

How Big Should Your Data Really Be?

Data-Driven Newsvendor: Learning One Sample at a Time

Omar Besbes, Omar Mouchtaki

Columbia University, Graduate School of Business

obesbes@columbia.edu, om2316@gsb.columbia.edu

first version: March 15, 2021; last revised: July 22, 2022

Abstract

We study the classical newsvendor problem in which the decision-maker must trade-off underage and overage costs. In contrast to the typical setting, we assume that the decision-maker does not know the underlying distribution driving uncertainty but has only access to historical data. In turn, the key questions are *how to map existing data to a decision* and what type of performance to expect *as a function of the data size*. We analyze the classical setting with access to past samples drawn from the distribution (e.g., past demand), focusing not only on asymptotic performance but also on what we call the *transient regime of learning*, i.e., performance for arbitrary data sizes. We evaluate the performance of any algorithm through its worst-case relative expected regret, compared to an oracle with knowledge of the distribution. We provide the first finite sample *exact* analysis of the classical Sample Average Approximation (SAA) algorithm for this class of problems across *all* data sizes. This allows to uncover novel fundamental insights on the value of data: it reveals that *tens of samples* are sufficient to perform very efficiently but also that more data can lead to worse out-of-sample performance for SAA. We then focus on the general class of mappings from data to decisions *without any* restriction on the set of policies and derive an optimal algorithm (in the minimax sense) as well as characterize its associated performance. This leads to significant improvements for limited data sizes, and allows to exactly quantify the value of historical information.

Keywords: Limited data, data-driven decisions, minimax regret, sample average approximation, empirical optimization, finite samples, distributionally robust optimization.

1 Introduction

The newsvendor problem is a prototypical model of decision-making under uncertainty that captures the trade-offs emerging in capacity or inventory decisions in the face of uncertainty in future

outcomes. For example, when setting inventory decisions, a decision-maker typically faces uncertainty with regard to the demand that will materialize. For a given decision, if the demand turns out to be lower, the decision-maker would incur overage costs and if the demand realization is higher than the inventory, then some underage costs would be incurred for unsatisfied demand. Such trade-offs for inventory decisions represent some of the most common operational problems faced by retailers. Newsvendor type trade-offs also emerge in a variety of other applications, e.g., in revenue management for capacity setting or overbooking, or in electricity markets for setting capacity levels.

The key to solving the trade-offs above and optimizing decisions is a statistical characterization of the uncertainty the decision-maker faces, typically captured by a distribution. In the inventory example above, this would correspond to the distribution of demand in the period between replenishments. In practice, the distribution is typically unknown and the only way to solve the trade-offs above, is through the data that has been collected. The main questions this paper focuses on are the following: How should a newsvendor decision-maker optimize decisions given the historical data they have collected? What is the optimal performance they can garner as a function of the data size? We are interested in understanding the full spectrum of performance of policies across data sizes and refer to this approach as the transient regime lens for learning.

In more detail, we are interested in analyzing central policies in the literature, in optimizing data-driven policies, and in understanding whether one could characterize the performance achievable *across data sizes*, small and large. The motivation to develop a framework for understanding performance for arbitrary data sizes has strong anchoring in both practice and theory. Despite the apparent wide availability of demand data, we posit that “relevant” data may be limited in practice due to the heterogeneity of market characteristics. For example, a year of weekly demand for a product only represents tens of samples, and assuming homogeneity of demand on a longer period of time may be too strong of a practical assumption. On the theory front, such a framework would provide a foundation for a bottom-up approach to data-driven decisions and would reveal the true robust value of data.

To analyze these questions, we focus on the typical data structure that comes in the form of past samples from the unknown distribution. This would correspond, for example, to demand observations in the inventory example. A data-driven policy is then a mapping from historical data to decisions. For any such policy, we evaluate its performance according to the worst-case (over all possible distributions) expected relative regret defined as the difference between the expected out-of-sample cost incurred by the data-driven policy and the expected optimal cost of an oracle

that knows the distribution, normalized by the latter cost.

It is important to highlight that evaluating the worst-case performance of a given data-driven algorithm against an arbitrary unknown distribution amounts to an intricate infinite dimensional optimization problem over the space of distributions. The characterization of an optimal algorithm and its performance, or even the exact performance of specific data-driven algorithms, have been mostly elusive to date.

1.1 Main contributions

Sample Average Approximation Analysis. A popular and central approach to such data-driven problems is the Sample Average Approximation (SAA) algorithm (also typically referred to as Empirical Optimization) which minimizes the expected cost according to the empirical distribution induced by the observed samples. This method has been introduced to solve various stochastic optimization problems and enjoys asymptotic guarantees (Kleywegt et al. (2002)). In the context of newsvendor decisions, state-of-the-art *instance-independent* results on the number of samples required for SAA to achieve a particular confidence level were derived in Levi et al. (2015) and in Cheung and Simchi-Levi (2019). While they capture the correct dependence on the confidence level as the number of samples grows large, state-of-the-art lower and upper bounds on the number of samples required to achieve a particular confidence level differ by orders of magnitude, leading to significant uncertainty on the value of information, or on the quality of SAA. As such, despite its wide use and central role in the literature and in practice, to date, a significant gap exists in the understanding of the *actual* performance of this policy and the value it can capture from data.

Our first main contribution is the characterization of the *exact* performance of SAA for newsvendor problems for arbitrary data sizes (Theorem 2). While determining the performance of the SAA algorithm is a priori an infinite dimensional optimization problem over the space of possible distributions, we actually establish that it is possible to reduce it to a one dimensional optimization problem and in turn derive a quasi closed-form solution that gives the exact worst-case relative regret of SAA. This worst-case is derived *for any number of samples* and can be computed efficiently using a line-search. Our method relies on the structure of the newsvendor problem and develops an analysis that leads to the SAA performance as a corollary. We establish that for any policy that can be expressed as an order statistic or a randomization of order statistics of the empirical distribution, one can transform the initial problem into a pointwise optimization problem of an appropriate functional, which ultimately leads to identifying the family of worst-case distributions for such policies (Theorem 1). In particular, we establish, quite interestingly, that across all distributions, the worst case is a Bernoulli distribution whose mean depends on the number of samples

observed. In turn, this yields the worst-case performance of the SAA policy as the solution of a one-dimensional search. While our analysis reveals that the worst-case distribution is a Bernoulli, the induced worst-case performance we obtain can be applied to bound the relative regret of SAA against *any* distribution.

Quite notably, this enables, for the first time, to fully characterize the spectrum of performances achievable by SAA *across data sizes*. These results highlight many fundamentally novel insights on the value of information. As examples, with 20 samples, SAA already leads to a relative regret of 26.8%, and with 100 samples, the relative regret shrinks to 8.1%. This highlights the possibility of making effective decisions already with very limited data. As a matter of fact, in Table 1 below, we show that the number of samples required by SAA to achieve a certain level of accuracy, as derived from the analysis in this paper, is actually two orders of magnitudes lower than the number induced by state-of-the art upper bounds on the expected relative regret in the existing literature (see Section 3.2.1 and Section 4.4).

		Expected relative regret target				
		25%	20%	15%	10%	5%
SAA	This paper	21	23	42	71	210
	Best known to date	5,088	7,780	13,530	29,762	100,000+
Optimal Algorithm	This paper	14	19	25	50	161

Table 1: **Number of samples that ensures a target relative regret.** The table reports the induced number of samples needed to reach a relative regret accuracy level. For SAA, we compare the best instance-independent known bounds to date (Levi et al., 2015) and the exact worst-case analysis developed in the present paper. We also report the number of samples needed by an optimal data-driven algorithm, derived in the present paper. Example with service level of 0.9.

Our analysis of the SAA policy also leads to another striking new insight: the relative regret is not monotone in the number of samples available. This implies that SAA is suboptimal but also that SAA is not able to accumulate information appropriately, sometimes “destroying” information. We highlight that the possibility to uncover this non-monotonic behavior has been enabled by the transient regime lens for learning we take.

Optimal data-driven policy. In turn, the next question we tackle pertains to optimal worst-case performance in the space of data-driven policies. Indeed, it is important to note that SAA is only one possible prescription among all possible mappings from data to decisions. Our next contribution lies in characterizing the minimal worst-case expected relative regret in the general space of data-driven policies and across all data sizes. In the remaining of the paper we refer to the algorithm achieving the minimax optimality for the worst-case expected relative regret as the optimal algorithm.

To prove this fundamental optimality result, we first establish a central reduction in the space of mechanisms. We show that, *without loss of optimality*, one may restrict attention to mixtures of order statistics (Theorem 3). In turn, we derive necessary conditions for optimality in that subspace. This leads to a candidate policy. The last step consists of establishing optimality of this candidate. For that, we introduce an alternative minimax problem in which we relax the space of distributions to be distributions over distributions, and show that the candidate, together with a proper mixture of distributions, is a saddle point for the alternate problem. This yields an optimal data-driven algorithm and its associated performance for the original problem (Theorem 4).

We establish that an optimal data-driven policy takes actually a simple form: it is a randomization between two consecutive order statistics, and we provide a procedure to compute its tuning parameters as well as its performance. As a corollary, we obtain that an alternative policy, which prescribes a convex combination of consecutive order statistics, is also minimax optimal while also (weakly) improving over the optimal mixture of order statistics policy for all possible demand distributions.

This result has significant implications. First, we can now assess the potential losses stemming from using the suboptimal SAA policy. We show that these can be significant for small data sizes and become smaller as the data size increases. Second, it allows to exactly *assess the value of the historical information* and to understand how effective one can be as a function of the data at hand without any assumption on the underlying distribution. This further emphasizes the possibility of operating effectively with limited data. In Table 1, we report the number of samples required to reach a particular level of accuracy, and one sees that, compared to SAA, the optimal algorithm reduces the amount of data needed to reach a particular level significantly (by 17% to 40% across the targets illustrated). We also note that, as a byproduct of our analysis, we also obtain the worst-case performance ratio for any Bayesian problem.

We highlight here that there has been significant work on data-driven policies. We discuss more these in the literature review and refer to Lam (2021) for a very recent overview of various subfamilies of policies considered in the literature. We also emphasize that when searching for optimal policies, we consider all possible mappings from data to decisions, and do not restrict attention to a subfamily of policies.

We note that, while our analysis is tailored around the worst-case relative regret, the minimax optimal policy that we derive is not overly conservative on “mild” instances. As a matter of fact, we show numerically in Section 6 that the performance of the optimal policy is typically on par or better than the one of SAA on a broad range of distributions. As a consequence, the “robustification” of

SAA in the worst-case comes at no cost and even typically translates into better performance on many common distributions.

Optimal asymptotic performance. When the data size becomes large, there are various existing results in the literature, and a corollary of these leads to upper bounds on the rate of convergence to zero of the expected relative regret of SAA as the data size n grows to infinity: it converges to zero at rate $O(1/\sqrt{n})$. While this makes progress on capturing the dependence in the data size n , even asymptotically, there is still limited understanding of the actual performance. In particular, there is no understanding of the *constant* characterizing the rate of convergence for SAA, nor for the best such constant achievable by a data-driven algorithm.

