncertainty, where the unknown distributional parameter is estimated from streaming data that arrive sequentially over time. Moreover, data may depend on the decision at the time when they are generated. For both decision-independent and decision-dependent uncertainties, we propose an approach to jointly estimate the distributional parameter via Bayesian posterior distribution and update the decision by applying stochastic gradient descent (SGD) on the Bayesian average of the objective function. Our approach converges asymptotically over time and achieves the convergence rates of classical SGD in the decision-independent case. We demonstrate the empirical performance of our approach on both synthetic test problems and a classical newsvendor problem.

Key words. Bayesian estimation, streaming input data, stochastic gradient descent, endogenous uncertainty

MSC code. 90C15

DOI. 10.1137/22M1478951

1. Introduction. Stochastic optimization is a mathematical framework that models decision making under uncertainty. It usually assumes that the decision maker has full knowledge about the underlying uncertainty through a known probability distribution and minimizes (or maximizes) a functional of the cost (or reward) function [55]. However, the probability distribution of the randomness in the system is rarely known in practice and is often estimated from historic data. The impact of the estimation accuracy and the subsequent distributional uncertainty have been widely studied in the literature. For example, [9] and [51] conduct perturbation analysis of the stochastic optimization problems and quantify the sensitivity of the optimal value (and/or solution) to the probability distribution. One popular approach to addressing this distributional uncertainty in stochastic optimization is distributionally robust optimization (DRO) (e.g., [14, 7, 61]). The DRO framework assumes that the underlying unknown probability distribution lies in an ambiguity set of probability distributions and then optimizes the problem with respect to the worst case in the ambiguity set. It has been successfully applied to a broad range of problems in statistics, optimization, and control, such as stochastic programming (e.g., [4, 37]), Markov Decision Processes (e.g., [67, 68]), stochastic control (e.g., [58, 69]), and ranking and selection (e.g., [27, 66, 65, 25]). To construct an appropriate ambiguity set that contains the true distribution with a probabilistic guarantee and ensures tractability of the optimization problem, various DRO methods have been developed, such as methods based on moment constraints (e.g., [14]), ϕ -divergence (e.g., [5]), and Wasser-

* Received by the editors February 17, 2022; accepted for publication (in revised form) October 13, 2023; published electronically January 25, 2024. A preliminary version of this paper appeared in Proceedings of the 2021 Winter Simulation Conference, IEEE, Piscataway, NJ, 2021. Tianyi Liu and Yifan Lin contributed equally to this work.

<https://doi.org/10.1137/22M1478951>

Funding: This research is funded by the Air Force Office of Scientific Research under grants FA9550-19-1-0283, FA9550-22-1-0244, and National Science Foundation under grant DMS2053489.

[†]H. Milton School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (tliu341@gatech.edu, ylin429@gatech.edu, enlu.zhou@isye.gatech.edu).

stein distance (e.g., [24]). In contrast to DRO, [72, 64] proposed a Bayesian risk optimization (BRO) framework, with the motivation to use the Bayesian posterior distribution (which encodes the likelihoods of all possibilities) to replace the ambiguity set (which treats every possibility inside the set with equal probability), and further take a risk functional with respect to the posterior distribution to allow a more flexible risk attitude.

Nearly all the aforementioned works that focus on stochastic optimization in static settings assume that the input data are given as one fixed batch. However, in many applications, data are often collected over time, and the decision maker often needs to make decisions in an online fashion given all the available data. For example, an inventory manager observes the customer demand on a daily or weekly basis, and adjusts his/her decision accordingly; a robot that searches for an unknown source receives signals from the source over time, and makes its move accordingly (e.g., [43]). Such streaming data have only been considered recently in stochastic simulation optimization, e.g., [63, 71, 62, 57]. While these recent works consider the streaming input data, their assumption is that the data are generated from an exogenous (decision-independent) distribution and hence are independent and identically distributed (i.i.d.). This assumption restricts their application to many real-world problems where the input data are endogenous (decision-dependent). For example, in live streaming e-commerce, there is usually a rolling banner that counts how many products are left, and customers are more likely to purchase the product that has only a few left since it is more popular. As another example, in the supermarket, tall stacks of a product impact its visibility, which leads more customers to purchase the product [30, 3].

Motivated by these real-world problems where data arrive sequentially and could even depend on the decision, in this work we consider stochastic optimization problems where the underlying distribution is unknown but data from the distribution arrive in batches over time. We assume a parameterized distributional model, and thus the distribution family is known but the true distributional parameter is unknown. It is also interesting to consider a nonparametric setting with a prior of Dirichlet process (see [59] for a nonparametric simulation optimization problem setting), though the associated analysis could be much more complicated. At each time stage, our procedure consists of two steps: (1) use the current batch of data to update the Bayesian posterior distribution of the distributional parameter, and (2) take the Bayesian average of the objective function and apply stochastic gradient descent (SGD) to this reformulated objective function. Our proposed approach can be viewed as an online extension of the BRO framework in [64]: BRO considers a fixed batch of data and only needs to solve the fixed BRO formulation; in contrast, we consider the setting where batches of data come in sequentially, and therefore, we update the stagewise BRO problem every time with the new incoming data; moreover, due to the limited time in each stage, we can only apply a few SGD iterations to solve each stagewise BRO problem. As a result, the convergence analyses of BRO and our paper are quite different and the results have distinct implications: the convergence of BRO shows that if the fixed batch of data has an infinite size, the BRO formulation recovers the true problem and BRO solutions are indeed the true optimal solutions; our convergence analysis shows that even though our algorithm applies SGD iterations to a sequence of estimated (Bayesian-average) problems, but the algorithm still converges to the true (local) optimal solution. Another related work [56] considers the same problem of fixed data batch as [72, 64] and uses Bayesian average to estimate the true problem,

but it also takes a robust approach with respect to the uncertainty associated with the parametric distributional model.

We consider both cases of exogenous and endogenous input data. In the former case, data follow a fixed distribution that only involves the distributional parameter. In the latter case, the data follow a time-varying distribution depending not only on the distributional parameter but also on the decision at the current time. It is worth noting that due to the correlation and nonstationarity of the decision-dependent data across time stages, the Bayesian estimation with such data is different from the classical Bayesian updating with i.i.d. data, which poses a great challenge to showing the consistency of the Bayesian posterior distribution. We consider the same problem as [57], but differ in two key aspects: first, we take a Bayesian approach to estimate the distributional parameter, whereas they estimate by maximum likelihood estimator (MLE) and solve the problem with the plug-in MLE; second, they only consider exogenous (decision-independent) uncertainty. Also note that compared to our preliminary conference version [44], this paper is a substantial extension in both theoretical analysis and numerical experiments. For the decision-independent uncertainty, we further show the convergence rate of the proposed algorithm. Apart from a synthetic test problem, we also evaluate the performance of the proposed algorithm in a classical newsvendor problem.

Our considered problem is related to online learning (e.g., [10, 53]). Online learning is often formulated as a repeated game: at each round, the learner makes a prediction and receives the true solution (or a cost function), with the goal to minimize the cumulative cost over time. Classical algorithms in online learning such as follow the leader and its variants, such as follow the perturbed Leader and follow the regularized leader, incorporate the learning process, which takes the information from previous rounds to improve prediction, into the algorithms in order to choose the next action that leads to the lowest cumulative cost. In contrast to the goal of minimizing the cumulative cost, our considered problem aims to find an optimal solution of a stationary objective function in the decision-independent case and a nonstationary objective function in the decision-dependent case, where the nonstationarity is only caused by the decision-dependent uncertainty. Since the online data in our problem are restricted to the randomness in the system that is generated from the (unknown) underlying distribution, it is natural to update our belief of the (unknown) distribution in a Bayesian way. In addition to the distinctive goal in our problem, it is worth noting the key differences between our approach and two closely related algorithms in online learning. The first one is the online gradient descent algorithm (see [73, 34, 16]), for which the cost function can vary completely arbitrarily over time and, hence, is unlike our SGD algorithm that makes use of the structure of the Bayesian average of the objective function over time. The second one is the Thompson sampling algorithm (see [1, 11]), which also assumes a parameterized model and updates the posterior distribution on the parameter in a Bayesian way. However, Thompson sampling makes the decision based on only one sample from the posterior distribution in each round, whereas our algorithms takes the entire posterior distribution into account and solve the Bayesian average of the original (unknown) objective function. Later in the numerical experiments, we show that the Bayesian average provides a better estimate of the original objective function compared to a point estimate.

As a final note, the endogenous uncertainty has been considered in many fields, including dynamic programming (e.g., [60]), robust optimization (e.g., [48, 40]), and stochastic optimization (e.g., [31, 17, 22, 35, 49, 45, 70]), with many applications in inventory control (e.g., [6, 41]), healthcare (e.g., [32]), and so on. However, almost none of the aforementioned work involving decision-dependent uncertainty take into consideration the additional input data. Only until recently, [38] and [46] study the performative prediction problem, which is essentially a stochastic optimization problem with streaming decision-dependent data; however, the goal is to find the so-called performatively stable point (or equilibrium point), which is in general different from the true optimal solution. Along the same line, [15] also considers static stochastic optimization under decision-dependent uncertainty, and proposes a proximal gradient method and its variants that converge to the performatively stable point under relatively strong assumptions (strong convexity, Lipschitz continuity, etc.). Asymptotic normality and optimality of the stochastic approximation algorithm are further studied in a follow-up work [13]. Most recently, [36] and [47] redesign the gradient algorithms in [38] by introducing a gradient correction term, and show the convergence to the true optimal solution. In particular, [36] also considers a parameterized model where the distributional parameter (as a function of the decision variable) can be estimated from streaming input data, and uses finite difference to estimate the gradient of the objective function. An important assumption in their approach is that the estimated distributional parameter has a constant error bound. Differently from their approach, we learn the distributional parameter with a Bayesian approach, and show the Bayesian consistency of the posterior distribution that finally leads to the convergence of the SGD algorithm to a stationary point of the original objective function (optimal solution if the problem is convex).

We summarize the contribution of this paper as follows. First, we propose a Bayesian stochastic gradient descent approach to the stochastic optimization problem with unknown underlying distribution and with streaming input data that could depend on the decision. This new approach is among the very few works [64, 56, 33] in the literature that take a Bayesian perspective on approaching distributional uncertainty in stochastic optimization. Second, we show the convergence of our approach in the decision-independent case and decision-dependent case, respectively. Under decision-independent uncertainty, our approach achieves the convergence rates of classical nonconvex SGD. Third, we show the consistency of the Bayesian posterior distribution with endogenous non-i.i.d. data under mild conditions; this result is applicable to a wide range of problems involving Bayesian estimation beyond the scope of this paper. Our nonasymptotic analysis of the Bayesian estimate with i.i.d. data is also new and could be potentially useful for analyzing other Bayesian algorithms.

The rest of the paper is organized as follows. We first propose Bayesian-SGD algorithms for stochastic optimization with decision-independent and decision-dependent streaming input data in section 2. We then analyze the convergence properties of the proposed algorithms for both cases in section 3. We verify the theoretical results and demonstrate the performance of our algorithms in the numerical experiments in section 4. Finally, we conclude the paper in section 5.

2. Bayesian SGD algorithms for stochastic optimization with streaming input data. We consider the following stochastic optimization problems with decision-independent uncertainty and decision-dependent uncertainty, respectively:

$$(2.1) \quad \min_{x \in \mathcal{X}} H(x) := \mathbb{E}_{f(\cdot; \theta^c)}[h(x, \xi)] \quad (\text{decision-independent uncertainty}),$$

$$(2.2) \quad \min_{x \in \mathcal{X}} H(x) := \mathbb{E}_{f(\cdot; x, \theta^c)}[h(x, \xi)] \quad (\text{decision-dependent uncertainty}),$$

where $x \in \mathcal{X} \subset \mathbb{R}^d$ is the decision vector, $\xi \in \Xi \subset \mathbb{R}^m$ is a random vector, $h : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a deterministic function. The expectation is taken with respect to (w.r.t.) the distribution of ξ , which is denoted by $f(\cdot; \theta^c)$ in the decision-independent case, and by $f(\cdot; x, \theta^c)$ in the decision-dependent case. The density function $f(\cdot; x, \theta^c)$ takes a general form, where the parameter θ^c does not depend on x . For example, $f(\xi; x, \theta^c) = \theta^c x \exp(-\theta^c x \xi)$ is the density function of the exponential distribution with rate $\theta^c x$. More assumptions on the density function will be discussed in section 3. We assume the distribution of ξ belongs to a parameterized family of distributions with parameter set $\Theta \subset \mathbb{R}^l$, and let θ^c be the true parameter value of the distribution.

In practice, the true distribution $f(\cdot; \theta^c)$ or, in other words, the true distributional parameter θ^c , is rarely known exactly and usually estimated from data. We consider an online setting where data arrive sequentially in time and decisions are updated at each time stage. It is natural to take a Bayesian approach for sequential estimation of the unknown parameter, since it is computationally convenient and the estimate is guaranteed with strong consistency with i.i.d. data (however, Bayesian consistency with non-i.i.d. data are much more complicated, which we will discuss later in section 3). With the Bayesian estimate of the distributional parameter, we apply iterations of the SGD algorithm on the estimated problem to update the decision, because the light computational effort of SGD makes it appealing for the online setting. On a high level, at each time stage t , after observing a new batch of data we carry out the following two steps:

- Update the Bayesian posterior distribution of the parameter with the new data.
- Use SGD on the Bayesian average of problem (2.1) or (2.2) to update the decision.