We leverage the exact finite sample analysis to derive, from the bottom up, the exact rate of convergence to zero, fully characterizing the constant for SAA and optimal performance. In particular, we show that the optimal relative regret asymptotically scales like C^*/\sqrt{n} where the number of samples n is large and provide a closed form expression for C^* (Theorem 5). This highlights how the rate of convergence is affected by the various economic parameters associated with the newsvendor decision. In addition, we establish that SAA asymptotically achieves rate optimality with the *same* limiting constant. As such, while SAA could lead to high suboptimality gaps for small data sizes, it satisfies a very strong notion of near-optimality for large data sizes.

Stepping back, one may see the present study as building a foundation for a “bottom-up” approach to data-driven decision-making in newsvendor settings. We highlight that our transient regime lens for learning and the associated exact analysis account for all the expected out-of-sample cost implications of deviations, small or large, that SAA (or an optimal policy) could generate compared to the oracle. As such, it allows to build an understanding of data-driven policies “one data point at a time.” Compared to existing approaches that are mostly anchored around large data regimes, this new perspective establishes that it is indeed possible to characterize performance across data sizes. It leads to new insights for small as well as large data regimes. We hope that the techniques developed here lead to further the understanding of the transient regime of learning in richer settings relating to newsvendors, but also across problem classes.

1.2 Related literature

Capacity management problems in the face of uncertainty are central across literatures and the present paper builds on and contributes to a vast existing literature.

Our work first relates to the study of this class of problems with limited information on the underlying distribution of demand. In early work, [Scarf \(1958\)](#) and [Gallego and Moon \(1993\)](#) characterizes the min-max optimal solution for the inventory problem when the mean and the

variance of the demand are known. [Perakis and Roels \(2008\)](#) derive robust policies that achieve minimax regret under various types of partial information on the demand function such as moments of the distribution, modes or symmetry. [Natarajan et al. \(2018\)](#) carries this robustness analysis for asymmetric distributions.

Information about the distribution may also be improved by a data-driven approach. [Xu et al. \(2021\)](#) construct ambiguity sets by using non-parametric information on the distribution along with observed samples. [Saghafian and Tomlin \(2016\)](#) develop a Maximum Entropy approach to leverage information from data combined with moment and tail bounds. [Liyanage and Shanthikumar \(2005\)](#); [Chu et al. \(2008\)](#) introduce the operational statistic framework which integrates estimation and optimization tasks for newsvendor problems under parametric classes of distributions. [Chu et al. \(2008\)](#), assuming that the decision-maker knows the distribution of demand up to a scale parameter, derives a mapping from data to decision that maximizes expected profit for any value of the unknown scale parameter. In the present work, we do not make any assumption on the underlying distribution of demand.

When no such information is initially available, the question becomes how to go from data to decisions. There are various dimensions associated with this problem, first on the level of uncertainty about the underlying distribution, and second on the offline or online aspect of the decision-making problem.

The present paper focuses on a non-parametric setting in which little, if anything is known about the underlying distribution and only data in the form of samples can be used to make decisions. A first foundational question for this class of problems is one pertaining to sample complexity: How many samples are needed to achieve a certain level of accuracy? Closest to our work are [Levi et al. \(2007\)](#), [Levi et al. \(2015\)](#) and [Cheung and Simchi-Levi \(2019\)](#) which establish bounds on probabilistic guarantees of the relative regret of Sample Average Approximation (SAA). In particular [Levi et al. \(2015\)](#) presents bounds that are problem-independent and apply to any distribution, and we compare to those in Section 3.2.1. [Levi et al. \(2015\)](#) also improve these bounds by deriving instance dependent guarantees in cases where the decision-maker has additional information about the class of distributions to which the demand belongs. In contrast, our work improves the characterization of the worst-case expected relative regret of SAA without any supplementary information about the distribution. [Cheung and Simchi-Levi \(2019\)](#) provides a lower bound on the number of samples required to achieve a target relative regret with a probability exceeding a given threshold. Their result implies that the upper bound derived in [Levi et al. \(2015\)](#) has the correct dependence in the problem parameters. [Ban \(2020\)](#) establishes consistent estimators for (s, S) policies for both

censored and uncensored information regimes. They derive bounds on the regret by constructing asymptotic confidence intervals around the policy.

In the offline setting, other related papers study the contextual version of the problem in which the decision-maker observes previous samples of demand along with features that give additional information on the environment (Ban and Rudin, 2019; Qi et al., 2021). Ban and Rudin (2019) proposes approaches based on Empirical Risk Minimization and kernel methods to derive generalization bounds for the cost of a feature-based data-driven decision. In the special case without contexts, their approach recovers the instance-independent bound derived by Levi et al. (2015). In contextual optimization, we also refer the reader to Elmachtoub and Grigas (2021) for a general data-driven approach that explicitly accounts for the nature of the optimization problem at hand.

Our paper also relates to the rich literature on the analysis of SAA. This approach has been applied broadly for discrete stochastic optimization problems when the underlying distribution is either unknown or when the expected objective function is hard to optimize, Kleywegt et al. (2002), or for multi-stage stochastic optimization problem (Swamy and Shmoys (2005); Shapiro (2008)). It has also been used in the specific context of multi-stage inventory planning (Levi et al. (2007); Cheung and Simchi-Levi (2019)) or for the newsvendor model (Levi et al. (2015); Besbes and Muharremoglu (2013)).

Gupta and Kallus (2022) share the motivation of limited data sizes, and explore the possibilities associated with pooling data across products. Bertsimas et al. (2018), Esfahani and Kuhn (2018) develop robust approaches enjoying probabilistic guarantee over the out-of-sample error. Bertsimas et al. (2018) considers ambiguity sets containing all distributions that pass a statistical goodness-of-fit test for given historical data. Esfahani and Kuhn (2018) proposes a data-driven distributionally robust approach by constructing an uncertainty ball around the empirical distribution. They show that the worst-case expectation over a Wasserstein ambiguity set can in fact be computed efficiently via convex optimization techniques for various loss functions. In contrast to these strategies, we highlight that we do not restrict the space of policies to those that construct uncertainty sets, but explore the entire space of mappings from data to decisions.

More broadly, our work relates to sequential decision-making under uncertainty. In the class of inventory decisions, a rich line of work on dynamic decisions has been developed. Different information structures (observable or censored demand) are studied and adaptive algorithms with desirable asymptotic properties derived. Godfrey and Powell (2001); Huh and Rusmevichientong (2009); van Ryzin and McGill (2000) develop gradient based methods to solve this sequential problem whereas Huh et al. (2011) uses the Kaplan-Meier estimator and Maglaras and Eren (2015) studies

a maximum entropy approach to dynamically adjust capacity levels. [Besbes and Muharremoglu \(2013\)](#) studies the price of demand censoring in these sequential decision making problem with stationary demand, and [Lugosi et al. \(2021\)](#) study censoring in a setting when demand is non-stationary. [Chen et al. \(2021\)](#) studies the interplay of pricing and inventory decisions. We note that in all these studies, the performance of policies are characterized asymptotically up to multiplicative constants, but there is no characterization of exact optimal performance. We hope that the exact performance characterization, and optimality results, developed in this work in the offline case with demand observations, will lead to future progress in this related class of problems.

Our work is also connected to the rich literature in statistics which focuses on the performance of various quantile estimators, and our problem may be reframed as the one of deriving minimax quantile estimators for a particular metric (here the relative regret between newsvendor losses). Depending on the application and the desired properties, many quantile estimators have been derived, either as L-estimators based on order statistics ([Harrell and Davis, 1982](#); [Kalgh and Lachenbruch, 1982](#); [Yang, 1985](#)) or by using different methods such as Stochastic Approximation ([Tierney, 1983](#)). This profusion of heuristics motivated the study of estimators achieving certain forms of optimality. In parametric settings, [Rukhin and Strawderman \(1982\)](#), and [Rukhin \(1983\)](#) derive minimax equivariant quantile estimators for a normalized squared loss. In non-parametric settings, [Zieliński \(1999\)](#) restricts attention to the set of equivariant estimators and derives an estimator uniformly better with respect to a particular metric, the worst-case F-Mean Absolute Deviation. Our work differs from [Zieliński \(1999\)](#) along various crucial dimensions. We focus on the commonly studied objective of minimax relative regret, on a newsvendor cost, and we do not restrict the space of decisions.

Our work is also remotely related to the understanding of the learning curve defined as the expected generalization performance, i.e., the out-of-sample performance, of a learner as a function of the size of the training set. We refer the reader to the recent review of [Viering and Loog \(2021\)](#). Our approach complements this line of work and gives a theoretical understanding of the robust (worst-case) value of data sizes for the newsvendor problem.

Finally, we also note that another related line of work pertains to modeling uncertainty differently. Another framework that has been widely studied is parametric and Bayesian, in which the decision-maker is assumed to have access to a prior about an underlying unknown parameter that characterizes the distribution. The seminal work of [Scarf \(1959\)](#) analyses a bayesian setting in which the decision maker has a prior belief on the nature of the distribution and updates his belief as he observes samples from the demand. The goal is then to analyze methods that use historical

data to prescribe inventory decisions on the fly. This line of work has also a rich literature dealing with different information structures (censored versus uncensored) observations. See, e.g., [Azoury \(1985\)](#), [Lariviere and Porteus \(1999\)](#), [Ding et al. \(2002\)](#) (and the related notes by [Lu et al. \(2005\)](#) and [Bensoussan et al. \(2009\)](#)) and [Besbes et al. \(2022\)](#).

2 Problem Formulation

We consider a newsvendor problem in which the decision maker decides on a capacity/inventory decision x in the face of uncertainty on the underlying demand D that will materialize. Any excess inventory leads to overage costs $h > 0$ per unit, and any demand that is not satisfied leads to underage cost of $b > 0$ per unit. In turn, the cost associated with decision x is given by

$$c(x, D) := b(D - x)^+ + h(x - D)^+.$$

Decision-making with knowledge of the distribution of D . Suppose that demand D is drawn from a distribution F supported on \mathbb{R}_+ . Then, in the classical newsvendor problem, the decision-maker minimizes the expected cost given by

$$c_F(x) := \mathbb{E}_{D \sim F} [b(D - x)^+ + h(x - D)^+]. \quad (1)$$

Let \mathcal{G} denote the set of distributions (cdf) with non-negative support. By convention, we assume that all cdf's are cadlag. Furthermore, we let

$$\mathcal{F} = \{F \in \mathcal{G} : \mathbb{E}_F[D] < \infty\}$$

denote the set of distributions with bounded first moment. We will assume that D is drawn from a distribution in \mathcal{F} , so that the above cost always admits a well defined expectation.

When the distribution F is known, the inventory decision minimizing the cost $c_F(\cdot)$ is given by

$$x_F^* := \min\{x \geq 0 : F(x) \geq q\},$$

where

$$q = \frac{b}{b + h}.$$

We will refer to q as the critical quantile, and we will let

$$\text{opt}(F) := c_F(x_F^*)$$

denote the minimal achievable cost with knowledge of the distribution F .