We now discuss the details of these two steps in the following. Let's first focus on the decision-independent case. Suppose at each time stage t we observe a batch of data $\mathbf{y}_t = \{y_{t,j}, j = 1, \dots, D\}$, where $\{y_{t,j}\}$ are i.i.d. according to $f(\cdot; \theta^c)$ and D is the batch size. By viewing the unknown distributional parameter as a random vector θ and assuming a prior distribution π_0 on θ , the posterior distribution of θ is updated by the Bayes rule as follows:

$$(2.3) \quad \pi_t(\theta) = \frac{\pi_{t-1}(\theta) f(\mathbf{y}_t; \theta)}{\int \pi_{t-1}(\theta) f(\mathbf{y}_t; \theta) d\theta} = \frac{\pi_{t-1}(\theta) \prod_{j=1}^D f(y_{t,j}; \theta)}{\int \pi_{t-1}(\theta) \prod_{j=1}^D f(y_{t,j}; \theta) d\theta}.$$

The objective function (2.1) can be viewed as a function of θ , so we define the following function

$$H(x, \theta) := \mathbb{E}_{f(\cdot; \theta)}[h(x, \xi)].$$

To estimate the true objective function (2.1), we consider the Bayesian average of the objective function,

$$(2.4) \quad \min_{x \in \mathcal{X}} \mathbb{E}_{\pi_t}[H(x, \theta)],$$

where the expectation is taken w.r.t. the posterior distribution π_t defined in (2.3). Then we apply SGD on (2.4) for K iterations within each time stage, where K is a user choice or limited by the time length of the current stage before the next batch of data come in. The key element in SGD is the stochastic gradient estimator, and an unbiased gradient estimator of the objective function in (2.4) can be computed by the infinitesimal perturbation analysis [26] as

$$(2.5) \quad \nabla_x h(x, \xi), \quad \xi \sim f(\cdot; \theta) \text{ and } \theta \sim \pi_t.$$

Now let's focus on the decision-dependent case. With a slight abuse of notations, we use the same notations as in the decision-independent case unless defined otherwise. Unlike the decision-independent case where the data batches are i.i.d. over time from the fixed distribution $f(\cdot; \theta^c)$, in the decision-dependent case data batches $\{\mathbf{y}_t\}_t$ are correlated and differently distributed across time stages, since \mathbf{y}_t depends on the decision x_t which is in turn updated from previous data over time. Regardless of the nonstationarity of the data batches, we still use Bayesian posterior distribution to estimate θ :

$$(2.6) \quad \pi_t(\theta) = \frac{\pi_{t-1}(\theta) f(\mathbf{y}_t; x_t, \theta)}{\int \pi_{t-1}(\theta) f(\mathbf{y}_t; x_t, \theta) d\theta} = \frac{\pi_{t-1}(\theta) \prod_{j=1}^D f(y_{t,j}; x_t, \theta)}{\int \pi_{t-1}(\theta) \prod_{j=1}^D f(y_{t,j}; x_t, \theta) d\theta}.$$

Due to the nonstationarity of data batches, the consistency of the posterior distribution is a question here; we will characterize the conditions needed for strong consistency of π_t in section 3. The Bayesian average of the objective function is

$$(2.7) \quad \mathbb{E}_{\pi_t}[H(x, \theta)] = \mathbb{E}_{\pi_t}[\mathbb{E}_{f(\cdot; x, \theta)}[h(x, \xi)]].$$

An unbiased gradient estimator of the objective function (2.7) is

$$(2.8) \quad \nabla_x h(x, \xi) + h(x, \xi) \frac{\nabla_x \hat{f}_t(\xi; x)}{\hat{f}_t(\xi; x)}, \quad \xi \sim f(\cdot; \theta) \text{ and } \theta \sim \pi_t,$$

where $\hat{f}_t(\cdot; x) := \mathbb{E}_{\pi_t}[f(\cdot; x, \theta)]$, $\nabla_x \hat{f}_t(\cdot; x) := \nabla_x \mathbb{E}_{\pi_t}[f(\cdot; x, \theta)]$. The derivation of the gradient estimators (2.5) and (2.8) will be shown in section 3. Informally, (2.8) is obtained by taking derivative of $h(x, \xi)f(\xi; x, \theta)$ w.r.t. x . In the algorithms we assume that the posterior distribution π_t and the expectation in $\hat{f}_t(\cdot; x)$ and $\nabla_x \hat{f}_t(\cdot; x)$ can be exactly computed, which is often the case when we choose a conjugate prior distribution for Bayesian updating. For general posterior distributions, we can use general Markov chain Monte Carlo methods, such as the Langevin algorithm ([23, 20]), to sample from the posterior and use these samples to approximate the expectation. It is worth noting that the first term in (2.8) is the same as the stochastic gradient estimator (2.5) in the decision-independent case, and the second term is unique here and caused by the dependence of the distribution on the decision x .

The algorithms, named Bayesian SGD (Bayesian-SGD), for stochastic optimization with decision-independent uncertainty and decision-dependent uncertainty are shown in Algorithms 2.1 and 2.2, respectively. Please note that to accelerate algorithm convergence, variants of SGD methods could be used instead of the plain SGD iterations in these algorithms.

Algorithm 2.1. Bayesian-SGD (decision-independent uncertainty).

input: data batch size D , number of SGD iterations K , step size sequence $\{a_{t,j}, t=1, 2, \dots; j=0, \dots, K-1\}$, time horizon T .
initialization: choose an initial decision x_1 and prior distribution $\pi_0(\theta)$.
for $t = 1 : T$ **do**
 -A batch of data $y_{t,1}, \dots, y_{t,D} \stackrel{\text{i.i.d.}}{\sim} f(\cdot; \theta^c)$ arrives;
 -Posterior update: compute $\pi_t(\theta)$ according to (2.3).
 -Decision update:

- set $x_{t,0} := x_t$;
- for $j = 0, \dots, K-1$, draw samples $\theta_{t,j} \sim \pi_t(\theta)$ and $\xi_{t,j} \sim f(\cdot; \theta_{t,j})$, and carry out SGD iteration:

$$(2.9) \quad x_{t,j+1} := \text{Proj}_{\mathcal{X}} \{x_{t,j} - a_{t,j} \nabla_x h(x_{t,j}, \xi_{t,j})\},$$

where $\text{Proj}_{\mathcal{X}}$ is a projection operator that projects the iterate to the set \mathcal{X} .

- set the updated decision as $x_{t+1} := x_{t,K}$;

end for
return x_{T+1}

Algorithm 2.2. Bayesian-SGD (decision-dependent uncertainty).

input: data batch size D , number of SGD iterations K , step size sequence $\{a_{t,j}, t=1, 2, \dots; j=0, \dots, K-1\}$, time horizon T .
initialization: choose an initial decision x_1 and prior distribution $\pi_0(\theta)$.
for $t = 1 : T$ **do**
 -A batch of data $y_{t,1}, \dots, y_{t,D} \stackrel{\text{i.i.d.}}{\sim} f(\cdot; x_t, \theta^c)$ arrives;
 -Posterior update: compute $\pi_t(\theta)$ according to (2.6).
 -Decision update:

- set $x_{t,0} := x_t$;
- for $j = 0, \dots, K-1$, draw sample $\theta_{t,j} \sim \pi_t(\theta)$ and $\xi_{t,j} \sim f(\cdot; x_{t,j}, \theta_{t,j})$, and carry out SGD iteration:

$$(2.10) \quad x_{t,j+1} := \text{Proj}_{\mathcal{X}} \left\{ x_{t,j} - a_{t,j} \left(\nabla_x h(x_{t,j}, \xi_{t,j}) \right. \right. \\ \left. \left. + h(x_{t,j}, \xi_{t,j}) \frac{\nabla_x \hat{f}_t(\xi_{t,j}; x_{t,j})}{\hat{f}_t(\xi_{t,j}; x_{t,j})} \right) \right\},$$

where $\text{Proj}_{\mathcal{X}}$ is a projection operator that projects the iterate to the set \mathcal{X} .

- set the updated decision as $x_{t+1} := x_{t,K}$;

end for
return x_{T+1}

3. Convergence analysis. In this section, we show asymptotic convergence of Algorithms 2.1 and 2.2. Towards this end, we first need to show the consistency of the Bayesian posterior distribution and then show the convergence of SGD when applied to the nonstationary Bayesian average stochastic optimization problems (2.4) and (2.5). In addition, we show the convergence rate in the decision-independent case.

3.1. Convergence analysis for the decision-independent case. Let's first consider the decision-independent case. The probability space is constructed as follows. Define the Bayesian prior π_0 on $(\Theta, \mathcal{B}_\Theta)$, where \mathcal{B}_Θ is the Borel σ -algebra on Θ . Let $\mathcal{Y} \subset \mathbb{R}^m$ denote the data (observation) space. The data y take values in \mathcal{Y} equipped with a Borel σ -algebra $\mathcal{B}_\mathcal{Y}$ and a probability measure $\{\mathbb{P}_{\theta^c}\}$, such that $\mathbb{P}_{\theta^c}(y \in A) = \int_A f(y; \theta^c) dy, \forall A \in \mathcal{B}(\mathcal{Y})$. For the sequence $y_1, y_2, \dots, y_n \stackrel{\text{i.i.d.}}{\sim} f(\cdot; \theta^c)$, the probability measure is denoted by $\mathbb{P}_{\theta^c}^n$. As for the infinite sequence $\{y_1, y_2, \dots\}$, the probability measure $\mathbb{P}_{\theta^c}^\infty$ can be constructed by Kolmogorov's extension theorem (cf. Theorem A.3.1 in [21]). In the following, w.p.1 (or almost surely) means that the considered property holds with probability one w.r.t. the probability measure $\mathbb{P}_{\theta^c}^\infty$. Finally, let $\mathcal{F}_t := \sigma\{(y_\tau), \tau \leq t\}$ be the σ -filtration generated by the data. We have the convergence of the posterior distribution $\{\pi_t\}$ that is updated according to (2.3) under the following assumptions.

Assumption 3.1 (see [56, Assumption 3.1]). (i) The set Θ is convex and compact with nonempty interior. (ii) $\ln \pi_0(\theta)$ is bounded on Θ . (iii) $f(\xi|\theta) > 0$ for all $\xi \in \Xi$ and $\theta \in \Theta$. (iv) $f(\xi|\theta)$ is continuous in $\theta \in \Theta$. (v) $\ln f(\xi|\theta), \theta \in \Theta$, is dominated by an integrable (w.r.t. $\xi \sim f(\cdot; \theta^c)$) function. (vi) The data batches are i.i.d. over time from the fixed distribution $f(\cdot; \theta^c)$.

We refer the readers to [56] for detailed explanations of the above assumptions. The next lemma shows the Bayesian consistency under Assumption 3.1, which implies the distributional uncertainty diminishes as $t \rightarrow \infty$.

DEFINITION 3.2 (weak convergence). *A sequence of distributions $\mathbb{P}_n \Rightarrow P$ if and only if $\int g d\mathbb{P}_n \rightarrow \int g dP$ as $n \rightarrow \infty$ for all g bounded and continuous.*

LEMMA 3.3 (see [56, Lemma 3.2]). *Under Assumption 3.1, $\pi_t(\theta) \Rightarrow \delta_{\theta^c}(\theta)$ w.p.1, where δ_{θ^c} is the Dirac delta function concentrated on the true parameter θ^c .*

We then study the asymptotic behavior of Algorithm 2.1 by the ordinary differential equation (ODE) method (please refer to [39] for a detailed exposition on the ODE method for stochastic approximation). The main idea is that SGD can be viewed as a noisy discretization of an ODE. Under certain conditions, the noise in SGD averages out asymptotically, such that the SGD iterates converge to the solution trajectory of the ODE. For simplicity, we consider the case where $K = 1$ and rewrite the SGD iteration (2.9) as

$$(3.1) \quad x_{t+1} = x_t - a_t \nabla_x h(x_t, \xi_t) + a_t z_t,$$

where $a_t z_t$ is the projection term, i.e., the vector of shortest Euclidean length needed to keep the decision x_{t+1} from leaving the decision space \mathcal{X} . We first show that under certain mild conditions, the proposed gradient estimator in (3.1) is unbiased.

Assumption 3.4. $h(x, \xi)$ is C^1 -smooth in x for all $\xi \in \Xi$, and the map $\xi \rightarrow \nabla_x h(x, \xi)$ is L_h -Lipschitz continuous for any $x \in \mathcal{X}$.

Assumption 3.4 is a commonly used smooth assumption in the stochastic approximation literature (cf. [28, 15]). An important consequence is that for any probability measure, $\mathbb{E}h(x, \xi)$ is differentiable in x with gradient $\mathbb{E}\nabla_x h(x, \xi)$ (cf. [15]).

LEMMA 3.5. *Under Assumption 3.4, $\nabla_x h(x, \xi)$ with $\xi \sim f(\cdot; \theta)$ and $\theta \sim \pi_t$ is an unbiased gradient estimator of the objective function in (2.4).*

Proof. For every fixed $x \in \mathcal{X}$,

$$\begin{aligned}\mathbb{E}_{\pi_t} [\mathbb{E}_{f(\cdot; \theta)} [\nabla_x h(x, \xi)]] &= \mathbb{E}_{\pi_t} [\nabla_x \mathbb{E}_{f(\cdot; \theta)} [h(x, \xi)]] \\ &= \nabla_x \mathbb{E}_{\pi_t} [\mathbb{E}_{f(\cdot; \theta)} [h(x, \xi)]],\end{aligned}$$

where the first equality holds because the gradient $\nabla_x h(x, \xi)$ is Lipschitz continuous, and the interchange between expectation and differentiation is justified by the dominated convergence theorem (DCT). Similarly, the second equality above is again justified by the DCT. Therefore, the proposed estimator in (3.1) is an unbiased gradient estimator of the objective function in (2.4). \square

Assumption 3.6.

- The step size $\{a_t\}$ satisfies $\sum_{t=1}^{\infty} a_t^2 < \infty$, $\sum_{t=1}^{\infty} a_t = \infty$, $\lim_{t \rightarrow \infty} a_t = 0$, $a_t > 0 \forall t > 0$.
- The decision space $\mathcal{X} \subset \mathbb{R}^d$ is compact and convex.

The above assumptions on the step size and the compact and convex decision space are often used in SGD (cf. [39]). The first assumption essentially requires that the step size diminishes to zero not too slowly ($\sum_{t=1}^{\infty} a_t^2 < \infty$) nor too quickly ($\sum_{t=1}^{\infty} a_t = \infty$). For example, we can choose $a_t = \frac{a}{t}$ for some $a > 0$.

Before proceeding to our main convergence result, we introduce the continuous-time interpolations of the decision sequence $\{x_t\}$. Define $t_1 = 1$ and $t_n = 1 + \sum_{i=1}^{n-1} a_i, n \geq 2$. For $t \geq 1$, let $N(t)$ be the unique n such that $t_n \leq t < t_{n+1}$. For $t < 1$, set $N(t) = 1$. Define the interpolated continuous process X as $X(1) = x_1$ and $X(t) = x_{N(t)}$ for any $t > 1$, and the shifted process as $X^n(s) = X(s + t_n)$. We then show in the following theorem that Algorithm 2.1 converges w.p.1.

THEOREM 3.7. *Let $\mathcal{D}^d[0, \infty)$ be the space of \mathbb{R}^d -valued operators which are right continuous and have left-hand limits for each dimension. Under Assumptions 3.1, 3.4, and 3.6, there exists a process $X^*(\cdot)$ to which the subsequence of $\{X^n(\cdot)\}_n$ converges w.p.1 in the space $\mathcal{D}^d[0, \infty)$, where $X^*(\cdot)$ satisfies the following ODE,*

$$(3.2) \quad \dot{X} = -\nabla H(X, \theta^c) + z, \quad z \in -\mathcal{C}(X), \quad X(1) = x_1,$$

where $\mathcal{C}(X)$ is the Clarke's normal cone to \mathcal{X} , i.e., for any $x \in \mathcal{X}$, $\mathcal{C}(x) = \{c : c^T x \geq c^T y \forall y \in \mathcal{C}\}$. z is the projection term: it is the vector of shortest Euclidean length needed to keep the trajectory of the ODE $X(\cdot)$ from leaving the decision space \mathcal{X} . The sequence $\{x_t\}_t$ in (3.1) also converges w.p.1 to the limit set of the ODE (3.2).