Data-driven decision-making. In the present paper, we consider a setting in which the distribution F is unknown to the decision-maker and only data is available in the form of past demand observations. The decision-maker observes n historical samples of demand, $\mathbf{D}_1^n := (D_1, \dots, D_n)$, where D_i are independently sampled from F . In this context, an admissible policy π is a mapping from observed demand \mathbf{D}_1^n to inventory decision x^π . Formally, we consider the class of policies Π_n of mappings from \mathbb{R}_+^n into the set of distributions \mathcal{F} . In particular, a policy π is a mapping

$$\pi : \mathbf{D}_1^n \mapsto G_{\mathbf{D}_1^n},$$

where $G_{\mathbf{D}_1^n} \in \mathcal{F}$. That is to say π maps previous demand observations to a randomized inventory decision. We note that, even when the policy π is a deterministic function of the observed demand \mathbf{D}_1^n , the inventory decision x^π is a random variable as it depends on \mathbf{D}_1^n .

When using a policy π , given n samples from an underlying distribution F , the out-of-sample expected cost incurred is defined as,

$$\mathcal{C}(\pi, F, n) := \mathbb{E}_{\mathbf{D}_1^n \sim F} \left[\mathbb{E}_{x \sim \pi(\mathbf{D}_1^n)} [c_F(x)] \right].$$

Note that the dependence between the underlying demand distribution and the expected cost of a data-driven algorithm is an intricate one in general, as *the demand distribution F affects the history the decision-maker observes, but also the out-of-sample performance of data-driven decisions.*

Objective. We are interested in understanding and quantifying the performance of data-driven algorithms for newsvendor problems. To that end, we evaluate the performance of a policy $\pi \in \Pi_n$ through the relative regret defined for every $F \in \mathcal{F}$ as¹,

$$\mathcal{R}_n(\pi, F) := \frac{\mathcal{C}(\pi, F, n) - \text{opt}(F)}{\text{opt}(F)}.$$

Note that the ratio above is always greater or equal than 0 and takes value in $[0, \infty) \cup \{\infty\}$. Given that the decision-maker does not know the distribution, we evaluate its performance through the

¹Note that $\text{opt}(F) = 0$ if and only if the distribution F has all its mass at a single point. In such a case, we set, by convention, $\mathcal{R}_n(\pi, F) = 0$, for any policy $\pi \in \Pi_n$ such that $\mathcal{C}(\pi, F, n) = \text{opt}(F) = 0$.

worst-case relative regret defined as follows

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F). \quad (2)$$

It represents the relative loss stemming from the gap between observing data of size n and full information on the demand distribution. We will be interested in characterizing the performance of specific policies considered in the literature, but also in the optimal achievable performance

$$\mathcal{R}_n^* := \inf_{\pi \in \Pi_n} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F). \quad (3)$$

We note that the above problem involves two infinite dimensional optimization problems, that of the decision-maker when selecting a policy, and that of nature when selecting a worst-case distribution. We also remark that there are no restrictions on the class of policy used in (3).

Notation. For any μ in $[0, 1]$, we let $\mathcal{B}(\mu)$ denote the distribution of a Bernoulli with mean μ . For any set A , $\Delta(A)$ denotes the set of distributions on A . We further let Δ_n denote the simplex in n dimensions, i.e., $\Delta_n = \{\boldsymbol{\lambda} \in \mathbb{R}^n : \lambda_i \geq 0, i \in \{1, \dots, n\}, \sum_{i=1}^n \lambda_i = 1\}$. For any deterministic sequences $(a_k)_{k \in \mathbb{N}}$ and $(b_k)_{k \in \mathbb{N}}$ both indexed by a common index k that goes to ∞ , we say that $a_k = o(b_k)$ if $a_k/b_k \rightarrow 0$, $a_k = \mathcal{O}(b_k)$ if there exists a finite $M > 0$ such that $|a_k| \leq M|b_k|$ for k large enough, $a_k = \omega(b_k)$ if $|a_k/b_k| \rightarrow \infty$, $a_k = \Omega(b_k)$ if there exist a finite $M > 0$, such that $|a_k| > M|b_k|$ for k large enough, $a_k = \Theta(b_k)$ if $a_k = \mathcal{O}(b_k)$ and $a_k = \Omega(b_k)$, and $a_k \sim b_k$ if $a_k/b_k \rightarrow 1$.

All proofs are deferred to the Online Appendix.

3 Sample Average Approximation: Performance Analysis across Data Sizes

The data-driven newsvendor problem is a particular instance of a data-driven stochastic optimization problem. One of the most common approaches to solve this type of problems is the Sample Average Approximation (SAA). As highlighted earlier, this approach has been applied to a broad set of problems. Previous works derived convergence guarantees for this method (Kleywegt et al. (2002)) but also bounds on probabilistic finite sample performance (Levi et al. (2007, 2015); Cheung and Simchi-Levi (2019)). In the newsvendor setting, SAA is also related to the broader literature of Distributionally Robust Optimization as it happens to be equivalent to the distributionally robust policy over the Wasserstein ball as noted in (Esfahani and Kuhn, 2018, Remark 6.7).

In the context of the newsvendor problem, SAA consists in solving the optimization problem

$$\min_x \frac{1}{n} \left[\sum_{i=1}^n b(D_i - x)^+ + h(x - D_i)^+ \right]. \quad (4)$$

This approach approximates the expectation in (1) with the empirical expectation, and solves the resulting problem. In particular, the solution of problem (4) is the q^{th} -empirical quantile. More precisely, let us define the order statistics of the historical dataset of demands observed as

$$D_{1:n} \leq \dots \leq D_{n:n}.$$

The SAA policy, which we will denote by π^{SAA} prescribes the $\lceil qn \rceil^{th}$ order statistic. With some abuse of notation, we have²

$$\pi^{\text{SAA}}(\mathbf{D}_1^n) = D_{\lceil qn \rceil:n}. \quad (5)$$

As highlighted in the introduction, this policy has been extensively studied. In particular, as the number of demand samples n grows, it is known that SAA leads to a solution that ensures that its worst-case relative regret approaches 0 as n grows to ∞ . Previous approaches derived upper bounds on the rate at which such convergence takes place, leveraging large deviations bounds. However, despite its widespread use, there is no characterization of its *actual* performance for a finite number of samples, and as a result, there is no robust quantification of the amount of data needed to achieve a particular level of performance.

Analyzing exactly the worst-case performance of SAA, or any policy, as in Problem (2), is an infinite dimensional optimization problem over a non-parametric class of distributions. For any policy $\pi \in \Pi_n$, let $G_{\mathbf{D}_1^n}^\pi := \pi(\mathbf{D}_1^n)$ be the distribution induced by π on the inventory level conditional on observing historical demand \mathbf{D}_1^n . The cost incurred by a policy $\pi \in \Pi_n$ against a distribution $F \in \mathcal{G}$ can be expressed as follows

$$\mathcal{C}(\pi, F, n) = \int_{\mathbf{D}_1^n \in [0, \infty)^n} \int_0^\infty \int_0^\infty c(x, D) dF(D) dG_{\mathbf{D}_1^n}^\pi(x) dF(D_1) \dots dF(D_n) \quad (6)$$

Therefore, the cost of a data-driven policy has in general a complex dependence on the demand distribution which appears in the integration measure.

In what follows, for any policy $\pi \in \Pi_n$, when solving for the worst-case performance, we will be

²Technically speaking, this is the policy that prescribes a point mass at $D_{\lceil qn \rceil:n}$.

working with the epigraph formulation of problem (2). Note that $\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) \geq 0$ and

$$\begin{aligned} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) &= \inf_{z \in \mathbb{R}_+} z \\ \text{s.t.} \quad &\mathcal{R}_n(\pi, F) \leq z \quad \forall F \in \mathcal{F}. \end{aligned}$$

It is easy to see that the problem can be further written as (this claim is formally established in Lemma E-1)

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) = \inf_{z \in \mathbb{R}_+} z \tag{8a}$$

$$\text{s.t.} \quad \mathcal{C}(\pi, F, n) - (z + 1)\text{opt}(F) \leq 0 \quad \forall F \in \mathcal{F}. \tag{8b}$$

This problem thus involves infinitely many constraints, and each of the constraints has a complex dependence in F as highlighted in (6). In Section 3.1, we analyze a general class of order statistic policies, of which SAA is a special case, and show that for these, the optimization problem (8) can be significantly simplified and as a matter of fact exactly solved. In particular, we leverage the structure of order statistic policies to simplify the expression of the cost function described in (6). This allows us to reduce the set of constraints (8b) to constraints parametrized by a one dimensional set.

3.1 Order Statistic Policies and Structural Results

As presented in (5), in the context of the newsvendor problem, SAA prescribes an inventory level according to an order statistic of the samples observed, the $\lceil qn \rceil^{\text{th}}$ order statistic. This can be seen as a special case of prescribing an order statistic or even a randomization over order statistics. To that end, we next define general order statistics policies which will also play a central role when we discuss optimal performance in Section 4.

Definition 1 (Mixture of Order Statistics). *Fix $n \geq 1$. For every $i \in \{1, \dots, n\}$, we let π^{OS_i} denote the policy that uses the i^{th} order statistic $D_{i:n}$ with probability one. Formally,*

$$\pi^{OS_i} : \mathbf{D}_1^n \mapsto \mathbb{1} \{x \geq D_{i:n}\}.$$

For any $\lambda \in \Delta_n$. We let π^λ denote the mixture of order statistics policy defined as follows: π^λ

prescribes the i^{th} order statistic $D_{i:n}$ with probability λ_i . Formally,

$$\pi^\lambda : \mathbf{D}_1^n \mapsto \sum_{i=1}^n \lambda_i \mathbb{1}\{x \geq D_{i:n}\}.$$

In the following, we denote by Π_n^{OS} the space of mixture of order statistics policies with n samples.

Another important class of policies are convex combinations of order statistics policies, which prescribe a deterministic convex combination instead of randomizing between different order statistics. We will relate the performance of policies in this class to the one for mixtures of order statistics policies in Section 4.3.

For a policy $\pi \in \Pi_n$, the expression of the relative regret involves the ratio between $\mathcal{C}(\pi, F, n)$ and $\text{opt}(F)$. In general, both quantities require to compute complex integral expressions in which the dependence on the demand distribution is intricate. Our first structural result establishes that for a mixture of order statistics policy, the cost incurred against any demand distribution F can be expressed as a single integral in which the integrand is a polynomial of the demand distribution and the integrating measure is the Lebesgue measure. We similarly show that the cost of the oracle is the integral of a piecewise-linear function of the demand distribution. Formally, we show the following.

Proposition 1. *For any $F \in \mathcal{F}$, any $n \geq 1$, and any mixture of order statistics policy π^λ we have,*

$$\begin{aligned} \mathcal{C}(\pi^\lambda, F, n) &= (b+h) \left[\int_0^\infty \sum_{i=1}^n \lambda_i ((1 - B_{i,n}(F(y)))(F(y) - q) + q(1 - F(y))) dy \right], \\ \text{opt}(F) &= (b+h) \int_0^\infty \min\{(1-q)F(y), q(1-F(y))\} dy, \end{aligned}$$

where $B_{i,n}$ is a Bernstein polynomial defined for any $y \in [0, 1]$ as

$$B_{i,n}(y) = \sum_{j=i}^n b_{j,n}(y),$$

with $b_{j,n}(y) = \binom{n}{j} y^j (1-y)^{n-j}$.

Recall the expression in (6), Proposition 1 shows that for mixture of order statistics policies, the cost function can be expressed as an integral in which the dependence in F only appears in the integrand but does not appear in the integrating measure anymore. As we will see, this is a key step towards the understanding of worst case distributions for this family of policies.

The main step in the proof of this result follows from Riemann–Stieltjes integration by part and from exploiting the special form of the cumulative distribution function of an order statistic.

Indeed, for any $F \in \mathcal{F}$, $n \geq 1$ and $r \in \{1, \dots, n\}$, the cumulative distribution of $D_{r:n}$ denoted by $F_{r:n}$ satisfies for $x \in \mathbb{R}_+$,

$$F_{r:n}(x) = B_{r,n}(F(x)).$$

The new expressions in Proposition 1 imply that the epigraph formulation (8) can be simplified. In particular, by rewriting the set of constraints (8b), we obtain the following formulation.

$$\inf_{z \in \mathbb{R}} \quad z \tag{9a}$$

$$\text{s.t.} \quad \sup_{F \in \mathcal{F}} \int_0^\infty \Psi_z^\lambda(F(y)) dy \leq 0, \tag{9b}$$

where Ψ_z^λ is a continuous mapping from $[0, 1]$ to \mathbb{R} . The constraint of (9) now involves a nonparametric optimization problem for which the demand distribution only appears in the integrand. This expression allows us to reduce the functional optimization problem over the class of distributions \mathcal{F} to a pointwise optimization problem. We now present formally this result as our first main contribution.

Recall that $\mathcal{B}(\mu)$ denotes a Bernoulli distribution with mean μ . Our first main result is a characterization of the exact performance of any mixture of order statistics policy.

Theorem 1. *Fix $n \geq 1$ and $\pi^\lambda \in \Pi_n^{OS}$. The worst-case performance of the policy π^λ satisfies*

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^\lambda, F) = \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)).$$

Furthermore for every $\mu \in [0, 1]$,

$$\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) = \sum_{i=1}^n \lambda_i \frac{(1 - B_{i,n}(1 - \mu))(1 - \mu - q) + q \cdot \mu}{\min\{(1 - q)(1 - \mu), q \cdot \mu\}} - 1.$$

This result has many implications. First, it establishes the notable fact that for any data size n , the worst-case performance of mixture of order statistics policies over the *entire* space of distributions \mathcal{F} is achieved at a Bernoulli distribution.

Second, in evaluating the worst-case performance of these policies, Theorem 1 leads to a reduction from a non-parametric general space of distributions \mathcal{F} to a space of distributions parametrized by a single parameter, the mean of the Bernoulli distribution. Moreover, in the case of Bernoulli distributions, the relative regret has a *closed-form* expression. Therefore, the *exact* worst-case performance of mixture of order statistics policies can be computed for *any* number of samples n through a *simple line search*. In Section 3.2, we analyze the implications of this result for SAA. This

result will also be instrumental when we analyze optimal policies in the entire space of mappings from data to decisions in Section 4.

Remark. At first glance, the result of Theorem 1 may seem counter-intuitive as one would expect that in the broad family of distributions \mathcal{F} , a “hard” instance for order statistics policies would have unbounded support. This result proves that on the contrary, the difficulty of the data-driven newsvendor problem does not stem from the tail of the distribution. Indeed, for distributions that are hard to learn, such as heavy-tail distributions, oracle costs are also large. Bernoulli distributions are flexible enough to lead to a low cost for the oracle due to the simple structure of the distribution, but also exacerbates the cost of mistakes for a decision-maker that does not know the distribution as the problem boils down to deciding between two extreme actions: prescribing 0 or 1.

Remark (Absolute vs. relative regret). We highlight here that a similar argument as the one developed to prove Theorem 1 can be used to establish a parallel result for the worst-case expected *absolute* regret metric, $\mathcal{C}(\pi, F, n) - \text{opt}(F)$. In such a case, if one restricts attention to the space of distributions supported on a bounded interval $[0, M]$ (for some real value M), then one can show that the worst-case expected absolute regret for any mixture of order statistics is achieved for a two point distribution with mass at 0 and M .

3.2 Performance Analysis of SAA across Data Sizes

A direct and important consequence of Theorem 1 is the ability to characterize the transient regime of learning, or exact performance associated with the central algorithm SAA for an *arbitrary number of samples*. Since SAA is a special case of mixture of order statistics policy, the following theorem is a direct corollary.

Theorem 2 (SAA Finite Sample Performance). *For any $n \geq 1$, the performance of the SAA policy is given by*

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{SAA}, F) = \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^{SAA}, \mathcal{B}(\mu)) = \sup_{\mu \in [0,1]} \frac{(1 - B_{[qn],n}(1 - \mu))(1 - \mu - q) + q \cdot \mu}{\min\{(1 - q)(1 - \mu), q \cdot \mu\}} - 1.$$

Theorem 2 leads to the notable result that one may characterize *exactly* the worst-case performance of the central SAA algorithm across *all data sizes* by performing a simple line search! As such, it allows to exactly measure the implications of all possible *out-of-sample* “mistakes” (compared to the oracle) made by SAA, and this is for any data size. Next, we analyze the implications

of this result on the value of data, compare this result to earlier bounds in the literature, as well as uncover novel insights on the quality of SAA as a data-driven policy in this class of problems.

3.2.1 Performance of SAA and Comparison to Existing Related Results

As mentioned earlier, SAA has been widely studied in various settings. In the context of the newsvendor problem, [Levi et al. \(2007\)](#) establish bounds relying on large deviations arguments to derive probabilistic results, which were later improved in [Levi et al. \(2015\)](#), with associated relative regret guarantees. More formally, ([Levi et al., 2015](#), Theorem 2) show that for any $\epsilon > 0$, any $n \geq 1$ and any demand distribution F , the relative-regret of SAA satisfies

$$\mathbb{P}_{\pi^{\text{SAA}}} \left(\frac{c_F(x) - \text{opt}(F)}{\text{opt}(F)} > \epsilon \right) \leq 2 \exp \left(-\frac{n\epsilon^2}{18 + 8\epsilon} \min(q, (1 - q)) \right),$$

with the associated bound on relative regret given by

$$\begin{aligned} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{\text{SAA}}, F) &= \sup_{F \in \mathcal{F}} \mathbb{E}_{\pi^{\text{SAA}}} \left[\frac{c_F(x) - \text{opt}(F)}{\text{opt}(F)} \right] \\ &= \sup_{F \in \mathcal{F}} \int_0^\infty \mathbb{P}_{\pi^{\text{SAA}}} \left(\frac{c_F(x) - \text{opt}(F)}{\text{opt}(F)} > \epsilon \right) d\epsilon \\ &\leq \int_0^\infty 2 \exp \left(-\frac{n\epsilon^2}{18 + 8\epsilon} \min(q, (1 - q)) \right) d\epsilon =: U(n). \end{aligned}$$

The function $U(n)$ represents state of the art instance-independent bounds for the worst-case performance of SAA as a function of the data size in the literature to date. Our result in [Theorem 2](#) allows to characterize $\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{\text{SAA}}, F)$, the actual worst-case performance of SAA. In [Table 2](#), we present a comparison of the number of samples required to guarantee various levels of relative regret (25%, 20%, ..., 5%) for different values of the critical fractile. We do so using the induced number based on [Theorem 2](#) in this paper, and based on $U(n)$. Formally, for a given performance threshold $\tau \geq 0$, we define

$$\begin{aligned} N^{\text{exact-SAA}}(\tau) &:= \min \left\{ m \geq 1 \mid \forall n \geq m, \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{\text{SAA}}, F) \leq \tau \right\} \\ N^{\text{UB}}(\tau) &:= \min \left\{ m \geq 1 \mid \forall n \geq m, U(n) \leq \tau \right\}. \end{aligned}$$

Notably, the exact analysis developed in the present paper yields a number of samples two orders of magnitude lower than the best known guarantee to date. The improvements above stem from the novel type of analysis conducted that enables to quantify the implications of all out-of-sample

q	Bound used	Expected relative regret target (τ)				
		25%	20%	15%	10%	5%
.7	$N^{\text{UB}}(\tau)$ (best known to date)	1,696	2,594	4,510	9,921	38,779
	$N^{\text{exact-SAA}}(\tau)$ (this paper)	8	11	15	31	84
.8	$N^{\text{UB}}(\tau)$	2,544	3,890	6,765	14,881	58,168
	$N^{\text{exact-SAA}}(\tau)$	11	16	21	41	116
.9	$N^{\text{UB}}(\tau)$	5,088	7,780	13,530	29,762	100,000+
	$N^{\text{exact-SAA}}(\tau)$	21	23	42	71	210

Table 2: **Number of samples ensuring that SAA achieves a target relative regret.** The table reports induced number of samples needed to reach a relative regret accuracy level, comparing the best instance-independent known bounds to date $U(n)$ (Levi et al., 2015) and the exact worst-case analysis of SAA developed in Theorem 2, for different values of the critical fractile q .

“mistakes” (compared to the oracle) that SAA could do, compared to the existing approaches for SAA analysis that are anchored around large deviations bounds to ensure near-optimality of the SAA solution.

Another fundamental insight from Table 2 stems from the actual values of the minimum number of samples N^{exact} to ensure a particular relative regret level. For example, less than 71 samples are sufficient to achieve a relative regret of 10% for the various critical fractiles above! Theorem 2 and the associated bounds enable to develop a new understanding of the value of data sizes, highlighting that smaller data sizes are extremely valuable and lead to very effective decisions for this class of problems. In practice, even in a data-rich environments such as online retail, the time granularity used to evaluate the demand is usually at a weekly level. As a consequence, a year of demand data for a single product may only represent tens of samples. The above table highlights that such data sizes already ensure very strong performance.

3.2.2 Transient Regime of Learning for SAA and Non-Monotonicity

In Figure 1, we depict the exact worst-case performance of π^{SAA} for sample sizes ranging from 2 to 100, with a critical fractile of q in $\{0.7, 0.8, 0.9\}$. We emphasize that the performance depicted is the *exact worst-case* relative regret of SAA and not a bound on it. Various observations are striking.