Proof. Note that

$$\begin{aligned}\mathbb{E} [\nabla_x h(x_t, \xi_t) | x_1, y_s, \xi_s, s < t] &= \mathbb{E}_{\pi_t} [\mathbb{E}_{f(\cdot; \theta)} [\nabla_x h(x_t, \xi)]] \\ &= \nabla_x H(x_t, \theta^c) + (\mathbb{E}_{\pi_t} [\mathbb{E}_{f(\cdot; \theta)} [\nabla_x h(x_t, \xi)]] - \nabla_x H(x_t, \theta^c)) \\ &= \nabla_x H(x_t, \theta^c) + (\mathbb{E}_{\pi_t} [\mathbb{E}_{f(\cdot; \theta)} [\nabla_x h(x_t, \xi)]] - \mathbb{E}_{\delta_{\theta^c}} [\mathbb{E}_{f(\cdot; \theta)} [\nabla_x h(x_t, \xi)]]).\end{aligned}$$

Let $\epsilon_t = \mathbb{E}_{\pi_t} [\mathbb{E}_{f(\cdot; \theta)} [\nabla_x h(x_t, \xi)]] - \mathbb{E}_{\delta_{\theta^c}} [\mathbb{E}_{f(\cdot; \theta)} [\nabla_x h(x_t, \xi)]]$. By Lemma 3.3, $\pi_t(\theta) \Rightarrow \delta_{\theta^c}(\theta)$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$), and by Theorem 3.1 in [64], $\epsilon_t \rightarrow 0$ as $t \rightarrow \infty$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$). We can then directly apply Theorem 5.2.3 in [39] and obtain the result. \square

Remark 3.8. The SGD iterates specified in (2.9) approach the solution trajectory of the ODE (3.2) and eventually converge to a limit point of the ODE, which is a point x^* satisfying $\nabla H(x^*, \theta^c) = 0$ if the point is in the interior of \mathcal{X} . Hence, such

a point is a stationary point of problem (2.1) for the decision-independent case and can be a local optimal solution if it is stable. On a related note, stochastic gradient Langevin dynamic (SGLD), a popular variant of SGD, adds properly scaled isotropic Gaussian noise to an unbiased estimate of the gradient at each iteration, which allows the solution trajectory to escape local minimum and guarantees asymptotic convergence to a global minimizer for sufficiently regular nonconvex objectives (see [52, 18] and references therein). It is an interesting future direction to apply SGLD to our considered stochastic optimization problem with streaming input data.

Next, we investigate the convergence rate of Algorithm 2.1 for the unconstrained case, i.e., without the projection term $a_t z_t$ under the following additional assumptions.

Assumption 3.9.

- The parameter space Θ is finite, i.e., $\Theta = \{\theta_1, \dots, \theta_k\}$. Moreover, $\theta^c \in \Theta$.
- There exists $0 < L_H < \infty$ such that $\|\nabla_x H(x, \theta_1) - \nabla_x H(x, \theta_2)\|_2 \leq L_H \|\theta_1 - \theta_2\|_2$ for all $\theta_1, \theta_2 \in \Theta$ and for all $x \in \mathcal{X}$.
- Sampling variance is bounded by σ^2 , i.e., $\mathbb{E}[\|\nabla_x h(x, \xi) - \nabla_x H(x, \theta)\|_2^2 | \theta] \leq \sigma^2$, for all $\theta \in \Theta$.

Due to technical challenges, in Assumption 3.9 we only consider a finite parameter space, which is practical in many real-world problems. For example, it can be viewed as a discrete approximation of a continuous parameter set, and the discretization can be chosen of any precision. The second assumption essentially requires $H(x, \theta)$ is C^1 -smooth in θ for all $x \in \mathcal{X}$ and is a common assumption in stochastic approximation literature (cf. [57]). The bounded sampling variance is also a common assumption in nonconvex SGD convergence analysis (cf. [54]).

Under Assumption 3.9, we can show the bias term (the difference between $\mathbb{E}_{\pi_t} \nabla_x H(x, \theta)$ and $\mathbb{E}_{\pi_t} \nabla_x H(x, \theta^c)$) can be upper bounded with high probability, which serves as a key lemma in showing the convergence rate of the decision-independent algorithm.

LEMMA 3.10. *Under Assumption 3.9, there exists a constant $C_1 > 0$ such that for any $\delta > 0$, with probability at least $1 - \delta$ we have*

$$\|\mathbb{E}_{\pi_t} \nabla_x H(x, \theta) - \mathbb{E}_{\pi_t} \nabla_x H(x, \theta^c)\|_2^2 \leq C_1 \frac{\log Dt + \log \frac{1}{\delta}}{Dt} \quad \forall x \in \mathcal{X} \quad \forall t > 0.$$

The proof of Lemma 3.10 can be found in Appendix A. Next, we show the convergence rate of Algorithm 2.1. To simplify the analysis and also be consistent with the convergence analysis of smooth nonconvex SGD, we consider a variant of SGD where the final output is randomly chosen as follows: let $z_T = x_t$ with probability $\frac{a_t}{\sum_{t=1}^T a_t}$, $t = 1, \dots, T$. The randomization scheme helps with the analysis of the expected gradient of the final output under the true parameter θ^c , and has been widely used in the smooth nonconvex SGD literature (cf. [28]). We then have the following theorem giving the convergence rate of the randomized output algorithm under different step sizes.

THEOREM 3.11. *Under Assumptions 3.1, 3.4, 3.6, and 3.9, for any $\delta > 0$, we have with probability at least $1 - \delta$, for any $T > 0$, the following bound on the expected gradient of the final output under the true parameter θ^c :*

(i) If the step size satisfies $a_t = \frac{a}{\sqrt{T}}$ $\forall t \leq T$ for some constant $a < \frac{\sqrt{T}}{L_h}$, then

$$\begin{aligned} & \mathbb{E}[\|\nabla_x H(z_T, \theta^c)\|_2^2] \\ & \leq \left[\frac{2(H(x_1, \theta^c) - \min_{x \in \mathcal{X}} H(x, \theta^c))}{a\sqrt{T}} \right] \\ & \quad + \left[\frac{A_1}{T} + \frac{A_2 \log T}{T} + \frac{A_3 \log^2 T}{T} \right] + \frac{L_h a \sigma^2}{\sqrt{T}}, \end{aligned}$$

where $A_1 = \frac{C_1(\log D - \log \delta)}{L_h D}$, $A_2 = \frac{C_1(\log D - \log \delta)}{L_h D} + \frac{C_1}{L_h D}$, $A_3 = \frac{C_1}{L_h D}$.

(ii) If the step size satisfies $a_t = \frac{a}{t}$ $\forall t \leq T$ for some constant $a < \frac{1}{L_h}$, then

$$\begin{aligned} & \mathbb{E}[\|\nabla_x H(z_T, \theta^c)\|_2^2] \\ & \leq \left[\frac{2(H(x_1, \theta^c) - \min_{x \in \mathcal{X}} H(x, \theta^c))}{a} + \frac{6C_1 + \pi^2 C_1 (\log D - \log \delta)}{6D} \right. \\ & \quad \left. + \frac{\pi^2 L_h a \sigma^2}{6} \right] \frac{1}{\log T}. \end{aligned}$$

(iii) If the step size satisfies $a_t = \frac{a}{\sqrt{t}}$ $\forall t \leq T$ for some constant $a < \frac{1}{L_h}$, then

$$\begin{aligned} & \mathbb{E}[\|\nabla_x H(z_T, \theta^c)\|_2^2] \\ & \leq \left[\frac{2(H(x_1, \theta^c) - \min_{x \in \mathcal{X}} H(x, \theta^c))}{a\sqrt{T}} + \frac{3C_1(\log D - \log \delta) + 4C_1}{D\sqrt{T}} + \frac{L_h a \sigma^2}{\sqrt{T}} \right] \\ & \quad + \frac{L_h a \sigma^2 \log T}{\sqrt{T}}. \end{aligned}$$

The proof of Theorem 3.11 can be found in Appendix B. Theorem 3.11 shows that for the constant step size $a_t = \frac{a}{\sqrt{T}}$, the convergence rate is $O(\frac{1}{\sqrt{T}})$. Note that in case (i), the first term in the convergence rate depends on the initialization of the solution (difference between $H(x_1, \theta^c)$ and $\min_x H(x, \theta^c)$); the last term depends on the Lipschitz constant and sampling variance. These two terms are consistent with the classical smooth nonconvex SGD (cf. [28]). The second, third, and fourth terms are caused by the difference between $\mathbb{E}_{\pi_t} \nabla_x H(x, \theta)$ and $\mathbb{E}_{\pi_t} \nabla_x H(x, \theta^c)$, which is due to the Bayesian estimation that is unique to the considered problem. As for the classical decreasing step size $a_t = \frac{a}{t}$, the convergence rate is $O(1/\log T)$. For the bigger decreasing step size $a_t = \frac{a}{\sqrt{t}}$, the convergence rate is $O(\log T/\sqrt{T})$.

3.2. Convergence analysis for the decision-dependent case. In this section, we theoretically study the convergence behavior of Algorithm 2.2. We follow the approach in [13] to construct the probability space for the decision-dependent case. Note that the data y take values in the space \mathcal{Y} equipped with a Borel σ -algebra $\mathcal{B}_{\mathcal{Y}}$ and a probability measure $\mathbb{P}_{\theta^c}(\cdot|x)$ such that $\mathbb{P}_{\theta^c}(y \in A|x) = \int_A f(y; x, \theta^c) dy \forall A \in \mathcal{B}_{\mathcal{Y}}$. Suppose that there is a probability space $(\mathcal{S}, \mathcal{H}, \mu)$ and a measurable map $F : \mathcal{S} \times \mathcal{X} \rightarrow \mathcal{Y}$ such that for every set $A \in \mathcal{B}_{\mathcal{Y}}$, the $\mathbb{P}_{\theta^c}(\cdot|x)$ -measure of A is equal to the μ -measure of the set $\{s \in \mathcal{S} : F(s, x) \in A\}$. Then we define $(\Omega, \mathcal{F}, \mathbb{P}_{\theta^c}^\infty)$ as the countable product $(\mathcal{S}, \mathcal{H}, \mu)^\infty$. In the following, w.p.1 (or almost surely) means that the considered property holds with probability one w.r.t. the probability measure $\mathbb{P}_{\theta^c}^\infty$. Let $\mathcal{F}_t = \sigma\{(x_\tau, y_\tau), \tau \leq t\}$ be the σ -filtration generated by the data and decision sequences. For simplicity, we assume at each time stage the data batch size $D = 1$ and the number of SGD iterations $K = 1$. We have the convergence of

the posterior distribution $\{\pi_t\}$ that is updated according to (2.6) under the following assumptions.

Assumption 3.12.

- The parameter space Θ is discrete. Moreover, $\theta^c \in \Theta$.
- The prior distribution $\pi_0(\theta^c) > 0$.

The assumptions above are regularity conditions and easy to be verified in practice. Note that Algorithm 2.2 works for a general parameter space, but due to technical challenges, we assume a discrete parameter space for the convergence analysis. Note that for the decision-dependent case, the correlated and differently distributed data $\{\mathbf{y}_t\}$ pose a great challenge to analyzing the consistency of the Bayesian posterior distribution π_t . To prove the Bayesian consistency, we first show the following intermediate result. Let $D_{KL}(P\|Q) := \int \log(\frac{dP}{dQ})dP$ denote the Kullback–Leibler (K–L) divergence from distribution P to distribution Q .

LEMMA 3.13. *Suppose Assumption 3.12 holds. Recall $\hat{f}_t(\cdot; x) = \sum_{\theta} \pi_t(\theta) f(\cdot; x, \theta)$. Denote $f^*(\cdot; x) := f(\cdot; x, \theta^c)$, for any $x \in \mathcal{X}$. At decision x_{t+1} , the K–L divergence from f^* to \hat{f}_t is denoted as d_t , i.e., $d_t := D_{KL}(f^*(\cdot; x_{t+1})\| \hat{f}_t(\cdot; x_{t+1}))$. Then we have*

$$\lim_{t \rightarrow \infty} d_t = 0 \text{ and } \sum_{t=1}^{\infty} d_t < \infty \text{ w.p.1}(\mathbb{P}_{\theta^c}^{\infty}).$$

The proof of Lemma 3.13 can be found in Appendix C. Intuitively, Lemma 3.13 implies that with more observation data even at different decisions, we know more about the true parameter θ^c and are able to provide a more precise estimation of the density f^* at the next decision. Moreover, if we know that each θ is identifiable as rigorously defined in the following assumption, we can further prove the consistency of $\{\pi_t\}$ regardless of the correlation and nonstationarity of the observation data.

Assumption 3.14 (linear independence). For almost every x in \mathcal{X} for any $\mathcal{K} \subseteq \mathbb{N}$, where \mathbb{N} is the set of natural numbers, $\{f(\cdot; x, \theta_i)\}_{i \in \mathcal{K}}$ are linearly independent in \mathcal{Y} , i.e.,

$$\sum_{i \in \mathcal{K}} c_i f(y; x, \theta_i) = 0 \quad \forall y \in \mathcal{Y} \Rightarrow c_i = 0 \quad \forall i \in \mathcal{K}.$$

Assumption 3.14 intuitively requires that for almost every decision x , the observation distributions generated from different θ 's are distinguishable (or identifiable; cf. Definition 5.2 in [42]). For the ease of notation, we denote the density function as $f(\cdot; x, \theta) := f(\cdot; g(x, \theta))$, where $g : \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}^s$ is a mapping from $\mathcal{X} \times \Theta$ to the s -dimensional parameter space of the distribution. A necessary condition for Assumption 3.14 to hold is $g(x, \theta_1) \neq g(x, \theta_2)$ for almost every $x \in \mathcal{X}$ and for all $\theta_1 \in \Theta$, $\theta_2 \in \Theta$ such that $\theta_1 \neq \theta_2$. Under this necessary condition, Assumption 3.14 is satisfied by many distribution families. For example, the Wronskian determinant for exponential distributions with different parameters $g(x, \theta_1), \dots, g(x, \theta_n)$ is computed by $W(\xi) = \prod_{i=1}^n g(x, \theta_i) \exp(-\sum_{i=1}^n g(x, \theta_i)\xi) \prod_{i \neq j} (g(x, \theta_i) - g(x, \theta_j))$, which is nonzero for almost every $x \in \mathcal{X}$ and all $\xi \in \Xi$ when the θ_i 's are distinct, which directly implies the linear independence of $\{f(\cdot; x, \theta_i)\}_i$. For other exponential families, such as normal, gamma, and Poisson, a general solution to check the Wronskian determinant may not be readily available. Instead, one could check whether the components of the sufficient statistics are linearly independent, i.e., whether the exponential family is minimal (cf. Chapter 1.5 in [42]).

Assumption 3.15. The decision space $\mathcal{X} \subset \mathbb{R}^d$ is compact and convex.

We then have the following proposition on the consistency of the posterior distribution $\{\pi_t\}$.

PROPOSITION 3.16. *Under Assumptions 3.12, 3.14, and 3.15, $\pi_t \Rightarrow \delta_{\theta^*}$ w.p. 1 ($\mathbb{P}_{\theta^*}^\infty$).*

The proof of Proposition 3.16 can be found in Appendix D. Proposition 3.16 guarantees that although the observation at each time depends on the current decision, it can provide enough information to ensure the posterior distribution will eventually concentrate on the true parameter. In the following, we will show that the consistency of $\{\pi_t\}$ ensures that the gradient estimator is accurate enough and thus Algorithm 2.2 converges.

Remark 3.17. We note that the consistency of posterior distributions for non-i.i.d. observations is previously shown in [29]. However, they give very general convergence result with assumptions (such as existence of a testing function sequence) that are often abstract and hard to verify in practice. On the other hand, our Bayesian consistency result is built on assumptions (in particular Assumption 3.14) that are easy to verify and interpret.

We then study the asymptotic behavior of Algorithm 2.2 by the ODE method similar to the decision-independent case. We can rewrite the SGD iteration (2.10) as

$$(3.3) \quad x_{t+1} = x_t - a_t \left(\nabla_x h(x_t, \xi_t) + h(x_t, \xi_t) \frac{\nabla_x \hat{f}_t(\xi_t; x_t)}{\hat{f}_t(\xi_t; x_t)} \right) + a_t z_t,$$

where $a_t z_t$ is the projection term. We show that under certain mild conditions, the proposed gradient estimator (3.3) is unbiased.

Assumption 3.18. The density function $f(\xi; x, \theta)$ is C^1 -smooth in x for all $\xi \in \Xi$ and for all $\theta \in \Theta$.