First, we observe that the relative regret decays sharply even after observing very few samples n . Consider the case where $q = .9$. With 10 samples, SAA is guaranteed to achieve a relative regret of 49.3%, with 20 samples it achieves 26.8% and with 100 samples, 8.1%. It highlights again the impressive guarantees that SAA yields for the newsvendor problem even when the number of samples is small. It also shows how good SAA is at capturing information relevant to the underlying

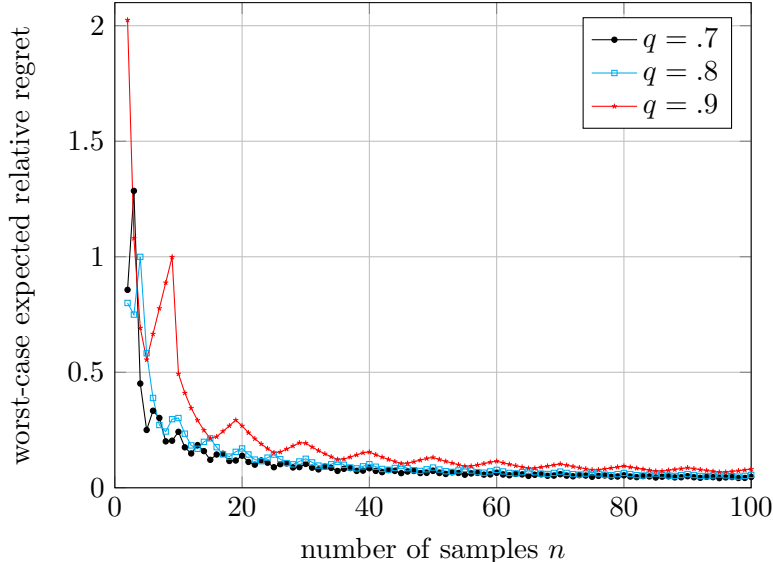


Figure 1: **SAA performance.** The figure depicts the performance of the SAA policy as a function of the number of samples n for different critical fractiles.

optimization problem. Indeed, one would not expect such a quick decay when trying to estimate the entire demand distribution. This in turn leads to a new understanding of the transient regime of learning and the performance possibilities across data sizes, small or large.

Another highly notable observation in Figure 1 is that the worst-case performance of SAA is non-monotone in the number of samples n . The performance curve admits various peaks. We emphasize that the peaks observed are not due to stochasticity when evaluating the performance of the policy but represent an actual deterioration of the performance in the worst-case for SAA when adding a sample. This result can seem counter-intuitive and establishes two notable facts: i) SAA is a suboptimal data-driven policy for various sample sizes; and ii) furthermore, more data is not synonymous with better worst-case performance when using SAA.

Remark (Non-monotonicity). Note that above, when considering the the worst-case relative regret, the worst-case distribution can change with the data size n . Another question could consist in assessing whether there exists a *fixed* distribution F and a data size n such that the performance deteriorates from n samples to $n + 1$ samples (from the same distribution F). In a few problem classes, examples have been exhibited such as pricing, from one to two samples (Babaioff et al., 2018) and misspecified linear regression (Loog et al., 2019). We next argue that the non-monotonicity observed in the worst-case performance in Figure 1 is actually a stronger

statement as it implies that there exists a distribution $F \in \mathcal{F}$ and a data size n such that,

$$\mathcal{R}_n(\pi^{\text{SAA}}, F) < \mathcal{R}_{n+1}(\pi^{\text{SAA}}, F).$$

Indeed, let n be such that the worst-case relative regret is increasing by adding an additional sample to n . In other words, we have

$$\delta := \sup_{F \in \mathcal{F}} \mathcal{R}_{n+1}(\pi^{\text{SAA}}, F) - \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{\text{SAA}}, F) > 0.$$

Then, consider a distribution $F^* \in \mathcal{F}$ such that $\mathcal{R}_{n+1}(\pi^{\text{SAA}}, F^*) > \sup_{F \in \mathcal{F}} \mathcal{R}_{n+1}(\pi^{\text{SAA}}, F) - \delta$. We have

$$\mathcal{R}_n(\pi^{\text{SAA}}, F^*) \leq \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{\text{SAA}}, F) = \sup_{F \in \mathcal{F}} \mathcal{R}_{n+1}(\pi^{\text{SAA}}, F) - \delta < \mathcal{R}_{n+1}(\pi^{\text{SAA}}, F^*).$$

This shows that the worst-case non-monotonicity of the relative regret implies the existence of an instance F^* for which the relative regret is non-monotonic.

In Section 4, we explore some intuition underlying this shortcoming of SAA, but also characterize an optimal data-driven algorithm.

4 Optimal Data-Driven Policy

While SAA is a natural and widely used data-driven policy, we observed in Figure 1 that the performance of SAA is not monotonically decreasing as a function of the number of samples n , implying that it is suboptimal from a minimax perspective. Therefore a natural question is how to improve upon SAA and more generally if it is possible to characterize an optimal data-driven policy in the general space of mappings from data to decisions. Recall that we refer to the optimal policy as the one that solves the minimax optimization problem defined in (3). In this section, we investigate the minimax relative regret \mathcal{R}_n^* presented in equation (3) and associated optimal policies. Compared to solving the worst-case distribution for a particular algorithm, solving (3) now involves two non-parametric and infinite dimensional optimization problems.

We approach the problem as follows. We first establish a fundamental reduction in the space of policies and show that one can restrict attention to mixture of order statistics policies (introduced in Definition 1), without loss of optimality. In this class, we leverage the structure of the problem (2) for mixture of order statistics that we established in Section 3 and we derive a necessary condition that a mixture of order statistics policy needs to satisfy to be optimal in this subclass. We then

show that it is possible to construct a “simple” policy that satisfies this necessary condition. This policy is our candidate optimal policy. The worst-case performance of this policy naturally leads to an upper bound on \mathcal{R}_n^* . To establish that this policy is actually optimal *in the entire class of data-driven policies* Π_n , we introduce an alternative minimax problem which is equal to \mathcal{R}_n^* and in which we extend the space of strategies that nature may take, to randomized ones. For this minimax problem, we construct a candidate prior over the space of distributions and show that the candidate policy above, together with the candidate prior, form a saddle point. This yields the optimality of the candidate policy but also a characterization of its performance.

4.1 Space Reduction from Arbitrary Mappings to Order Statistics

We first reduce the minimax optimization problem (3) involving two non-parametric infinite dimensional optimization problems to a minimax problem over two finite dimensional spaces.

Our next result shows that (3) is equivalent to an optimization problem over the space of mixture of order statistic policies which has a much simpler structure than the general set of mappings from data to decisions. More formally, we show the following.

Theorem 3. *For any $n \geq 1$,*

$$\inf_{\pi \in \Pi_n} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) = \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^\lambda, F).$$

Theorem 3 enables a crucial space reduction of the policy space. In particular, it allows us to reduce our optimization problem to the space of mixture of order statistics policies which is parametrized by the n -dimensional vector of probabilities λ . A notable step in the proof of the theorem consists in showing that,

$$\inf_{\pi \in \Pi_n} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi, \mathcal{B}(\mu)) = \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)). \quad (10)$$

This equation complements Theorem 1 which states that Bernoulli distributions are the worst-case distribution against mixture of order statistic policies. On the other hand, (10) implies that mixture of order statistics policies are the best data-driven policies when facing a Bernoulli distribution. This result is established through a series of reductions, without loss of optimality, in the space of policies. We first show that against Bernoulli distributions, one may restrict attention to policies prescribing inventory in the support of the historical demands. Second, we show that one may restrict attention to policies that prescribe identical inventory conditional on the number of ones observed. Third, we show that one may restrict attention to policies that prescribe a monotonically

increasing inventory as the number of ones observed grows. We finally show that for any policy in the latter class, there exists a mixture of order statistics policy incurring a (weakly) lower cost.

Moreover, by leveraging the characterization of worst-case performance for mixture of order statistics policies derived in Theorem 3, we obtain that

$$\mathcal{R}_n^* = \inf_{\pi \in \Pi_n} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) \stackrel{(a)}{=} \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^\lambda, F) \stackrel{(b)}{=} \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)).$$

where (a) holds by Theorem 3 and (b) follows from Theorem 1.

This implies that Problem (3) is equivalent to the following problem

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)), \quad (11)$$

which now involves optimization over two finite dimensional spaces. In Section 4.2 we construct a candidate optimal policy for (11).

4.2 Candidate Policy for Optimality

In general, prescribing a single order statistic policy can be suboptimal. However, there are particular cases in which extremal policies (either prescribing the minimum sample or the maximum one) achieve optimality. We first describe degenerate cases in which extremal order statistics are optimal.

Proposition 2. *For every n ,*

1. *If $\sup_{\mu \in [0,1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) > \sup_{\mu \in [1-q,1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu))$, then π^{OS_1} is optimal for Problem (3).*
2. *If $\sup_{\mu \in [0,1-q]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)) < \sup_{\mu \in [1-q,1]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu))$, then π^{OS_n} is optimal for Problem (3).*

This result implies that the optimal performance is obtained by extremal order statistics under some particular conditions. Note that these two conditions cannot hold simultaneously (we formally discuss this in Lemma E-2 in Appendix E). We highlight here that the conditions of Proposition 2 do not hold for all data sizes. As a matter of fact, we formally show in Lemma E-3, stated and proved in Appendix E, that these do not hold for any $n \geq \frac{2}{\min(q,1-q)^2}$. Next, we analyze the structure of optimal policies when the conditions do not hold. To that effect, we introduce the following assumption.

Assumption 1. We say that a data size n is non-degenerate if the following two conditions on the performance of extremal order statistics policies hold

$$\sup_{\mu \in [0, 1-q]} \mathcal{R}_n \left(\pi^{OS_1}, \mathcal{B}(\mu) \right) \leq \sup_{\mu \in [1-q, 1]} \mathcal{R}_n \left(\pi^{OS_1}, \mathcal{B}(\mu) \right) \quad (12)$$

$$\sup_{\mu \in [0, 1-q]} \mathcal{R}_n \left(\pi^{OS_n}, \mathcal{B}(\mu) \right) \geq \sup_{\mu \in [1-q, 1]} \mathcal{R}_n \left(\pi^{OS_n}, \mathcal{B}(\mu) \right). \quad (13)$$

In the case in which Assumption 1 holds, one may benefit from randomization. Next, we establish a necessary condition satisfied for a mixture of order statistics policies to solve (11).

Proposition 3. For every n such that Assumption 1 holds, for any solution $\pi^\lambda \in \Pi^{OS}$ that achieves the infimum in Problem (11), the solution π^λ must satisfy

$$\sup_{\mu \in [0, 1-q]} \mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right) = \sup_{\mu \in [1-q, 1]} \mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right). \quad (14)$$

Proposition 3 establishes a necessary condition for a mixture of order statistics policy to be optimal for (11). In particular, π^λ must balance between worst-cases among Bernoulli distributions with mean smaller than $1-q$ and ones with mean larger than $1-q$. Intuitively, if the policy does not satisfy this property, it is possible to improve it by adding mass on lower or higher order statistics.

In Section 3.2.2, we observed that the worst-case performance of SAA is not monotonic as the number of samples grows and deduced its suboptimality. Proposition 3 highlights why this is the case. One can show that SAA does not satisfy (14) in general, and by not doing so enables nature to exploit the imbalance in worst cases to “hurt” the decision-maker. We present in Appendix F.1 a more detailed discussion about the sub-optimality of SAA.

Our next result establishes that it is possible to construct a simple mixture of order statistics policy that satisfies (14), and randomizes between at most two consecutive order statistics.