Together with Assumption 3.4, Assumption 3.18 puts mild conditions that justify the interchange between differentiation and integral for the decision-dependent case.

LEMMA 3.19. *Under Assumptions 3.4 and 3.18, we have that $\nabla_x h(x_t, \xi) + h(x_t, \xi) \frac{\nabla_x \hat{f}_t(\xi; x_t)}{\hat{f}_t(\xi; x_t)}$ with $\xi \sim f(\cdot; x_t, \theta)$ and $\theta \sim \pi_t$ is an unbiased gradient estimator of the objective function in (2.7).*

The detailed derivation can be found in Appendix E. Note that in performative prediction literature (e.g., [15]), the gradient estimator is also derived using the chain rule similarly to (2.8). However, due to the difficulty in estimating the second term, most of the literature in performative prediction focus only on the first term, and show that under the biased gradient estimator, the solution converges to a so-called performative stable point which is in general different from the true optimal solution. In contrast, our approach provides a Bayesian way to estimate the second term under the parametric assumption and aims to converge to the true optimal solution of problem (2.2).

A final set of assumptions on the step size to show the convergence of Algorithm 2.2 is listed below.

Assumption 3.20. The step size a_t satisfies $\sum_{t=1}^{\infty} a_t = \infty, \lim_{t \rightarrow \infty} a_t = 0, a_t > 0 \forall t > 0$.

We then have the following theorem showing the weak convergence of Algorithm 2.2.

THEOREM 3.21. *Let $\mathcal{D}^d[0, \infty)$ be the space of \mathbb{R}^d -valued operators which are right continuous and have left-hand limits for each dimension. Under Assumptions 3.4, 3.12, 3.14, 3.15, 3.18, and 3.20, for each subsequence of $\{X^n(\cdot)\}_n$, there exists a further subsequence $\{X^{n_k}(\cdot)\}_{n_k}$ and a process $X^*(\cdot)$ such that $X^{n_k}(\cdot) \Rightarrow X^*(\cdot)$ in the weak sense as $t \rightarrow \infty$ in the space $D^d[0, \infty)$, where $X^*(\cdot)$ satisfies the following ODE,*

$$(3.4) \quad \dot{X} = -\nabla H(X, \theta^c) + z, \quad z \in -\mathcal{C}(X), \quad X(1) = x_1,$$

where $\mathcal{C}(X)$ is the Clarke's normal cone to \mathcal{X} , i.e., for any $x \in \mathcal{X}$, $\mathcal{C}(x) = \{c : c^T x \geq c^T y, \forall y \in \mathcal{C}\}$. z is the projection term: it is the vector of shortest Euclidean length needed to keep the trajectory of the ODE $X(\cdot)$ from leaving the decision space \mathcal{X} . Let $L_{\mathcal{X}}$ be the set of limit points of (3.4) in \mathcal{X} . Then there exist $\mu_n \rightarrow 0$ and $T_n \rightarrow \infty$ such that

$$\lim_n P \left\{ \sup_{t \leq T_n} \text{Dist}(X^n(t), L_{\mathcal{X}}) \geq \mu_n \right\} = 0,$$

where $\text{Dist}(x, \mathcal{E}) = \inf_{y \in \mathcal{E}} \|x - y\|_2$ for any set \mathcal{E} and point $x \in \mathcal{X}$. The sequence $\{x_t\}_t$ in (3.3) also converges weakly to the limit set of the ODE (3.4).

Remark 3.22. (3.4) shows the weak convergence of Algorithm 2.2. The SGD iterates specified in (2.10) approach the solution trajectory of the ODE (3.4) and eventually converge to a limit point of the ODE, which is a point x^* satisfying $\nabla H(x^*, \theta^c) = 0$ if the point is in the interior of \mathcal{X} . Hence, such a point is a stationary point of problem (2.2) for the decision-dependent case and can be a local optimal solution if it is stable. The weak convergence result implies that once the trajectory enters the domain of attraction of a local optimal solution, the chance of escaping from it goes to 0 in the limit.

Now we prove (3.4) below.

Proof. Recall that at time $t + 1$, Algorithm 2.2 takes the following update:

$$x_{t+1} = x_t - a_t \left(\nabla_x h(x_t, \xi_t) + h(x_t, \xi_t) \frac{\nabla_x \hat{f}_t(\xi_t; x_t)}{\hat{f}_t(\xi_t; x_t)} \right) + a_t z_t.$$

From the derivation of the unbiased gradient estimator in Appendix E, we have

$$\begin{aligned} & \mathbb{E}_{\pi_t} [\mathbb{E}_{f(\cdot; x_t, \theta)} [\nabla_x h(x_t, \xi)]] \\ &= \mathbb{E}_{\hat{f}_t(\cdot; x_t)} [\nabla_x h(x_t, \xi)] \\ &= \mathbb{E}_{f^*(\cdot; x_t)} \nabla_x h(x_t, \xi) + \left(\mathbb{E}_{\hat{f}_t(\cdot; x_t)} [\nabla_x h(x_t, \xi)] - \mathbb{E}_{f^*(\cdot; x_t)} [\nabla_x h(x_t, \xi)] \right) \\ &= \mathbb{E}_{f^*(\cdot; x_t)} [\nabla_x h(x_t, \xi)] + \beta_{t,1}, \end{aligned}$$

where $f^*(\cdot; x) = f(\cdot; x, \theta^c)$ for $x \in \mathcal{X}$, $\beta_{t,1} = \mathbb{E}_{\hat{f}_t(\cdot; x_t)} [\nabla_x h(x_t, \xi)] - \mathbb{E}_{f^*(\cdot; x_t)} [\nabla_x h(x_t, \xi)]$. Similarly, we have

$$\begin{aligned} & \mathbb{E}_{\pi_t} \left[\mathbb{E}_{f(\cdot; x_t, \theta)} \left[h(x_t, \xi) \frac{\nabla_x \hat{f}_t(\xi; x_t)}{\hat{f}_t(\xi; x_t)} \right] \right] \\ &= \int_{\Xi} h(x_t, \xi) \nabla_x \hat{f}_t(\xi; x_t) d\xi_t \end{aligned}$$

$$\begin{aligned}
&= \int_{\Xi} h(x_t, \xi_t) \nabla_x f^*(\xi; x_t) d\xi \\
&\quad + \left(\int_{\Xi} h(x_t, \xi) \nabla_x \hat{f}_t(\xi; x_t) d\xi - \int_{\Xi} h(x_t, \xi) \nabla_x f^*(\xi; x_t) d\xi \right) \\
&= \int_{\Xi} h(x_t, \xi) \nabla_x f^*(\xi; x_t) d\xi + \beta_{t,2}
\end{aligned}$$

where $\beta_{t,2} = \int_{\Xi} h(x_t, \xi) \nabla_x \hat{f}_t(\xi; x_t) d\xi - \int_{\Xi} h(x_t, \xi) \nabla_x f^*(\xi; x_t) d\xi$. Note that

$$\int_{\Xi} h(x_t, \xi) \nabla_x f^*(\xi; x_t) d\xi + \mathbb{E}_{f^*(\cdot; x_t)}[\nabla_x h(x_t, \xi)] = \nabla_x H(x, \theta^c),$$

and we can rewrite the update as

$$x_{t+1} = x_t - a_t \nabla_x H(x_t, \theta^c) - a_t \beta_{t,1} - a_t \beta_{t,2} - a_t \delta M_t + a_t z_t,$$

where

$$\delta M_t = \nabla_x h(x_t, \xi_t) + h(x_t, \xi_t) \frac{\nabla_x \hat{f}_t(\xi_t; x_t)}{\hat{f}_t(\xi_t; x_t)} - \mathbb{E}_{\hat{f}_t(\cdot; x_t)} \left[\nabla_x h(x_t, \xi) + h(x_t, \xi) \frac{\nabla_x \hat{f}(\xi; x_t)}{\hat{f}(\xi; x_t)} \right]$$

is a martingale difference sequence. Suppose that we can show $\lim_{t \rightarrow \infty} \beta_{t,1} = 0$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$) and $\lim_{t \rightarrow \infty} \beta_{t,2} = 0$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$), then the rest of the update is exactly the discretization of ODE (3.4). Then (3.4) is proved by a straightforward application of Theorem 7.2.1 in [39]. We conclude the proof with the following two lemmas showing that the two bias terms $\beta_{t,1}$ and $\beta_{t,2}$ vanish in the limit.

LEMMA 3.23. *Under Assumptions 3.4, 3.12, 3.18, and 3.20, we have $\lim_{t \rightarrow \infty} \beta_{t,1} = 0$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$).*

LEMMA 3.24. *Under Assumptions 3.4, 3.12, 3.14, and 3.18, we have $\lim_{t \rightarrow \infty} \beta_{t,2} = 0$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$).*

See Appendices F and G for the detailed proofs of the above two lemmas.

Finally, we summarize the main similarities and differences between decision-independent and decision-dependent cases below. Both cases require a compact and convex decision space \mathcal{X} and smoothness of the objective function $h(x, \xi)$ in x . For the decision-dependent case, we also require the density function $f(\xi; x, \theta)$ to be smooth in x , since the gradient estimator of the objective function in (2.7) involves the gradient of $f(\xi; x, \theta)$ and, moreover, we assume linear independence between densities in order to show the consistency of the posterior distribution with non-i.i.d. decision-dependent data. For the decision-independent case, we further impose some stronger conditions in order to show stronger results, including the finiteness of the parameter space Θ to show the convergence rate, and a stricter step size assumption to show the strong convergence of the solution sequence to the limit set of the ODE.

4. Numerical experiments.

4.1. Synthetic test problems. We first demonstrate the performance of Algorithms 2.1 and 2.2 on two synthetic test problems in a univariate setting and in a multivariate setting, respectively. Our method is abbreviated as Bayesian-SGD.

4.1.1. Decision-independent uncertainty. We first carry out numerical experiments on a simple quadratic problem in a univariate setting, $h(x, \xi) = (x - 5)^2 +$

$0.5\xi x$, where $\xi \sim \mathcal{N}(\theta^c, \sigma^2)$. The parameter values are as follows: $\sigma = 4$, $\theta^c = 9$, $D = 1$, $K = 1$, $\Theta = \{1, 2, \dots, 20\}$, $a_t = \frac{2}{t+5}$. It is easy to check $H(x, \theta^c) = x^2 - 5.5x + 25$, and the true optimal decision is taken at $x^* = 2.75$. At each time t , the gradient estimator in Algorithm 2.1 is $\nabla_x h(x_t, \xi_t) = 2x_t - 10 + 0.5\xi$. In Algorithm 2.1, we use the uniform distribution on Θ as the prior distribution and set the initial solution $x_1 = 0$.

As a benchmark, we assume the true parameter θ^c is known and use the plain SGD algorithm on the true problem (2.1). Obviously, with the knowledge of the true parameter value this algorithm should provide a lower bound on the objective value that can be achieved. We also compare this with the MLE method (cf. [57]), which uses the MLE $\hat{\theta}_t$ at each time stage to replace the unknown θ^c in the objective function (2.1) and then solves the corresponding optimization problem by SGD. For fair comparison, we use the same number of SGD iterations at each time stage for all three algorithms. We run all three algorithms (Algorithm 2.1, benchmark, MLE) for 100 times on the problem. The mean and standard deviation of the solution error $|x_t - x^*|$ over time are shown in Figure 1. The observations from Figure 1 can be summarized as follows.

- With decreasing step size, the solution sequence in Algorithm 2.1 converges to the true optimal solution.
- The benchmark algorithm (without parameter uncertainty) performs better than the proposed algorithm and the MLE algorithm, but in the long run (e.g., $t > 1000$ to be shown in a multivariate setting) the three algorithms behave similarly.
- In the initial time stages our algorithm performs slightly better than the MLE algorithm. This is due to the better estimation of the objective function by the Bayesian average in our algorithm than the point estimate in the MLE algorithm, when the data are limited.

We then carry out numerical experiments on a quadratic problem in a multivariate setting, $h(x, \xi) = (x_1 - 1)^2 + (x_2 - 2)^2 + \xi(x_1 + x_2)$, where ξ follows an exponential distribution with mean θ^c . The parameter values are as follows: $\theta^c = 4$, $D = 1$, $K = 1$, $\Theta = \{1, 2, \dots, 20\}$, $a_t = \frac{2}{t+5}$. It is easy to check $H(x, \theta^c) = (x_1 + 1)^2 + x_2^2 + 4$, and the true optimal decision is taken at $x^* = (-1, 0)$. At each time t , the gradient estimator in Algorithm 2.1 is $\nabla_x h(x_t, \xi_t) = (2x_1 - 2 + \xi, 2x_2 - 4 + \xi)$. We use the uniform distribution on Θ as the prior distribution and set the initial solution $x_1 = (5, 5)$. We again run all three algorithms (Algorithm 2.1, benchmark, MLE) for 100 times on

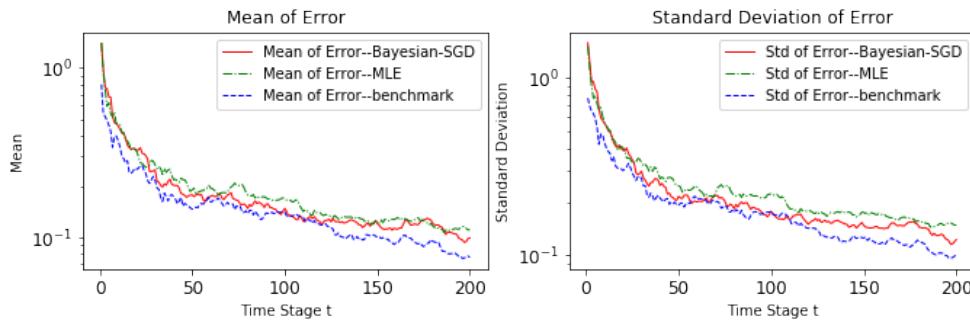


FIG. 1. Mean and standard deviation of $|x_t - x^*|$ of 100 runs of Algorithm 2.1 (Bayesian-SGD), MLE, and the benchmark algorithm in a univariate example. Note: color appears only in the online article.

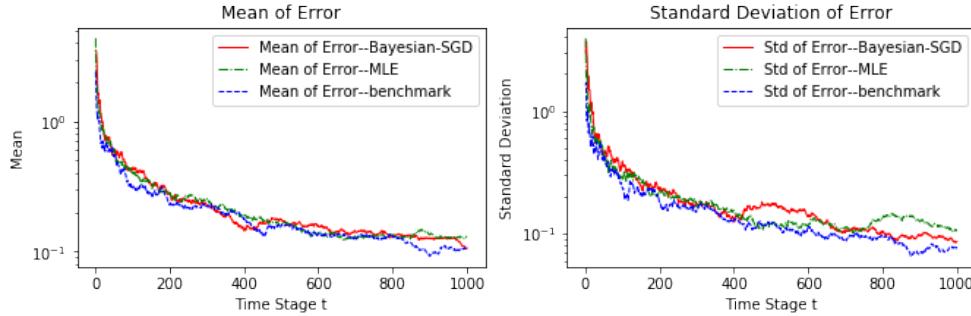


FIG. 2. Mean and standard deviation of $\|x_t - x^*\|_2$ of 100 runs of Algorithm 2.1 (Bayesian-SGD), MLE, and benchmark algorithm in a multivariate example. Note: color appears only in the online article.

the problem. The mean and standard deviation of the solution error $\|x_t - x^*\|_2$ over time are shown in Figure 2, from which we can draw the same conclusion as in the univariate setting.

4.1.2. Decision-dependent uncertainty. We carry out numerical experiments on a simple quadratic problem in a univariate setting: $h(x, \xi) = (x - 5)^2 + 0.5\xi x$, where $\xi \sim \mathcal{N}(x + \theta^c, \sigma^2)$. The parameters are as follows: $\sigma = 4$, $\theta^c = 4$, $D = 1$, $K = 1$, $\Theta = \{1, 2, \dots, 30\}$, $a_t = \frac{2}{t+5}$. It is easy to check $H(x, \theta) = (x - 5)^2 + 0.5(x + \theta^c)x = 1.5x^2 - 8x + 25$ and the true optimal decision is $x^* = \frac{8}{3}$. The gradient estimator in Algorithm 2.2 at each time t is $\nabla_x h(x_t, \xi_t) + h(x_t, \xi_t) \frac{\nabla_x \hat{f}_t(\xi_t; x_t)}{\hat{f}_t(\xi_t; x_t)}$, which can be computed as $(2x_t - 10 + 0.5\xi) + ((x_t - 5)^2 + 0.5\xi_t x_t) \frac{\sum_\theta \pi_t(\theta) \cdot \nabla_x f(\xi_t; x_t, \theta_t)}{\sum_\theta \pi_t(\theta) \cdot f(\xi_t; x_t, \theta_t)}$, where $f(\xi; x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\xi - (x + \theta))^2}{2\sigma^2}\right)$.

We use the uniform distribution on Θ as the prior distribution and set the initial solution $x_1 = 0$. We run Algorithm 2.2 and the benchmark algorithm for 100 times on the problem. Note that the MLE method in [57] is not applicable for the decision-dependent case. The mean and standard deviation of the solution error $|x_t - x^*|$ over time are shown in Figure 3. We further show the convergence of the posterior distribution under different data batch sizes D in Figure 4. Note that the benchmark algorithm (without parameter uncertainty) can be viewed as Algorithm 2.2 with $D = \infty$. The observations from Figures 3 and 4 are summarized as follows.

- With decreasing step size, the solution sequence in Algorithm 2.2 converges to the true optimal solution.
- Figure 4 shows that as we observe more data at each time stage, the Bayesian posterior distribution converges faster to the delta function concentrated on the true parameter θ^c .
- There is no significant difference in the convergence rate of Algorithm 2.2 under different data batch sizes, even though the posterior distribution converges faster with larger data batch size. It implies that the Bayesian average of the objective function (2.7) in this example is a good estimate of the true objective function despite the inaccuracy of the posterior distribution at the beginning time stages.

We then carry out numerical experiments on a quadratic problem in a multivariate setting, $h(x, \xi) = (x_1 - 1)^2 + (x_2 - 2)^2 + \xi$, where $x = (x_1, x_2)$ and ξ follows an exponential distribution with mean $(x_1 - 2)^2 + \theta^c$. The parameters are as follows:

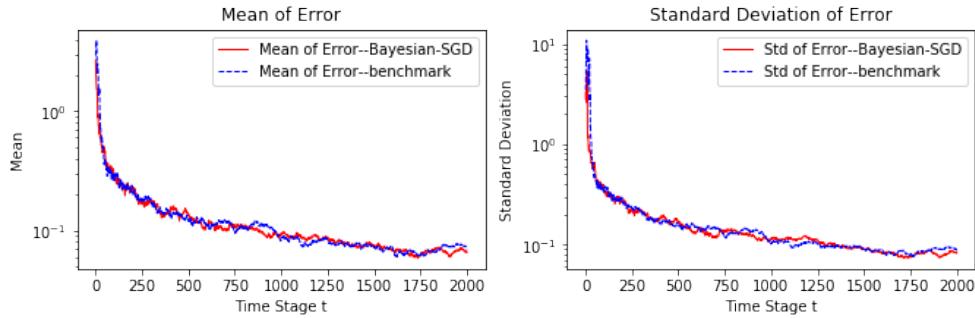


FIG. 3. Mean and standard deviation of $\|x_t - x^*\|_2$ of 100 runs of Algorithm 2.2 (Bayesian-SGD) and the benchmark algorithm in a univariate example. Note: color appears only in the online article.

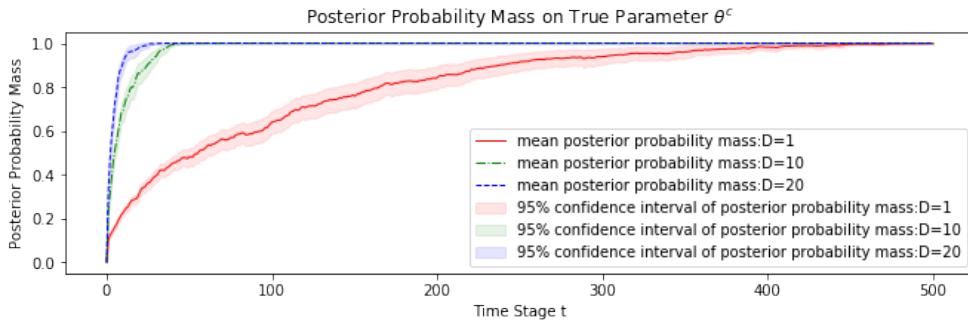


FIG. 4. Mean and 95% confidence interval of $\pi_t(\theta^c)$ of 100 runs of Algorithm 2.2 (Bayesian-SGD) under different data batch sizes. Note: color appears only in the online article.

$\theta^c = 4$, $D = 1$, $K = 1$, $\Theta = \{1, 2, \dots, 20\}$, $a_t = \frac{2}{t+5}$. It is easy to check $H(x, \theta^c) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 2x_1 - 4x_2 + 9$, and the true optimal decision is taken at $x^* = (\frac{4}{3}, \frac{5}{3})$. The gradient estimator in Algorithm 2.2 can be computed as $(2x_1 - 2, 2x_2 - 4) + ((x_1 - 1)^2 + (x_2 - 2)^2 + \xi) \frac{\sum_{\theta} \pi_t(\theta) \cdot \nabla_x f(\xi_t; x_t, \theta_t)}{\sum_{\theta} \pi_t(\theta) \cdot f(\xi_t; x_t, \theta_t)}$. Recall that $f(\xi; x, \theta) = \frac{1}{(x_1 - x_2)^2 + \theta} \exp - \frac{\xi_t}{(x_1 - x_2)^2 + \theta}$. We use the uniform distribution on Θ as the prior distribution and set the initial solution $x_1 = (5, 5)$. We run Algorithm 2.2 and the benchmark algorithm for 100 times on the problem. The mean and standard deviation of the solution error $\|x_t - x^*\|_2$ over time are shown in Figure 5, from which we can draw the same conclusion as the univariate setting.

4.2. Multi-item newsvendor problem. We consider a multi-item newsvendor problem and its variant with decision-dependent uncertainty. In the multi-item newsvendor problem, there are $d = 3$ different kinds of newspapers, and a newsboy orders $x \in \mathbb{R}_{\geq 0}^d$ units of newspaper to replenish the inventory at the beginning of a selling season. We assume $0 \leq x_i \leq M_i$, where M_i is the inventory capacity for newspaper $i \in [d]$. During the selling season, the newsboy observes customer demands, which are observations of a random vector $\xi \in (-\infty, \infty)^d$ following an unknown joint distribution F . Negative demand implies that some customers may have bought the newspaper somewhere else and drop it off after reading. The cost of purchasing a newspaper is c per unit, and the selling price is p per unit. At the end of the selling season the unsold newspaper has a salvage value of s per unit. Note that c, p, s are all 3-dimensional

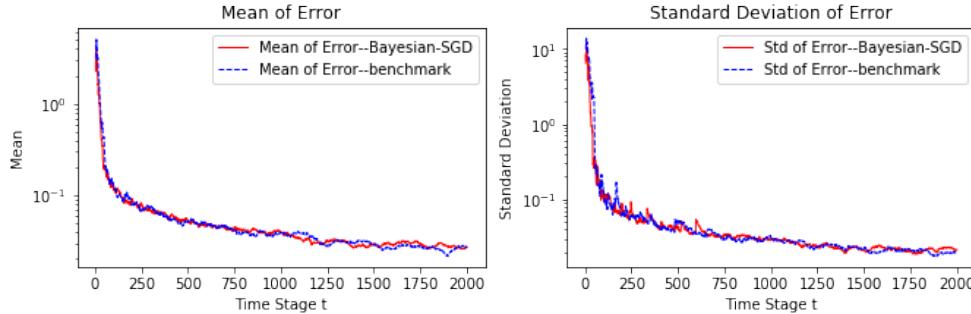


FIG. 5. Mean and standard deviation of $\|x_t - x^*\|_2$ of 100 runs of Algorithm 2.2 (Bayesian-SGD) and the benchmark algorithm in a multivariate example. Note: color appears only in the online article.

vectors. Also note that there is no replenishment of newspaper during the selling season. The cost function is given by $h(x, \xi) = c^T x - p^T \min(x, \max(0, \xi)) - s^T \max(0, x - \xi)$. Both min and max are elementwise operators. The newsboy aims to choose the amount x that minimizes the expected cost, where the expectation is taken w.r.t. the distribution of ξ .

4.2.1. Decision-independent uncertainty. We first consider the multi-item newsvendor problem with the decision-independent input uncertainty. We assume ξ follows a multivariate normal distribution with mean θ_μ^c and covariance matrix θ_Σ^c . Note that in this problem we have 9 unknown parameters, i.e., 3 mean parameters θ_μ^c , 3 variance parameters, and 3 correlation parameters $\theta_\Sigma^c := (\theta_{\text{var}}^c, \theta_{\text{corr}}^c)$. At each time t , the gradient estimator in Algorithm 2.1 is $\nabla_x h(x, \xi) = \begin{cases} c - p, x \leq \xi \\ c - s, x > \xi \end{cases}$. The parameters are as follows, $\theta_\mu^c = (10, 15, 20)$, $\theta_{\text{var}}^c = (3, 6, 9)$, $\theta_{\text{corr}}^c = (0.1, 0.3, 0.5)$, thus the true covariance matrix is $((3, 0.42, 1.56), (0.42, 6, 3.67), (1.56, 3.67, 9))$; parameter space $\Theta_\mu = \{5, 10, 15, 20, 25\}^3$, $\Theta_{\text{var}} = \{1, 3, 6, 9, 12\}^3$, $\Theta_{\text{corr}} = \{0.1, 0.2, 0.3, 0.4, 0.5\}^3$; $D = 2$, $K = 1$, $M = (100, 100, 100)$, $c = (2, 4, 6)$, $p = (4, 6, 8)$, $s = (1, 2, 3)$, $a_t = \frac{2}{t+5}$. We denote by x^* the optimal decision under the true parameters. We use the uniform distribution on Θ as the prior distribution and set the initial solution $x_1 = (15, 15, 15)$. We run all three algorithms (Algorithm 2.1, benchmark, MLE) for 100 times on the problem. The mean and standard deviation of the solution error $\|x_t - x^*\|_2$ over time are shown in Figure 6. We have similar observations as for the synthetic quadratic problem.

4.2.2. Decision-dependent uncertainty. We then consider the multi-item newsvendor problem with the decision-dependent input uncertainty, where the customer demand depends on the order amount x of the inventory. We follow the setting in [3], in which high inventory stimulates demand. We assume the demand ξ follows a multivariate normal distribution with mean $\theta_\mu^c + \alpha x^\beta$ and covariance matrix θ_Σ^c , where $\alpha > 0$, $0 < \beta < 1$ are vectors and $(\cdot)^\beta$ is an elementwise operator. Note that the mean function admits diminishing marginal utility, which says that the marginal increase in the mean demand diminishes as the inventory level increases. The gradient estimator in Algorithm 2.2 at each time stage t is given by

$$\nabla_x h(x_t, \xi_t) + h(x_t, \xi_t) \frac{\sum_\theta \pi_t(\theta) \cdot \nabla_x f(\xi_t; x_t, \theta_t)}{\sum_\theta \pi_t(\theta) \cdot f(\xi_t; x_t, \theta_t)}.$$

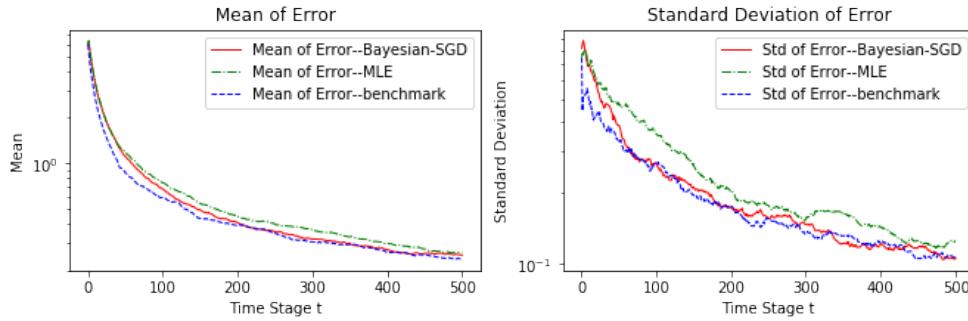


FIG. 6. Mean and standard deviation of $\|x_t - x^*\|_2$ of 100 runs of Algorithm 2.1 (Bayesian-SGD), MLE, and the benchmark algorithm in the multi-item newsvendor problem with decision-independent data. Note: color appears only in the online article.

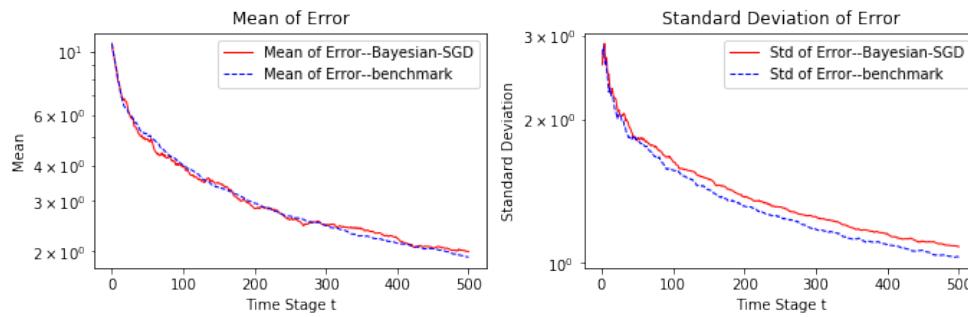


FIG. 7. Mean and standard deviation of $\|x_t - x^*\|_2$ of 100 runs of Algorithm 2.2 (Bayesian-SGD) and the benchmark algorithm in the multi-item newsvendor problem with decision-dependent data. Note: color appears only in the online article.