Proposition 4. For every n such that Assumption 1 holds, there exist $k \in \{2, \dots, n\}$ and $\gamma \in [0, 1]$ such that the policy $\pi^{k,\gamma}$ that prescribes the order statistic $D_{k:n}$ w.p γ and $D_{k-1:n}$ w.p $1-\gamma$ satisfies (14) i.e., there exist $\mu^- \in [0, 1-q]$ and $\mu^+ \in [1-q, 1]$ such that,

$$\mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu^-) \right) = \mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu^+) \right) = \sup_{\mu \in [0, 1]} \mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu) \right). \quad (15)$$

Moreover, k satisfies

$$\sup_{\mu \in [1-q, 1]} \mathcal{R}_n \left(\pi^{OS_{k-1}}, \mathcal{B}(\mu) \right) \geq \sup_{\mu \in [0, 1-q]} \mathcal{R}_n \left(\pi^{OS_{k-1}}, \mathcal{B}(\mu) \right) \quad (16)$$

$$\sup_{\mu \in [1-q, 1]} \mathcal{R}_n \left(\pi^{OS_k}, \mathcal{B}(\mu) \right) \leq \sup_{\mu \in [0, 1-q]} \mathcal{R}_n \left(\pi^{OS_k}, \mathcal{B}(\mu) \right). \quad (17)$$

In other words, Proposition 4 intuitively characterizes the simplest candidate optimal mixture of order statistics policy one could consider when no single order statistic policy satisfies the necessary condition (14).

This candidate policy alleviates the imbalance of the expected relative regret incurred by single order statistic policies. Indeed, letting k denote the largest order statistic prescribed by the candidate policy, we have by (17) that the worst case performance of π^{OS_k} on Bernoulli distributions with relatively small mean supersedes the one for Bernoulli distributions with large ones. On the contrary, according to (16), this imbalance is reverted for $\pi^{OS_{k-1}}$.

Based on Proposition 4, we have a candidate policy $\pi^{k,\gamma}$ satisfying a necessary condition for optimality for Problem (11). This policy induces an upper bound on the value of (11) as we have

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0, 1]} \mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right) \leq \sup_{\mu \in [0, 1]} \mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu) \right).$$

In Section 4.3, we show that the candidate policy $\pi^{k,\gamma}$ not only satisfies a necessary optimality condition for order statistic policies, but is actually optimal in this space of policies which, by Theorem 3, implies its optimality in the general space of data-driven policies Π_n .

4.3 Optimal Data-Driven Policy and its Performance

After deriving a candidate optimal policy, we now show that this policy is optimal for the initial Problem (3) by proving its optimality for (11). Remark that for $n \geq 1$, Problem (11) is equivalent to the following problem in which we extend the space of Bernoulli distributions to the space of distributions over Bernoulli distributions

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{p \in \Delta([0, 1])} \mathbb{E}_{\mu \sim p} \left[\mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right) \right], \quad (18)$$

where $\Delta([0, 1])$ is the set of distributions supported on $[0, 1]$. Furthermore, we have

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{p \in \Delta([0, 1])} \mathbb{E}_{\mu \sim p} \left[\mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right) \right] \geq \sup_{p \in \Delta([0, 1])} \inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p} \left[\mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right) \right].$$

To derive a lower bound matching the upper bound of Section 4.1, it is sufficient to show that there exists a prior p^* , such that the policy $\pi^{k,\gamma}$ introduced in Proposition 4 satisfies,

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p^*} \left[\mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right) \right] = \mathbb{E}_{\mu \sim p^*} \left[\mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu) \right) \right], \quad (19)$$

$$\mathbb{E}_{\mu \sim p^*} \left[\mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu) \right) \right] = \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu) \right). \quad (20)$$

Equality (19) would imply that the policy $\pi^{k,\gamma}$ presented in Proposition 4 is the best response when Nature selects prior p^* . Equality (20) would ensure that prior p^* leads to the worst-case performance of $\pi^{k,\gamma}$.

Consider $\mu^- \in [0, 1 - q]$ and $\mu^+ \in [1 - q, 1]$ as introduced in Proposition 4. Note that (15) implies that for any prior p_0 supported on $\{\mu^-, \mu^+\}$, we have

$$\mathbb{E}_{\mu \sim p_0} \left[\mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu) \right) \right] = \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu) \right).$$

It follows that (20) holds for any such prior. This motivates restricting attention to the set of priors supported on two Bernoulli distributions. Our next result shows that in the class of priors over two Bernoulli distributions, there exists a prior for which (19) holds. Formally we show the following.

Proposition 5. *For any $k \in \{2, \dots, n\}$, $\gamma \in [0, 1]$, $\mu^- \in (0, 1 - q)$ and $\mu^+ \in (1 - q, 1)$, there exists a prior p^* on $\{\mu^-, \mu^+\}$ such that,*

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p^*} \left[\mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right) \right] = \mathbb{E}_{\mu \sim p^*} \left[\mathcal{R}_n \left(\pi^{k,\gamma}, \mathcal{B}(\mu) \right) \right].$$

We are now in a position to state our next main result. The next result provides a characterization of an optimal policy and its performance. In particular, we build on Proposition 5 and on the upper bound derived in Section 4.1 to establish that, when Assumption 1 holds, an optimal data-driven policy, in the *entire* space of possible mappings from data to decision, is given by a randomization over at most two consecutive order statistics of the historical demand samples (in the case where one of the conditions does not hold, we have already established that an extremal order statistic is optimal). Formally we show the following.

Theorem 4 (Optimal Data-Driven Policy). *For every n such that Assumption 1 holds, there exists $k \in \{2, \dots, n\}$ and $\gamma \in [0, 1]$ such that the policy $\pi^{k,\gamma}$ that prescribes the order statistic $D_{k:n}$ w.p γ and $D_{k-1:n}$ w.p $1 - \gamma$ satisfies*

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n \left(\pi^{k,\gamma}, F \right) = \mathcal{R}_n^*.$$

Moreover, k satisfies (16) and (17).

Furthermore, if (12) does not hold, π^{OS_1} is optimal for Problem (3). Similarly, if (13) does not hold, π^{OS_n} is optimal for Problem (3).

This result provides a full characterization of an *optimal* data-driven policy across data sizes. Notably, *i.*) an optimal policy and associated optimal performance can be characterized for this class of problems; and *ii.*) the optimal data-driven policy takes a surprisingly simple structure: it randomizes between two consecutive order statistics. This result allows not only to obtain an optimal algorithm but also to *quantify exactly the robust value of data* associated with historical demand for this class of problems.

Remark. (A “better” minimax optimal policy) A corollary of Theorem 4 is that the *deterministic* policy which selects the inventory level equal to $(1 - \gamma)D_{k-1:n} + \gamma D_{k:n}$ is also minimax optimal (where k and γ are the parameters defined in Theorem 4). In addition, this policy is uniformly better (across all instances) than the minimax optimal mixture of order statistics policy, and its performance coincides with the latter against Bernoulli distributions on which it yields the same worst-case relative regret. We formalize this in Corollary 1 below.

Corollary 1. For every $n \geq 1$. Let $\pi^{k,\gamma}$ be the minimax optimal policy defined in Theorem 4 and let $\pi^{cvx(k,\gamma)}$ be the policy which prescribes the inventory level $\gamma D_{k-1:n} + (1 - \gamma)D_{k:n}$. Then,

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{k,\gamma}, F) = \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{cvx(k,\gamma)}, F) = \mathcal{R}_n^*.$$

Furthermore, for every $F \in \mathcal{F}$,

$$\mathcal{R}_n(\pi^{cvx(k,\gamma)}, F) \leq \mathcal{R}_n(\pi^{k,\gamma}, F).$$

Remark. We also observe that a byproduct of our analysis has implications for the value of data in the Bayesian newsvendor problem. Our analysis shows that there is no gap between the frequentist problem (3) and its bayesian counterpart in the sense that, against the worst prior (which is a randomization between two Bernoulli distributions), the Bayesian problem is as hard (in the sense of the value of data) as the frequentist one and achieves the same worst-case relative regret.

4.4 Optimal Performance and the Robust Value of Data

Algorithm 1 (presented in Appendix F.3) enables us to compute the performance of the optimal policy defined in Theorem 4. Figure 2 presents a comparison of the performance of SAA and the best achievable performance for a data-driven policy for different critical fractiles.

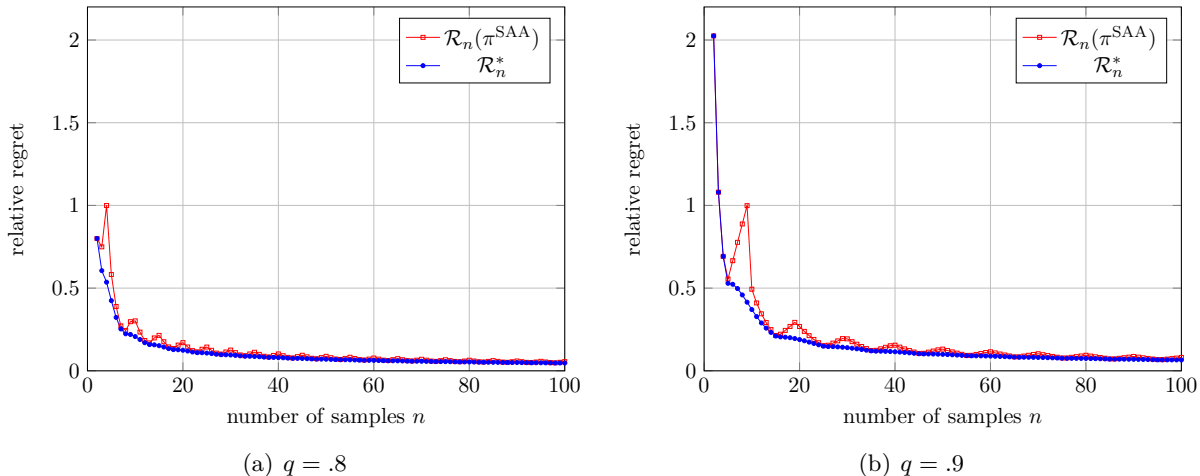


Figure 2: **Optimal performance.** The figure depicts optimal performance versus the performance of SAA as a function of the number of samples n for different critical fractiles.

In this plot, the curve associated to the optimal policy describes the exact value of historical demand data in the newsvendor problem. It gives a clear sense of the inherent hardness of this class of data-driven problems. Deriving the full spectrum of performances for both SAA and the optimal data-driven policy shows that SAA can be considerably improved when the number of samples is relatively small. In particular, when $q = .9$, the relative regret for SAA at $n = 9$ can be reduced by more than 50% by using the optimal policy, and for $n = 19$, it can be reduced by 33%. We also remark that the performance of SAA matches more closely the optimal one as n becomes large. The amplitudes of the peaks decrease as the number of samples increases. We further explore the asymptotic performance in Section 5.