$f(\xi; x, \theta) = \frac{\exp(-\frac{1}{2}(\xi - (\theta_\mu + \alpha x^\beta))^T \theta_\Sigma^{-1}(\xi - (\theta_\mu + \alpha x^\beta)))}{\sqrt{(2\pi)^d |\theta_\Sigma|}}$, $\nabla_x f(\xi; x, \theta) = f(\xi; x, \theta) \theta_\Sigma^{-1}(\xi - (\theta_\mu + \alpha x^\beta)) \alpha \beta x^{\beta-1}$. The parameters are as follows: $\theta_\mu^c = (10, 15, 20)$, $\theta_{\text{var}}^c = (3, 6, 9)$, $\theta_{\text{corr}}^c = (0.1, 0.3, 0.5)$, the true covariance matrix is $((3, 0.42, 1.56), (0.42, 6, 3.67), (1.56, 3.67, 9))$. $\Theta_\mu = \{5, 10, 15, 20, 25\}^3$, $\Theta_{\text{var}} = \{1, 3, 6, 9, 12\}^3$, $\Theta_{\text{corr}} = \{0.1, 0.2, 0.3, 0.4, 0.5\}^3$. $D = 2$, $K = 1$, $M = (100, 100, 100)$, $c = (2, 4, 6)$, $p = (4, 6, 8)$, $s = (1, 2, 3)$, $\alpha = 1$, $\beta = 0.5$, $a_t = \frac{2}{t+5}$. We denote by x^* the optimal decision under the true parameters. We use the uniform distribution on Θ as the prior distribution and set the initial solution $x_1 = (15, 15, 15)$. We run Algorithm 2.2 and the benchmark algorithm for 100 times on the problem. The mean and standard deviation of the solution error $\|x_t - x^*\|_2$ over time are shown in Figure 7. We have similar observations as for the synthetic quadratic problem.

As a final note, the good performance of our proposed algorithms on the multi-dimensional newsvendor problem shows promise of the applicability of our proposed approaches to large-scale problems. However, it should be noted that most of the computational time is devoted to the posterior updating, especially for the high-dimensional problem where there is no conjugate prior. It would be interesting to adapt our algorithms to such a high-dimensional setup, where we could leverage the recent theoretical results of Bayesian procedures in high dimensions (cf. [19, 12]).

5. Conclusions. In this paper, we propose a Bayesian-SGD approach to stochastic optimization with streaming input data, and present two algorithms for decision-independent and decision-dependent uncertainty, respectively. We show the asymptotic convergence of both algorithms, and derive the convergence rate in the decision-independent case based on the nonasymptotic analysis of the Bayesian estimate. Our consistency result of Bayesian posterior distribution with decision-dependent input data could be of independent interest to Bayes estimation. Note that our approach can be viewed as an online extension of the BRO framework [72, 64], and it would be interesting to adapt our approach to other risk functionals (such as value-at-risk and conditional value-at-risk) with respect to the unknown distributional parameter.

Appendix A. Proof of Lemma 3.10.

Proof. Define the Hellinger distance between θ_1 and θ_2 as

$$d(\theta_1, \theta_2) = \sqrt{\frac{1}{2} \int_{\mathcal{Y}} (\sqrt{f(y; \theta_1)} - \sqrt{f(y; \theta_2)})^2}.$$

One can easily verify that there exists a constant A such that $\|\theta_1 - \theta_2\| \leq Ad(\theta_1, \theta_2)$, where $\|\cdot\|$ is the Euclidean norm. Let $B_k^t = B(\theta^c, k/\sqrt{Dt})$ be a ball centered at θ^c with radius k/\sqrt{Dt} under distance d . Since Θ is finite, we can directly apply Proposition 1 in [8]. Then for $t \leq T, \epsilon, \delta \in (0, 1)$ with probability at least $1 - \frac{6\delta}{\pi^2 t^2}$ with respect to $\mathbb{P}_{\theta^c}^t$, we have

$$\pi_t(B_{k(t)}^t) \geq 1 - \epsilon,$$

where

$$k(t) = \inf \left\{ j \geq 1 \left| \sum_{i \geq j} |\Theta| e^{-i^2} \leq \frac{6\delta}{\pi^2 t^2} \sqrt{\epsilon \pi_0(\theta^c)} \right. \right\}.$$

Note that since $\sum_{i \geq j} e^{-i^2} \leq \frac{e}{e-1} e^{-j^2}$, we can set $k(t)$ to be the solution of the next equation:

$$\frac{e}{e-1} |\Theta| e^{-k(t)^2} = \frac{6\delta}{\pi^2 t^2} \sqrt{\epsilon \pi_0(\theta^c)}.$$

By a simple calculation, we have $k(t) = \sqrt{\log \frac{e|\Theta|\pi^2 t^2}{6\delta(e-1)\sqrt{\epsilon\pi_0(\theta^c)}}}$. Now we are ready to bound the bias in the gradient estimator:

$$\begin{aligned} & \|\mathbb{E}_{\pi_t} \nabla_x H(x, \theta) - \mathbb{E}_{\pi_t} \nabla_x H(x, \theta^c)\|_2^2 \\ &= \left\| \int (\nabla_x H(x, \theta) - \nabla_x H(x, \theta^c)) \pi_t(\theta) d\theta \right\|_2^2 \\ &\leq \int \|(\nabla_x H(x, \theta) - \nabla_x H(x, \theta^c))\|_2^2 \pi_t(\theta) d\theta \\ &\leq \int L_H^2 \|\theta - \theta^c\|_2^2 \pi_t(\theta) d\theta \\ &= \int_{B_{k(t)}^t} L_H^2 \|\theta - \theta^c\|_2^2 \pi_t(\theta) d\theta + \int_{(B_{k(t)}^t)^c} L_H^2 \|\theta - \theta^c\|_2^2 \pi_t(\theta) d\theta \end{aligned}$$

$$\begin{aligned} &\leq A^2 L_H^2 \frac{k(t)^2}{Dt} \int_{B_{k(t)}^t} \pi_t(\theta) d\theta + L_H^2 \max_{\theta \in \Theta} \|\theta - \theta^c\|_2^2 \int_{(B_{k(t)}^t)^c} \pi_t(\theta) d\theta \\ &\leq A^2 L_H^2 \frac{k(t)^2}{Dt} + L_H^2 \max_{\theta \in \Theta} \|\theta - \theta^c\|_2^2 \epsilon. \end{aligned}$$

Recall that D is the data batch size. Take $\epsilon = \frac{1}{Dt}$ and note that $k(t) = \sqrt{\log \frac{e|\Theta|\pi^2 t^2 \sqrt{Dt}}{6\delta(e-1)\sqrt{\pi_0(\theta^c)}}}$. We further have

$$\begin{aligned} \|\mathbb{E}_{\pi_t} \nabla_x H(x, \theta) - \mathbb{E}_{\pi_t} \nabla_x H(x, \theta^c)\|_2^2 &\leq A^2 L_H^2 \frac{k(t)^2}{Dt} + L_H^2 \max_{\theta \in \Theta} \|\theta - \theta^c\|_2^2 \epsilon \\ &\leq 2A^2 L_H^2 \max_{\theta \in \Theta} \|\theta - \theta^c\|_2^2 \frac{\log \frac{e|\Theta|\pi^2 t^2 \sqrt{Dt}}{6\delta(e-1)\sqrt{\pi_0(\theta^c)}}}{Dt} \\ &= O\left(\frac{\log Dt + \log \frac{1}{\delta}}{Dt}\right). \end{aligned}$$

Let \mathcal{E}_t denote the event that the above inequality holds, and \mathcal{E}_t^c denote the complement event. Then we have $\mathbb{P}(\mathcal{E}_t^c) \leq \frac{6\delta}{\pi^2 t^2}$. Therefore,

$$\begin{aligned} \mathbb{P}(\bigcap_{t=1}^{\infty} \mathcal{E}_t) &= 1 - \mathbb{P}\left(\bigcup_{t=1}^{\infty} \mathcal{E}_t^c\right) \\ &\geq 1 - \sum_{t=1}^{\infty} \mathbb{P}(\mathcal{E}_t^c) \quad (\text{union bound}) \\ &\geq 1 - \sum_{t=1}^{\infty} \frac{6\delta}{\pi^2 t^2} \\ &= 1 - \delta. \end{aligned} \quad \square$$

Appendix B. Proof of Theorem 3.11.

Proof. By the update (3.1), we know that for any $t \leq T$,

$$\begin{aligned} x_{t+1} &= x_t - a_t \nabla_x H(x_t, \theta^c) - a_t [\mathbb{E}_{\pi_t} \nabla_x H(x_t, \theta) - \nabla_x H(x_t, \theta^c)] \\ &\quad - a_t [\nabla_x h(x_t, \xi_t) - \mathbb{E}_{\pi_t} \nabla_x H(x_t, \theta)] \\ &= x_t - a_t \nabla_x H(x_t, \theta^c) - a_t B_t - a_t N_t, \end{aligned}$$

where B_t is the bias and N_t is the noise. By Lemma 3.10, we know $\mathbb{E}[\|B_t\|_2^2] \leq C_1 \frac{\log Dt + \log \frac{1}{\delta}}{Dt}$. From Assumption 3.9 we have $\mathbb{E}[\|N_t\|_2^2] \leq \sigma^2$. By the proof of Lemma 2 in [2], we know that

$$\mathbb{E}[H(x_{t+1}, \theta^c)] - H(x_t, \theta^c) \leq -\frac{a_t}{2} \|\nabla_x H(x_t, \theta^c)\|_2^2 + \frac{a_t}{2} C_1 \frac{\log Dt + \log \frac{1}{\delta}}{Dt} + \frac{a_t^2}{2} L_h \sigma^2.$$

Rearranging the terms in the inequality above, summing over t from 1 to T , and noting that $H(x_t, \theta^c) \leq \min_{x \in \mathcal{X}} H(x, \theta^c) \forall t$, we have

$$\begin{aligned} \sum_{t=1}^T a_t \mathbb{E}[\|\nabla_x H(x_t, \theta^c)\|_2^2] &\leq 2(H(x_1, \theta^c) - \min_{x \in \mathcal{X}} H(x, \theta^c)) + C_1 \sum_{t=1}^T a_t \frac{\log Dt + \log \frac{1}{\delta}}{Dt} \\ &\quad + L_h \sigma^2 \sum_{t=1}^T a_t^2. \end{aligned}$$

Dividing both sides of the above inequality by $\sum_{t=1}^T a_t$, and noting that

$$\mathbb{E}[\|\nabla_x H(z_T, \theta^c)\|_2^2] = \frac{1}{\sum_{t=1}^T a_t} \sum_{t=1}^T a_t \mathbb{E}[\|\nabla_x H(x_t, \theta^c)\|_2^2],$$

we have

$$\begin{aligned} \mathbb{E}[\|\nabla_x H(z_T, \theta^c)\|_2^2] &\leq \frac{1}{\sum_{t=1}^T a_t} \left[2(H(x_1, \theta^c) - \min_{x \in \mathcal{X}} H(x, \theta^c)) \right. \\ &\quad \left. + C_1 \sum_{t=1}^T a_t \frac{\log Dt + \log \frac{1}{\delta}}{Dt} + L_h \sigma^2 \sum_{t=1}^T a_t^2 \right]. \end{aligned}$$

(i) $a_t = \frac{a}{\sqrt{T}}$ $\forall t \leq T$ for some constant $a < \frac{\sqrt{T}}{L_h}$. Note that $\sum_{t=1}^T \frac{1}{t} \leq \log T + 1$ and $\sum_{t=1}^T \frac{\log t}{t} \leq \log(\log T + 1)$. Then

$$\begin{aligned} &\mathbb{E}[\|\nabla_x H(z_T, \theta^c)\|_2^2] \\ &\leq \frac{2(H(x_1, \theta^c) - \min_x H(x, \theta^c))}{a\sqrt{T}} + \frac{C_1(\log D - \log \delta)(\log T + 1)}{L_h DT} + \frac{C_1 \log T (\log T + 1)}{L_h DT} \\ &= \frac{2(H(x_1, \theta^c) - \min_x H(x, \theta^c))}{a\sqrt{T}} + \frac{C_1(\log D - \log \delta)}{L_h DT} + \frac{C_1(\log D - \log \delta) \log T}{L_h DT} \\ &\quad + \frac{C_1 \log^2 T}{L_h DT} + \frac{L_h a \sigma^2}{\sqrt{T}}. \end{aligned}$$

(ii) $a_t = \frac{a}{t}$ $\forall t \leq T$ for some constant $a < \frac{1}{L_h}$. Let $M_T = \sum_{t=1}^T \frac{1}{t}$. Note that

$$\sum_{t=1}^T \frac{\log t}{t^2} < \sum_{t=1}^{\infty} \frac{\log t}{t^2} = \frac{\pi^2}{6} (12 \ln A - \gamma - \ln 2\pi) < 1,$$

where $A \approx 1.28$ is the Glaisher–Kinkelin constant and $\gamma \approx 0.58$ is the Euler–Mascheroni constant. Then we have

$$\begin{aligned} &\mathbb{E}[\|\nabla_x H(z_T, \theta^c)\|_2^2] \\ &\leq \frac{2(H(x_1, \theta^c) - \min_{x \in \mathcal{X}} H(x, \theta^c))}{a M_T} + \frac{C_1}{M_T} \sum_{t=1}^T \frac{\log Dt + \log \frac{1}{\delta}}{Dt^2} + \sum_{t=1}^T \frac{L_h a \sigma^2}{M_T t^2} \\ &\leq \left[\frac{2(H(x_1, \theta^c) - \min_{x \in \mathcal{X}} H(x, \theta^c))}{a} + \frac{6C_1 + \pi^2 C_1 (\log D - \log \delta)}{6D} \right. \\ &\quad \left. + \frac{\pi^2 L_h a \sigma^2}{6} \right] \frac{1}{\log T}. \end{aligned}$$

(iii) $a_t = \frac{a}{\sqrt{t}}$ $\forall t \leq T$ for some constant $a < \frac{1}{L_h}$. Let $Q_t = \sum_{t=1}^T \frac{1}{\sqrt{t}}$. Note that $\sum_{t=1}^{\infty} \frac{1}{t\sqrt{t}} = \zeta(1.5) \approx 2.61 < 3$, $\sum_{t=1}^{\infty} \frac{\log t}{t\sqrt{t}} < 4$, $\sum_{t=1}^T \frac{1}{\sqrt{t}} \geq \sqrt{T}$, where $\zeta(\cdot)$ is the Riemann's zeta function. Then we have

$$\begin{aligned}
& \mathbb{E}[\|\nabla_x H(z_T, \theta^c)\|_2^2] \\
& \leq \frac{2(H(x_1, \theta^c) - \min_x H(x, \theta^c))}{aQ_T} + \frac{C_1(\log D - \log \delta)}{DQ_T} \sum_{t=1}^T \frac{1}{t\sqrt{t}} \\
& \quad + \frac{C_1}{DQ_T} \sum_{t=1}^T \frac{\log t}{t\sqrt{t}} + \frac{L_h a \sigma^2}{Q_T} \sum_{t=1}^T \frac{1}{t} \\
& \leq \left[\frac{2(H(x_1, \theta^c) - \min_x H(x, \theta^c))}{a\sqrt{T}} + \frac{3C_1(\log D - \log \delta) + 4C_1}{D\sqrt{T}} + \frac{L_h a \sigma^2}{\sqrt{T}} \right] \\
& \quad + \frac{L_h a \sigma^2 \log T}{\sqrt{T}}.
\end{aligned}$$

□

Appendix C. Proof of Lemma 3.13.

Proof. Define $w_t = -\log \pi_t(\theta^c)$. One can easily verify that $w_t \geq 0$. Then we have

$$\begin{aligned}
\mathbb{E}[w_{t+1}] &= \mathbb{E}[\mathbb{E}[w_{t+1} | \mathcal{F}_t, x_{t+1}]] \\
&= \mathbb{E}\left[\mathbb{E}\left[-\log \frac{\pi_t(\theta^c) f(y_{t+1}; x_{t+1}, \theta^c)}{\sum_\theta \pi_t(\theta) f(y_{t+1}; x_{t+1}, \theta)} | \mathcal{F}_t, x_{t+1}\right]\right] \\
&= \mathbb{E}\left[-\log \pi_t(\theta^c) - \mathbb{E}\left[\log \frac{f(y_{t+1}; x_{t+1}, \theta^c)}{\sum_\theta \pi_t(\theta) f(y_{t+1}; x_{t+1}, \theta)} | \mathcal{F}_t, x_{t+1}\right]\right] \\
&= \mathbb{E}[w_t] - \mathbb{E}[D_{KL}(f^*(\cdot; x_{t+1}) || \hat{f}_t(\cdot; x_{t+1}))].
\end{aligned}$$

This implies that $\mathbb{E}[d_t] = \mathbb{E}[w_t] - \mathbb{E}[w_{t+1}]$. For any $T > 0$, we have

$$\sum_{t=0}^T \mathbb{E}[d_t] = \sum_{t=0}^T \mathbb{E}[w_t] - \mathbb{E}[w_{t+1}] = w_0 - \mathbb{E}[w_{T+1}] \leq w_0 < \infty.$$

Then we have $\sum_{t=0}^\infty \mathbb{E}[d_t] \leq w_0$. $\forall \epsilon > 0$, we have

$$\sum_{t=0}^\infty \mathbb{P}(d_t \geq \epsilon) \leq \frac{1}{\epsilon} \sum_{t=0}^\infty \mathbb{E}[d_t] < \infty.$$

By the Borel–Cantelli lemma, we know that $\mathbb{P}(d_t \geq \epsilon, i.o.) = 0$, where *i.o.* stands for infinitely often. It then implies $\lim_{t \rightarrow \infty} d_t = 0$, w.p.1 ($\mathbb{P}_{\theta^c}^\infty$). Moreover, since $d_t \geq 0$, by Tonelli's theorem, we have

$$\mathbb{E}\left[\sum_{t=0}^\infty d_t\right] = \sum_{t=0}^\infty \mathbb{E}[d_t] \leq w_0.$$

Since $\sum_{t=0}^\infty d_t$ has bounded expectation, it must be finite w.p.1 ($\mathbb{P}_{\theta^c}^\infty$). □

Appendix D. Proof of Proposition 3.16.

Proof. Without loss of generality, we assume that $\theta^c = \theta_1$. Recall that $f^*(\xi; x_{t+1}) = f^*(\xi; x_{t+1}, \theta_1)$ and $\hat{f}_t(\xi; x_{t+1}) = \sum_i \pi_t(\theta_i) f(\xi; x_{t+1}, \theta_i)$. Then we have

$$(D.1) \quad f^*(\xi; x_{t+1}) - \hat{f}_t(\xi; x_{t+1}) = (1 - \pi_t(\theta_1)) f(\xi; x_{t+1}, \theta_1) - \sum_{i>1} \pi_t(\theta_i) f(\xi; x_{t+1}, \theta_i).$$

Note that for any $t > 0$, $(\pi_t(\theta_1), \pi_t(\theta_2), \dots)$ is an infinitely dimensional bounded vector with all components in the interval $[0, 1]$ and sum up to 1 (normalized), we can take a

subsequence $\{\pi_{t_k}\}$ such that for each component j , $\pi_{t_k}(\theta_j)$ converges to a limit which is denoted by $\pi_\infty(\theta_j)$, which is also known as weak convergence (of a deterministic sequence). Next, we will show that $\pi_\infty(\theta)$ is a normalized vector. For any $j \in \mathbb{N}$, $\lim_{t_k \rightarrow \infty} \pi_{t_k}(\theta_j) = \pi_\infty(\theta_j)$, which is equivalent to

$$\forall \epsilon_j > 0, \exists N \in \mathbb{N}, \text{s.t.} \forall n \geq N, |\pi_\infty(\theta_j) - \pi_n(\theta_j)| \leq \epsilon_j.$$

Therefore, we have

$$(D.2) \quad -\epsilon_j < \pi_\infty(\theta_j) - \pi_n(\theta_j) < \epsilon_j, \quad j = 1, 2, \dots$$

According to the Bayesian update rule, we know $\sum_{j=1}^{\infty} \pi_n(\theta_j) = 1$. It then follows that $\forall \epsilon > 0$, taking $\epsilon_j = \frac{\epsilon}{2^j}$ and summing over (D.2) for all $j \in \mathbb{N}$, we get

$$-\left(\frac{\epsilon}{2^1} + \frac{\epsilon}{2^2} + \dots\right) < \sum_{j=1}^{\infty} \pi_\infty(\theta_j) - 1 < \left(\frac{\epsilon}{2^1} + \frac{\epsilon}{2^2} + \dots\right),$$

which indicates $\forall \epsilon > 0$, $|\sum_{j=1}^{\infty} \pi_\infty(\theta_j) - 1| < \epsilon$, and it implies that $\sum_{j=1}^{\infty} \pi_\infty(\theta_j) = 1$. So the limit is also a valid probability simplex. Since every weakly convergent sequence in L^1 is strongly convergent (cf. Chapter 2 in [50]), we can take any convergent subsequence of $\{\pi_{t_k}\}$ with limit (p_1^*, p_2^*, \dots) . Since \mathcal{X} is also bounded, from this subsequence, we could take a further subsequence $\{\pi_{\tau_k}\}$ with time stage τ_1, τ_2, \dots , such that $\{x_{\tau_k}\}$ converges to some x' . Then taking the limit over (D.1) along τ_1, τ_2, \dots , we have

$$f^*(\xi; x_{\tau_k}) - \hat{f}_{\tau_k}(\xi; x_{\tau_k}) \rightarrow (1 - (p_1^*))f(\xi; x', \theta_1) - \sum_{i>1} p_i^* f(\xi; x', \theta_i).$$

Moreover, since K-L divergence dominates total variation distance between two distributions, we have

$$(D.3) \quad \int_{\Xi} \left| f^*(\xi; x_{t+1}) - \hat{f}_t(\xi; x_{t+1}) \right| d\xi \leq d_t.$$

From (D.3) and Lemma 3.13, we know that $\int_{\Xi} \left| f^*(\xi; x_{t+1}) - \hat{f}_t(\xi; x_{t+1}) \right| d\xi \rightarrow 0$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$). By the DCT, we have

$$\int_{\Xi} \left| (1 - (p_1^*))f(\xi; x', \theta_1) - \sum_{i>1} p_i^* f(\xi; x', \theta_i) \right| d\xi = 0,$$

which implies:

$$(1 - (p_1^*))f(\xi; x', \theta_1) - \sum_{i>1} p_i^* f(\xi; x', \theta_i) = 0 \quad \forall \xi.$$

By linear independence, we know $p_1^* = 1, p_2^* = p_3^* = \dots = 0$. Since every convergent subsequence of $\{(\pi_t(\theta_1), \pi_t(\theta_2), \dots)\}_t$ has the same limit, we have $\pi_t \Rightarrow \delta_{\theta^c}$ w.p.1\break ($\mathbb{P}_{\theta^c}^{\infty}$). \square

Appendix E. Derivation of unbiased estimator in decision-dependent case.

$$\begin{aligned}
\nabla_x \mathbb{E}_{\pi_t}[H(x, \theta)] &= \mathbb{E}_{\pi_t}[\nabla_x \mathbb{E}_{f(\cdot; x, \theta)}[h(x, \xi)]] \\
&= \mathbb{E}_{\pi_t}\left[\int_{\Xi} \nabla_x h(x, \xi) f(\xi; x, \theta) d\xi\right] + \mathbb{E}_{\pi_t}\left[\int_{\Xi} h(x, \xi) \nabla_x f(\xi; x, \theta) d\xi\right] \\
&= \mathbb{E}_{\pi_t}[\mathbb{E}_{f(\cdot; x, \theta)}[\nabla_x h(x, \xi)]] + \int_{\Theta} \left(\int_{\Xi} h(x, \xi) \nabla_x f(\xi; x, \theta) d\xi\right) \pi_t(\theta) d\theta \\
&= \mathbb{E}_{\pi_t}[\mathbb{E}_{f(\cdot; x, \theta)}[\nabla_x h(x, \xi)]] + \int_{\Xi} h(x, \xi) \left(\int_{\Theta} \pi_t(\theta) \nabla_x f(\xi; x, \theta) d\theta\right) d\xi \\
&= \mathbb{E}_{\pi_t}[\mathbb{E}_{f(\cdot; x, \theta)}[\nabla_x h(x, \xi)]] + \int_{\Xi} h(x, \xi) \nabla_x \hat{f}(\xi; x) d\xi \\
&= \mathbb{E}_{\pi_t}[\mathbb{E}_{f(\cdot; x, \theta)}[\nabla_x h(x, \xi)]] + \int_{\Xi} h(x, \xi) \frac{\nabla_x \hat{f}(\cdot; x)}{\hat{f}(\cdot; x)} \hat{f}(\cdot; x) d\xi \\
&= \mathbb{E}_{\pi_t}[\mathbb{E}_{f(\cdot; x, \theta)}[\nabla_x h(x, \xi)]] + \mathbb{E}_{\hat{f}(\cdot; x)}\left[h(x, \xi) \frac{\nabla_x \hat{f}(\cdot; x)}{\hat{f}(\cdot; x)}\right] \\
&= \mathbb{E}_{\pi_t}\left[\mathbb{E}_{f(\cdot; x, \theta)}[\nabla_x h(x, \xi) + h(x, \xi) \frac{\nabla_x \hat{f}(\cdot; x)}{\hat{f}(\cdot; x)}]\right].
\end{aligned}$$

From Assumptions 3.4 and 3.18, we know that both the objective function $h(x, \xi)$ and the density function $f(\xi; x, \theta)$ are C^1 -smooth. The Lipschitz continuous gradient implies both $h(x, \xi)$ and $f(\xi; x, \theta)$ are integrable functions; $\nabla_x h(x, \xi)$ and $\nabla_x f(x, \xi)$ are dominated by some integrable functions. Using the chain rule, we have $\nabla_x h(x, \xi) f(\xi; x, \theta) = \nabla_x h(x, \xi) \cdot f(\xi; x, \theta) + h(x, \xi) \cdot \nabla_x f(\xi; x, \theta)$, and thus $\nabla_x h(x, \xi) f(\xi; x, \theta)$ is dominated by some integrable function. The second equality holds as the interchange between expectation and differentiation is justified by the DCT. The first equality is also justified by the DCT in a similar manner. Also note that since $h(x, \xi) \nabla_x f(\xi; x, \theta)$ is dominated by some integrable function, it is also absolutely integrable, hence the fourth equality is justified by the Fubini–Tonelli theorem.

Appendix F. Proof of Lemma 3.23.

Proof. We bound $|\beta_{t,1}|$ as follows:

$$\begin{aligned}
|\beta_{t,1}| &= |\mathbb{E}_{\hat{f}_t(\cdot; x_t)} \nabla_x h(x_t, \xi) - \mathbb{E}_{f^*(\cdot; x_t)} \nabla_x h(x_t, \xi)| \\
&\leq \max_{x, \xi} |\nabla_x h(x, \xi)| \int_{\Xi} |f^*(\xi; x_t) - \hat{f}_t(\xi; x_t)| d\xi \\
&\leq L'_h \int_{\Xi} |f^*(\xi; x_t) - f^*(\xi; x_{t+1})| + |f^*(\xi; x_{t+1}) - \hat{f}_t(\xi; x_{t+1})| \\
&\quad + |\hat{f}_t(\xi; x_t) - \hat{f}_t(\xi; x_{t+1})| d\xi.
\end{aligned}$$

From Assumption 3.4 we know $h(x, \xi)$ is continuously differentiable, which implies it has a bounded gradient, such that $|\nabla_x h(x, \xi)| \leq L'_h$ for some $L'_h > 0$. From Assumption 3.18, we know $f(\xi; x, \theta)$ is continuously differentiable, which implies it has a bounded gradient, such that $|\nabla_x f(\xi; x, \theta)| \leq L'_f$ for some $L'_f > 0$. Therefore, for every $\xi \in \Xi$,

$$(F.1) \quad |f^*(\xi; x_t) - f^*(\xi; x_{t+1})| \leq L'_f |x_t - x_{t+1}| \leq L'_f D_f a_t,$$

$$(F.2) \quad |\hat{f}_t(\xi; x_t) - \hat{f}_t(\xi; x_{t+1})| \leq L'_f D_f a_t$$

for some $D_f > 0$. Let $q_t^*(\xi) = f^*(\xi; x_t) - f^*(\xi; x_{t+1})$. Since $\lim_{t \rightarrow \infty} a_t = 0$, and by (F.1), we have $\lim_{t \rightarrow \infty} |q_t^*(\xi)| = 0$ for every $\xi \in \Xi$. By the absolute value theorem, we have $\lim_{t \rightarrow \infty} q_t^*(\xi) = 0$ for every $\xi \in \Xi$, that is, q_t^* converges pointwise to 0. By the DCT, we have

$$(F.3) \quad \lim_{t \rightarrow \infty} \int_{\Xi} |f^*(\xi; x_t) - f^*(\xi; x_{t+1})| = 0.$$

Similarly, let $\hat{q}_t(\xi) = \hat{f}_t(\xi; x_t) - \hat{f}_t(\xi; x_{t+1})$. From Assumption 3.20, we have $\lim_{t \rightarrow \infty} a_t = 0$ and, by (F.2), we have $\lim_{t \rightarrow \infty} |\hat{q}_t(\xi)| = 0$ for every $\xi \in \Xi$. By the absolute value theorem, we have $\lim_{t \rightarrow \infty} \hat{q}_t(\xi) = 0$ for every $\xi \in \Xi$, that is, \hat{q}_t converges pointwise to 0. By the DCT, we have

$$(F.4) \quad \lim_{t \rightarrow \infty} \int_{\Xi} |\hat{f}_t(\xi; x_t) - \hat{f}_t(\xi; x_{t+1})| = 0.$$

Moreover, since K-L divergence dominates total variation distance between two distributions, we have

$$(F.5) \quad \int_{\Xi} |f^*(\xi; x_{t+1}) - \hat{f}_t(\xi; x_{t+1})| d\xi \leq d_t.$$

From Lemma 3.13, we have $\lim_{t \rightarrow \infty} d_t = 0$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$). Combining (F.3), (F.4), and (F.5) together, we know that $\lim_{t \rightarrow \infty} |\beta_{t,1}| = 0$ w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$). \square

Appendix G. Proof of Lemma 3.24.