In Table 3, we present a comparison of the number of samples required to guarantee various levels of relative regret for different values of the critical fractile for both SAA and the optimal policy. Recall the definition of $N^{\text{exact-SAA}}$ presented in Section 3.2.1. We similarly define, the number of samples required to achieve a given performance threshold $\tau \geq 0$ when using the optimal policy, as

$$N^{\text{opt}}(\tau) := \min \left\{ m \geq 1 \mid \forall n \geq m, \mathcal{R}_n^* \leq \tau \right\}.$$

We observe that the number of samples required to ensure a particular level of accuracy across

q	Bound used	Expected relative regret target (τ)				
		25%	20%	15%	10%	5%
.7	$N^{\text{exact-SAA}}(\tau)$	8	11	15	31	84
	$N^{\text{opt}}(\tau)$	5	8	12	21	68
.8	$N^{\text{exact-SAA}}(\tau)$	11	16	21	41	116
	$N^{\text{opt}}(\tau)$	8	11	16	28	91
.9	$N^{\text{exact-SAA}}(\tau)$	21	23	42	71	210
	$N^{\text{opt}}(\tau)$	14	19	25	50	161

Table 3: **Number of samples required by SAA and by the optimal policy to ensure a target relative regret.** The table reports the *exact* number of samples needed to reach a relative regret accuracy level, comparing the exact worst-case analysis of SAA developed in Theorem 2, to the optimal minimax performance presented in Theorem 4 for different values of the critical fractile q .

all distributions can be reduced by 17 to 40 % (across the targets tested) when moving from SAA to the minimax optimal policy.

Remark (Structure of the optimal policy). While Theorem 4 does not provide an exact characterization of the parameter k , we have observed numerically that k is either equal to $\lceil qn \rceil$ or $\lceil qn \rceil + 1$. We discuss in more details this aspect in Appendix F.2. As a consequence, the optimal policy can be interpreted as a correction of SAA.

5 Asymptotic Analysis of Optimal Performance

We derived in Section 4 a characterization of an optimal data-driven policy for an arbitrary finite number of samples. We now provide a simple approximation of the optimal performance as the number of samples grows large and derive the exact convergence rate of the minimax relative regret to 0 with its associated multiplicative constant.

In this section, as we study what happens when n changes, we will introduce the notion of a policy sequence. A policy sequence is defined as a sequence $\boldsymbol{\pi} := (\pi_n)_{n \geq 1}$ of mappings where for every $n \geq 1$, we have $\pi_n \in \Pi_n$. For example, $\boldsymbol{\pi}^{\text{SAA}}$ denotes the sequence of policies such that for any $n \geq 1$,

$$\pi_n^{\text{SAA}}(\mathbf{D}_1^n) = D_{\lceil qn \rceil : n}.$$

There are three types of asymptotic results one could consider. A first characterization of the performance, which is typically referred to as consistency or first order optimality states that the cost of a data-driven policy converges to the optimal cost as the number of samples goes to infinity.

In our setting, a policy sequence $\boldsymbol{\pi}$ is said to be consistent if

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\boldsymbol{\pi}, F) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

A second, more refined characterization consists in establishing the rate of convergence of the worst-case performance of a data-driven policy. In the data-driven newsvendor model, for a given sequence $(u_n)_{n \in \mathbb{N}}$ converging to 0, we say that the cost of a data-driven policy-sequence $\boldsymbol{\pi}$ converges to zero at rate u_n if

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\boldsymbol{\pi}, F) = \mathcal{O}(u_n) \quad \text{as } n \rightarrow \infty.$$

At the rate level, the best result one can aim at is to prove that a policy-sequence converges at a rate of \mathcal{R}_n^* , in which case we say that the policy achieves rate-optimality.

A third, yet more refined characterization enables a sharper understanding of the asymptotic performance of a data-driven policy. It consists in deriving a sequence equivalent to the relative regret as the number of samples goes large. In particular, for a given sequence $(u_n)_{n \in \mathbb{N}}$ we say that the performance of a data-driven policy-sequence $\boldsymbol{\pi}$ is asymptotically equivalent to u_n if,

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\boldsymbol{\pi}, F) = u_n + o(u_n) \quad \text{as } n \rightarrow \infty.$$

Deriving an equivalent sequence is a much stronger result than the rate of convergence as it requires to characterize the convergence rate *as well as* the multiplicative constant associated with the rate. When a policy-sequence has a performance asymptotically equivalent to \mathcal{R}_n^* , we say that it is rate-optimal at the multiplicative constant level.

From the work of [Levi et al. \(2015\)](#) one may derive consistency results and the rate of convergence for $\sup_{F \in \mathcal{F}} \mathcal{R}_n(\boldsymbol{\pi}^{\text{SAA}}, F)$. In particular, we show in [Lemma E-5](#), stated and proved in [Appendix E](#), that their bound implies that $\sup_{F \in \mathcal{F}} \mathcal{R}_n(\boldsymbol{\pi}^{\text{SAA}}, F)$ scales at a $\mathcal{O}(1/\sqrt{n})$ rate.

The next result characterizes the asymptotic equivalent of the relative regret for the optimal data-driven policy and establishes that SAA is not only rate-optimal but also rate-optimal at the multiplicative constant level.

Theorem 5 (Optimal Asymptotic Behavior). *i.) The optimal performance \mathcal{R}_n^* converges to zero and satisfies*

$$\mathcal{R}_n^* = \frac{C^*}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \rightarrow \infty, \tag{21}$$

where

$$C^* := \frac{1}{\sqrt{q(1-q)}} \max_{p \geq 0} p(1 - \Phi(p)) \approx \frac{.17}{\sqrt{q(1-q)}},$$

with Φ denoting the cdf of a standard Gaussian distribution.

ii.) In addition, the policy sequence associated with SAA is rate-optimal at the multiplicative constant level. In particular,

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{SAA}, F) = \frac{C^*}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \rightarrow \infty.$$

This result describes the exact rate of convergence of the optimal relative regret as the number of samples goes to infinity. While Sections 3 and 4 yield the first exact results for arbitrary sample sizes, the significant novelty in this section lies in explicitly characterizing more finely the rate of convergence of the performance of the optimal data-driven policy as data grows. Indeed, we derive a semi-closed form expression of the exact constant C^* associated with the rate of convergence of the optimal policy for this class of problems. This expression highlights the role of the critical fractile q in affecting optimal relative regret performance. Problems with high and low values of q are “harder” in that they lead to higher constant C^* , and in turn slower convergence to zero.

In addition, we are able to establish that, while SAA was suboptimal for finite samples in general, it satisfies a very strong form of near-optimality when the number of samples is large. While SAA leads to relative regret that converges to zero at rate $\mathcal{O}(1/\sqrt{n})$, it also leads to the *optimal* constant that one could achieve at this rate of convergence.

By leveraging our novel analysis across all data sizes, we derive new insights in the asymptotic regime. Therefore, understanding more finely the performance of data-driven policies with finite data also improves our understanding of their performance as the number of samples goes to infinity.

6 Instance-Dependent Performance

Our approach enables us to develop a sharp understanding of the robust performance of SAA and of a minimax optimal policy for the data-driven newsvendor problem. Our analysis quantifies exactly the worst-case performance of these algorithms when the worst-case is taken over the whole class of distributions with finite first moment without any shape restriction. In this section, to illustrate the range of possible performances that could emerge, we compute the empirical performances of both algorithms against various distributions: Uniform, Exponential, Lognormal and Pareto. Note that these distribution families are the ones used in (Levi et al., 2015, Table 1) which are supported on $[0, \infty)$.

We compute numerically the expected regret of a data-driven policy π that uses n samples against a *fixed* distribution by repeating independently $M = 10^5$ times the following procedure. For

every $m \in \{1, \dots, M\}$, we first draw n independent samples $\{d_1^m, \dots, d_n^m\}$ representing in-sample demand realizations. We then draw independently $\{\tilde{d}_1^m, \dots, \tilde{d}_K^m\}$ (where $K = 1000$) samples to compute the out-of-sample cost. We finally draw a decision realization \tilde{x}^m from the distribution $\pi(d_1^m, \dots, d_n^m)$ and compute the average realized cost $\tilde{c}^m = \frac{1}{K} \sum_{k=1}^K c(\tilde{x}^m, \tilde{d}_k^m)$. Our estimator of the expected relative regret for policy π is finally defined as

$$\frac{1}{M} \sum_{m=1}^M \left(\frac{\tilde{c}^m}{\text{opt}(F)} - 1 \right).$$

Table 4 presents the number of samples³ required to achieve a target accuracy level for both SAA and the minimax optimal policy presented in Corollary 1.

Policy	Distribution	Expected relative regret target (τ)				
		25%	20%	15%	10%	5%
SAA	Worst-case (Bernoulli)	21	23	42	71	210
	Uniform(0,1)	6	11	12	14	25
	Exponential(1)	7	10	13	20	40
	Log-normal($\mu = 1, \sigma = 1.805$)	10	10	10	20	40
	Pareto($\alpha = 1.5, x_m = 1$)	10	16	16	20	93
Minimax Optimal Policy	Worst-case (Bernoulli)	14	19	25	50	161
	Uniform(0,1)	6	7	10	13	22
	Exponential(1)	6	8	10	18	37
	Log-normal($\mu = 1, \sigma = 1.805$)	9	11	14	19	36
	Pareto($\alpha = 1.5, x_m = 1$)	10	16	16	18	93

Table 4: **Number of samples required by SAA and by the minimax optimal policy $\pi^{\text{cvx}(k,\gamma)}$ (cf. Corollary 1) to achieve a target relative regret.** The table reports a numerical estimation of the number of samples needed to reach a relative regret accuracy level against several fixed distributions. The worst-case line indicates the *exact* number of samples required to achieve a certain target performance level.

While the minimax policy is optimized relatively to a rather conservative measure, it is notable that its performance is on par or most often better than the one of SAA even in “mild” cases. In other words, the “robustification” of SAA provides significant benefits in the worst case along with improvements against a variety of “mild” distributions.

7 Conclusion

In this paper, we investigate the central class of data-driven newsvendor problems. We analyze the performance of the central SAA algorithm across all data sizes and establish a characteriza-

³We report the minimum number of samples such that the upper bound of the 95% confidence interval is below the desired relative regret target.

tion of its *actual* worst-case performance. The exact performance characterization of this widely studied policy leads to a new understanding of the economics of data sizes, highlighting the very strong performance achievable with limited data. At the same time, it also demonstrates a notable phenomenon: when using SAA, more data is not synonymous with better worst-case performance.

In turn, we optimize over the entire space of data-driven algorithms that maps data to decisions and derive an optimal algorithm (in the minimax sense) and its associated performance. This provides the first optimality result in this class of data-driven problems. It also perfectly quantifies the value of data and the potential associated with corrections to the classical SAA algorithm, especially with smaller data sizes. It further emphasizes that for this class of problems, a decision-maker may operate efficiently even in environments with limited data.

Finally, we provide a simple approximation of the optimal worst-case performance achievable by a data-driven algorithm when the number of samples is large. In particular, we leverage our exact analysis across all data sizes to characterize the exact rate of convergence of the minimax relative regret and characterize in semi-closed form the multiplicative constant associated with it. We further show that while SAA is suboptimal in general, it is rate-optimal at the multiplicative constant level when the number of samples is large.