Proof. We bound $|\beta_{t,2}|$ as follows. From Assumption 3.4 we know $h(x, \xi)$ is continuously differentiable, which implies it is an integrable function of ξ for every $x \in \mathcal{X}$. Thus, $\int_{\Xi} h(x, \xi) d\xi = U_h$ for some $-\infty < U_h < \infty$. From Assumption 3.18 we know $f(\xi; x, \theta)$ is continuously differentiable, which implies it has a bounded gradient, such that $|\nabla_x f(\xi; x, \theta)| \leq L'_f$ for some $L'_f > 0$.

$$\begin{aligned} |\beta_{t,2}| &= \left| \int_{\Xi} h(x_t, \xi) \nabla_x \hat{f}_t(\xi; x_t) d\xi - \int_{\Xi} h(x_t, \xi) \nabla_x f^*(\xi; x_t) d\xi \right| \\ &= \left| \int_{\Xi} h(x_t, \xi) \left(\nabla_x \hat{f}_t(\xi; x_t) - \nabla_x f^*(\xi; x_t) \right) d\xi \right| \\ &= \left| \int_{\Xi} h(x_t, \xi) \left(\sum_{\theta \in \Theta} (\pi_t(\theta) - \delta_{\theta^c}(\theta)) \nabla_x f(\xi; x_t, \theta) \right) d\xi \right| \\ &\leq |U_h| \cdot L'_f \left| \sum_{\theta \in \Theta} \pi_t(\theta) - \delta_{\theta^c}(\theta) \right| \rightarrow 0, \end{aligned}$$

w.p.1 ($\mathbb{P}_{\theta^c}^{\infty}$) as $t \rightarrow \infty$, using the consistency of $\pi_t(\theta)$ from Proposition 3.16. \square

REFERENCES

- [1] S. AGRAWAL AND N. GOYAL, *Analysis of Thompson sampling for the multi-armed bandit problem*, in Proceedings of the 25th Annual Conference on Learning Theory, Vol. 23, 2012, pp. 39.1–39.26.
- [2] A. AJALLOEIAN AND S. U. STICH, *Analysis of SGD with Biased Gradient Estimators*, in Workshop on “Beyond First Order Methods in ML Systems” at the 37th International Conference on Machine Learning, preprint, arXiv:2008.00051, 2020.
- [3] A. BALAKRISHNAN, M. S. PANGBURN, AND E. STAVRULAKI, “*Stack them high, let’em fly*”: Lot-sizing policies when inventories stimulate demand, *Manag. Sci.*, 50 (2004), pp. 630–644.
- [4] G. BAYRAKSAN AND D. K. LOVE, *Data-driven stochastic programming using phi-divergences*, in The Operations Research Revolution, INFORMS, Hanover, MD, 2015, pp. 1–19.
- [5] A. BEN-TAL AND M. TEBOULLE, *Penalty functions and duality in stochastic programming via φ -divergence functionals*, *Math. Oper. Res.*, 12 (1987), pp. 224–240.
- [6] L. BENKHEROUF, A. BOUMENIR, AND L. AGGOUN, *A stochastic inventory model with stock dependent demand items*, *J. Appl. Math. Stoch. Anal.*, 14 (2001), pp. 317–328.
- [7] D. BERTSIMAS, V. GUPTA, AND N. KALLUS, *Data-driven robust optimization*, *Math. Program.*, 167 (2018), pp. 235–292.
- [8] L. BIRGÉ, *About the non-asymptotic behaviour of Bayes estimators*, *J. Statist. Plann. Inference*, 166 (2015), pp. 67–77.
- [9] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer Science & Busi, New York, 2013.
- [10] L. BOTTOU, *Online learning and stochastic approximations*, in Online Learning in Neural Networks, 17 (1998), pp. 9–42.
- [11] S. BUBECK, O. DEKEL, T. KOREN, AND Y. PERES, *Bandit convex optimization: \sqrt{T} regret in one dimension*, in Proceedings of the 28th Annual Conference on Learning Theory, Vol. 40, 2015, pp. 266–278.
- [12] N. CHOPIN, S. GADAT, B. GUEDJ, A. GUYADER, AND E. VERNET, *On some recent advances on high dimensional Bayesian statistics*, *ESAIM Proc. Surveys*, 51 (2015), pp. 293–319.
- [13] J. CUTLER, M. DÍAZ, AND D. DRUSVYATSKIY, *Stochastic Approximation with Decision-Dependent Distributions: Asymptotic Normality and Optimality*, preprint, arXiv:2207.04173, 2022.
- [14] E. DELAGE AND Y. YE, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, *Oper. Res.*, 58 (2010), pp. 595–612.
- [15] D. DRUSVYATSKIY AND L. XIAO, *Stochastic optimization with decision-dependent distributions*, *Math. Oper. Res.*, 48 (2023), pp. 954–998.
- [16] J. DUCHI, E. HAZAN, AND Y. SINGER, *Adaptive subgradient methods for online learning and stochastic optimization*, *J. Mach. Learn. Res.*, 12 (2011), pp. 2121–2159.
- [17] J. DUPACOVÁ, *Optimization under exogenous and endogenous uncertainty*, in Proceedings of the 24th International conference on Mathematical Methods in Economics, 2006, pp. 131–136.
- [18] A. DURMUS, S. MAJEWSKI, AND B. MIASOJEDOW, *Analysis of Langevin Monte Carlo via convex optimization*, *J. Mach. Learn. Res.*, 20 (2019), pp. 2666–2711.
- [19] A. DURMUS AND E. MOULINES, *High-dimensional Bayesian inference via the unadjusted Langevin algorithm*, *Bernoulli*, 25 (2019), pp. 2854–2882.
- [20] A. DURMUS, G. O. ROBERTS, G. VILMART, AND K. C. ZYGALAKIS, *Fast Langevin based algorithm for MCMC in high dimensions*, *Ann. Appl. Probab.*, 27 (2017), pp. 2195–2237.
- [21] R. DURRETT, *Probability: Theory and Examples*, Cambridge Set. Stat. Probab. Math. 49, Cambridge University Press, Cambridge, 2019.
- [22] T. EKIN, N. G. POLSON, AND R. SOYER, *Augmented nested sampling for stochastic programs with recourse and endogenous uncertainty*, *Naval Res. Logist.*, 64 (2017), pp. 613–627.
- [23] D. L. ERMAK AND H. BUCKHOLZ, *Numerical integration of the Langevin equation: Monte Carlo simulation*, *J. Comput. Phys.*, 35 (1980), pp. 169–182.
- [24] P. M. ESFAHANI AND D. KUHN, *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations*, *Math. Program.*, 171 (2018), pp. 115–166.
- [25] W. FAN, L. J. HONG, AND X. ZHANG, *Distributionally robust selection of the best*, *Manag. Sci.*, 66 (2020), pp. 190–208.
- [26] M. C. FU, *What you should know about simulation and derivatives*, *Naval Res. Logist.*, 55 (2008), pp. 723–736.
- [27] S. GAO, H. XIAO, E. ZHOU, AND W. CHEN, *Robust ranking and selection with optimal computing budget allocation*, *Automatica J. IFAC*, 81 (2017), pp. 30–36.

- [28] S. GHADIMI AND G. LAN, *Stochastic first-and zeroth-order methods for nonconvex stochastic programming*, SIAM J. Optim., 23 (2013), pp. 2341–2368.
- [29] S. GHOSAL AND A. VAN DER VAART, *Convergence rates of posterior distributions for noniid observations*, Ann. Statist., 35 (2007), pp. 192–223.
- [30] B. GIRI, S. PAL, A. GOSWAMI, AND K. CHAUDHURI, *An inventory model for deteriorating items with stock-dependent demand rate*, European J. Oper. Res., 95 (1996), pp. 604–610.
- [31] V. GOEL AND I. E. GROSSMANN, *A class of stochastic programs with decision dependent uncertainty*, Math. Program., 108 (2006), pp. 355–394.
- [32] L. V. GREEN, S. SAVIN, AND N. SAVVA, “*Nursevendor problem*”: *Personnel staffing in the presence of endogenous absenteeism*, Manag. Sci., 59 (2013), pp. 2237–2256.
- [33] V. GUPTA, *Near-optimal Bayesian ambiguity sets for distributionally robust optimization*, Manag. Sci., 65 (2019), pp. 4242–4260.
- [34] E. HAZAN, A. RAKHILIN, AND P. BARTLETT, *Adaptive online gradient descent*, in *Advances in Neural Information Processing Systems*, Vol. 20, J. Platt, D. Koller, Y. Singer, and S. Roweis, eds., MIT Press, Cambridge, MA, 2007.
- [35] L. HELLEMO, P. I. BARTON, AND A. TOMASGARD, *Decision-dependent probabilities in stochastic programs with recourse*, Comput. Manag. Sci., 15 (2018), pp. 369–395.
- [36] Z. IZZO, L. YING, AND J. ZOU, *How to learn when data reacts to your model: Performatice gradient descent*, in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, eds., Curran Associates, Red Hook, NY, 2021, pp. 4641–4650.
- [37] R. JIANG AND Y. GUAN, *Data-driven chance constrained stochastic program*, Math. Program., 158 (2016), pp. 291–327.
- [38] P. JUAN, Z. TIJANA, M.-D. CELESTINE, AND H. MORITZ, *Performatice prediction*, in *Proceedings of the 37th International Conference on Machine Learning*, H. D. III and A. Singh, eds., IEEE, Piscataway, NJ, 2020, pp. 7599–7609.
- [39] H. KUSHNER AND G. YIN, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer, New York, 2003.
- [40] N. H. LAPPAS AND C. E. GOUNARIS, *Robust optimization for decision-making under endogenous uncertainty*, Comput. Chem. Eng., 111 (2018), pp. 252–266.
- [41] S. LEE, T. HOMEM-DE MELLO, AND A. J. KLEYWEGT, *NewsVendor-type models with decision-dependent uncertainty*, Math. Methods Oper. Res., 76 (2012), pp. 189–221.
- [42] E. L. LEHMANN AND G. CASELLA, *Theory of Point Estimation*, Springer Science & Business Media, New York, 1998.
- [43] Y. LI, T. LIU, E. ZHOU, AND F. ZHANG, *Bayesian learning model predictive control for process-aware source seeking*, IEEE Control Syst. Lett., 6 (2022), pp. 692–697.
- [44] T. LIU, Y. LIN, AND E. ZHOU, *A Bayesian approach to online simulation optimization with streaming input data*, in *Proceedings of the 2021 Winter Simulation Conference*, S. Kim, B. Feng, K. S. S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds., IEEE, Piscataway, NJ, 2021.
- [45] F. LUO AND S. MEHROTRA, *Distributionally robust optimization with decision dependent ambiguity sets*, Optim. Lett., 14 (2020), pp. 2565–2594.
- [46] C. MENDLER-DÜNNER, J. PERDOMO, T. ZRNIC, AND M. HARDT, *Stochastic optimization for performatice prediction*, in *Advances in Neural Information Processing Systems*, Vol. 33, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds., Curran Associates, Red Hook, NY, 2020, pp. 4929–4939.
- [47] J. P. MILLER, J. C. PERDOMO, AND T. ZRNIC, *Outside the echo chamber: Optimizing the performatice risk*, in *Proceedings of the 38th International Conference on Machine Learning*, Vol. 139, M. Meila and T. Zhang, eds., Curran Associates, Red Hook, NY, 2021, pp. 7710–7720.
- [48] O. NOHADANI AND K. SHARMA, *Optimization under decision-dependent uncertainty*, SIAM J. Optim., 28 (2018), pp. 1773–1795.
- [49] N. NOYAN, G. RUDOLF, AND M. LEJEUNE, *Distributionally robust optimization with decision-dependent ambiguity set*, Optimization Online, <https://optimization-online.org/2018/09/6821/> (2018).
- [50] G. K. PEDERSEN, *Analysis Now*, Grad. Texts in Math. 118, Springer, New York, 1989.
- [51] S. T. RACHEV AND W. RÖMISCH, *Quantitative stability in stochastic programming: The method of probability metrics*, Math. Oper. Res., 27 (2002), pp. 792–818.
- [52] M. RAGINSKY, A. RAKHILIN, AND M. TELGARSKY, *Non-convex learning via stochastic gradient Langevin dynamics: A nonasymptotic analysis*, Proc. Mach. Learn. Res. (PMLR) 65, (2017), pp. 1674–1703.
- [53] S. SHALEV-SHWARTZ, ET AL., *Online learning and online convex optimization*, Found. Trends Mach. Learn., 4 (2011), pp. 107–194.

- [54] O. SHAMIR AND T. ZHANG, *Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes*, in Proceedings of the 30th International Conference on Machine Learning, Vol. 28, S. Dasgupta and D. McAllester, eds., Curran Associates, Red Hook, NY, 2013, pp. 71–79.
- [55] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYNSKI, *Lectures on Stochastic Programming: Modeling and Theory*, 3rd ed., SIAM, Philadelphia, 2021.
- [56] A. SHAPIRO, E. ZHOU, AND Y. LIN, *Bayesian distributionally robust optimization*, SIAM J. Optim., 33 (2023), pp. 1279–1304.
- [57] E. SONG AND U. V. SHANBHAG, *Stochastic approximation for simulation optimization under input uncertainty with streaming data*, in Proceedings of the 2019 Winter Simulation Conference, N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds., IEEE, Piscataway, NJ, 2019, pp. 3597–3608.
- [58] B. P. VAN PARYS, D. KUHN, P. J. GOULART, AND M. MORARI, *Distributionally robust control of constrained stochastic systems*, IEEE Trans. Automat. Control, 61 (2015), pp. 430–442.
- [59] H. WANG, X. ZHANG, AND S. H. NG, *A Nonparametric Bayesian Approach for Simulation Optimization with Input Uncertainty*, preprint, arXiv:2008.02154, 2020.
- [60] M. WEBSTER, N. SANTEN, AND P. PARPAS, *An approximate dynamic programming framework for modeling global climate policy under decision-dependent uncertainty*, Comput. Manag. Sci., 9 (2012), pp. 339–362.
- [61] W. WIESEMANN, D. KUHN, AND M. SIM, *Distributionally robust convex optimization*, Oper. Res., 62 (2014), pp. 1358–1376.
- [62] D. WU, Y. WANG, AND E. ZHOU, *Data-driven ranking and selection under input uncertainty*, Oper. Res., to appear, <https://doi.org/10.1287/opre.2022.2375>.
- [63] D. WU AND E. ZHOU, *Ranking and selection under input uncertainty: A budget allocation formulation*, in Proceedings of the 2017 Winter Simulation Conference, W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds., IEEE, Piscataway, NJ, 2017, pp. 2245–2256.
- [64] D. WU, H. ZHU, AND E. ZHOU, *A Bayesian risk approach to data-driven stochastic optimization: Formulations and asymptotics*, SIAM J. Optim., 28 (2018), pp. 1588–1612.
- [65] H. XIAO, F. GAO, AND L. H. LEE, *Optimal computing budget allocation for complete ranking with input uncertainty*, IIE Transactions, 52 (2020), pp. 489–499.
- [66] H. XIAO AND S. GAO, *Simulation budget allocation for selecting the top- m designs with input uncertainty*, IEEE Trans. Automat. Control, 63 (2018), pp. 3127–3134.
- [67] H. XU AND S. MANNER, *Distributionally robust Markov decision processes*, in Advances in Neural Information Processing Systems, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, eds., Curran Associates, Red Hook, NY, 2010, pp. 2505–2513.
- [68] I. YANG, *A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance*, IEEE Control Syst. Lett., 1 (2017), pp. 164–169.
- [69] I. YANG, *Wasserstein distributionally robust stochastic control: A data-driven approach*, IEEE Trans. Automat. Control, 66 (2020), pp. 3863–3870.
- [70] X. YU AND S. SHEN, *Multistage distributionally robust mixed-integer programming with decision-dependent moment-based ambiguity sets*, Math. Program., 196 (2022), pp. 1025–1064.
- [71] E. ZHOU AND T. LIU, *Online quantification of input uncertainty for parametric models*, in Proceedings of the 2018 Winter Simulation Conference, M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds., IEEE, Piscataway, NJ, 2018, pp. 1587–1598.
- [72] E. ZHOU AND W. XIE, *Simulation optimization when facing input uncertainty*, in Proceedings of the 2015 Winter Simulation Conference, L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds., IEEE, Piscataway, NJ, 2015, pp. 3714–3724.
- [73] M. ZINKEVICH, *Online convex programming and generalized infinitesimal gradient ascent*, in Proceedings of the 20th International Conference on Machine Learning, AAAI Press, Menlo Park, CA, 2003, pp. 928–936.