The present paper offers a new lens, that of the transient regime of learning, through which some data-driven problems may be approached, but also highlights the possibility to operate effectively with limited data. There are many avenues for future research, ranging from exploring the possibility of performance characterization and optimization across data sizes for sequential decision-making problems with different information structures (e.g., censoring) to exploring the transient regime of learning in contextual newsvendor problems, or more general stochastic problem classes.

Acknowledgment

The authors are grateful to the editor, the associate editor and two anonymous reviewers whose valuable suggestions lead to many significant improvements in the final version of the paper. They also thank Nick Arnosti, Santiago Balseiro, Gah-Yi Ban, Omar El Housni, Yale T. Herer, Nathan Kallus, Will Ma, and Dan Russo for their questions and comments which helped improve this work.

References

Azoury, K. S. (1985), ‘Bayes solution to dynamic inventory models under unknown demand distribution’, *Management Science* **31**(9).

- Babaioff, M., Gonczarowski, Y. A., Mansour, Y. and Moran, S. (2018), Are two (samples) really better than one?, in ‘Proceedings of the 2018 ACM Conference on Economics and Computation’, pp. 175–175.
- Ban, G.-Y. (2020), ‘Confidence intervals for data-driven inventory policies with demand censoring’, *Operations Research* **68**(2), 309–326.
- Ban, G.-Y. and Rudin, C. (2019), ‘The big data newsvendor: Practical insights from machine learning’, *Operations Research* **67**(1), 90–108.
- Bensoussan, A., Cakanyildirim, M. and Sethi, S. (2009), ‘Technical note: The censored newsvendor and the optimal acquisition of information’, *Operations Research* **57**, 791–794.
- Bertsimas, D., Gupta, V. and Kallus, N. (2018), ‘Robust sample average approximation’, *Mathematical Programming* **171**(1-2), 217–282.
- Besbes, O., Chaneton, J. and Moallemi, C. (2022), ‘The exploration-exploitation tradeoff in the newsvendor problem’, *Stochastic Systems (Articles in Advance)* .
- Besbes, O. and Muharremoglu, A. (2013), ‘On implications of demand censoring in the newsvendor problem’, *Management Science* **59**(6), 1407–1424.
- Chen, B., Chao, X. and Shi, C. (2021), ‘Nonparametric learning algorithms for joint pricing and inventory control with lost sales and censored demand’, *Mathematics of Operations Research* **46**(2), 726–756.
- Cheung, W. C. and Simchi-Levi, D. (2019), ‘Sampling-based approximation schemes for capacitated stochastic inventory control models’, *Mathematics of Operations Research* **44**(2), 668–692.
- Chu, L. Y., Shanthikumar, J. G. and Shen, Z.-J. M. (2008), ‘Solving operational statistics via a bayesian analysis’, *Operations Research Letters* **36**(1), 110–116.
- Ding, X., Puterman, M. L. and Bisi, A. (2002), ‘The censored newsvendor and the optimal acquisition of information’, *Operations Research* **50**(3).
- Durrett, R. (2019), *Probability: theory and examples*, Vol. 49, Cambridge university press.
- Elmachtoub, A. N. and Grigas, P. (2021), ‘Smart “predict, then optimize”’, *Management Science* .
- Esfahani, P. M. and Kuhn, D. (2018), ‘Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations’, *Mathematical Programming* **171**(1), 115–166.
- Gallego, G. and Moon, I. (1993), ‘The distribution free newsboy problem: review and extensions’, *Journal of the Operational Research Society* **44**(8), 825–834.
- Godfrey, G. A. and Powell, W. B. (2001), ‘An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution’, *Management Science* **47**(8), 1101–1112.
- Gupta, V. and Kallus, N. (2022), ‘Data pooling in stochastic optimization’, *Management Science* **68**(3), 1595–1615.
- Harrell, F. E. and Davis, C. (1982), ‘A new distribution-free quantile estimator’, *Biometrika* **69**(3), 635–640.
- Hoeffding, W. (1994), Probability inequalities for sums of bounded random variables, in ‘The Collected Works of Wassily Hoeffding’, Springer, pp. 409–426.

- Huh, W. T., Levi, R., Rusmevichientong, P. and Orlin, J. B. (2011), ‘Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator’, *Operations Research* **59**(4), 929–941.
- Huh, W. T. and Rusmevichientong, P. (2009), ‘A nonparametric asymptotic analysis of inventory planning with censored demand’, *Mathematics of Operations Research* **34**(1), 103–123.
- Kalgh, W. and Lachenbruch, P. A. (1982), ‘A generalized quantile estimator’, *Communications in Statistics - Theory and Methods* **11**(19), 2217–2238.
- Kleywegt, A. J., Shapiro, A. and Homem-de Mello, T. (2002), ‘The sample average approximation method for stochastic discrete optimization’, *SIAM Journal on Optimization* **12**(2), 479–502.
- Lam, H. (2021), ‘On the impossibility of statistically improving empirical optimization: A second-order stochastic dominance perspective’, *arXiv preprint arXiv:2105.13419* .
- Lariviere, M. and Porteus, E. L. (1999), ‘Stalking information: Bayesian inventory management with unobserved lost sales’, *Management Science* **45**(3), 346–363.
- Levi, R., Pál, M., Roundy, R. O. and Shmoys, D. B. (2007), ‘Approximation algorithms for stochastic inventory control models’, *Mathematics of Operations Research* **32**(2), 284–302.
- Levi, R., Perakis, G. and Uichanco, J. (2015), ‘The data-driven newsvendor problem: New bounds and insights’, *Operations Research* **63**(6), 1294–1306.
- Liyanaige, L. H. and Shanthikumar, J. G. (2005), ‘A practical inventory control policy using operational statistics’, *Operations Research Letters* **33**(4), 341–348.
- Loog, M., Viering, T. and Mey, A. (2019), Minimizers of the empirical risk and risk monotonicity, in ‘Advances in Neural Information Processing Systems’, pp. 7478–7487.
- Lu, X., Song, J.-S. and Zhu, K. (2005), ‘On “the censored newsvendor and the optimal acquisition of information”’, *Operations Research* **53**(6), 1024–1026.
- Lugosi, G., Markakis, M. and Neu, G. (2021), ‘On the hardness of learning from censored demand’, *Available at SSRN 3509255* .
- Maglaras, C. and Eren, S. (2015), ‘A maximum entropy joint demand estimation and capacity control policy’, *Production and Operations Management* **24**(3), 438–450.
- Natarajan, K., Sim, M. and Uichanco, J. (2018), ‘Asymmetry and ambiguity in newsvendor models’, *Management Science* **64**(7), 3146–3167.
- Perakis, G. and Roels, G. (2008), ‘Regret in the newsvendor model with partial information’, *Operations Research* **56**(1), 188–203.
- Qi, M., Cao, Y. and Shen, Z.-J. (2021), ‘Distributionally robust conditional quantile prediction with fixed design’, *Management Science* .
- Rukhin, A. L. (1983), ‘A class of minimax estimators of a normal quantile’, *Statistics & Probability Letters* **1**(5), 217–221.
- Rukhin, A. L. and Strawderman, W. E. (1982), ‘Estimating a quantile of an exponential distribution’, *Journal of the American Statistical Association* **77**(377), 159–162.
- Saghafian, S. and Tomlin, B. (2016), ‘The newsvendor under demand ambiguity: Combining data with moment and tail information’, *Operations Research* **64**(1), 167–185.
- Scarf, H. (1958), ‘A min-max solution of an inventory problem’, *Studies in the mathematical theory of inventory and production* .

- Scarf, H. (1959), ‘Bayes solutions of the statistical inventory problem’, *The annals of mathematical statistics* **30**(2), 490–508.
- Shapiro, A. (2008), ‘Stochastic programming approach to optimization under uncertainty’, *Mathematical Programming* **112**(1), 183–220.
- Swamy, C. and Shmoys, D. B. (2005), Sampling-based approximation algorithms for multi-stage stochastic optimization, in ‘46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)’, IEEE, pp. 357–366.
- Tierney, L. (1983), ‘A space-efficient recursive procedure for estimating a quantile of an unknown distribution’, *SIAM Journal on Scientific and Statistical Computing* **4**(4), 706–711.
- van Ryzin, G. and McGill, J. (2000), ‘Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels’, *Management Science* **46**(6), 760–775.
- Viering, T. and Loog, M. (2021), ‘The shape of learning curves: a review’, *arXiv preprint arXiv:2103.10948* .
- Xu, L., Zheng, Y. and Jiang, L. (2021), ‘A robust data-driven approach for the newsvendor problem with nonparametric information’, *Manufacturing & Service Operations Management* .
- Yang, S.-S. (1985), ‘A smooth nonparametric estimator of a quantile function’, *Journal of the American Statistical Association* **80**(392), 1004–1011.
- Zieliński, R. (1999), ‘Best equivariant nonparametric estimator of a quantile’, *Statistics & probability letters* **45**(1), 79–84.

Electronic Companion: Appendix for
How Big Should Your Data Really Be?
Data-Driven Newsvendor: Learning One Sample at a Time

A Proofs for Section 3

Proof of Proposition 1. Fix $F \in \mathcal{F}$. For every $r \in \{1, \dots, n\}$, let $F_{D_{r:n}}$ denote the distribution of the random variable $D_{r:n}$. We will use the following alternative expression for $c_F(x)$.

Lemma A-1. For any distribution $F \in \mathcal{F}$, and any $x \geq 0$,

$$c_F(x) = b(\mathbb{E}_F[D] - x) + (b + h) \int_0^x F(y) dy.$$

This result is proved in Appendix D. In what follows, we use \bar{F} to denote the complementary cumulative distribution, i.e., $\bar{F} = 1 - F$.

We have

$$\begin{aligned} \mathbb{E}_{x \sim F_{D_{r:n}}}[c_F(x)] &\stackrel{(a)}{=} b(\mathbb{E}_F[D] - \mathbb{E}_{F_{D_{r:n}}}[D_{r:n}]) + (b + h) \int_0^\infty \int_0^s F(y) dy dF_{D_{r:n}}(s) \\ &\stackrel{(b)}{=} b(\mathbb{E}_F[D] - \mathbb{E}_{F_{D_{r:n}}}[D_{r:n}]) + (b + h) \int_0^\infty \int_y^\infty dF_{D_{r:n}}(s) F(y) dy \\ &= b(\mathbb{E}_F[D] - \mathbb{E}_{F_{D_{r:n}}}[D_{r:n}]) + (b + h) \int_0^\infty \bar{F}_{D_{r:n}}(y) F(y) dy \\ &= b \left(\int_0^\infty \bar{F}(y) dy - \int_0^\infty \bar{F}_{D_{r:n}}(y) dy \right) + (b + h) \int_0^\infty \bar{F}_{D_{r:n}}(y) F(y) dy \\ &= (b + h) \left[q \left(\int_0^\infty \bar{F}(y) dy - \int_0^\infty \bar{F}_{D_{r:n}}(y) dy \right) + \int_0^\infty \bar{F}_{D_{r:n}}(y) F(y) dy \right] \\ &= (b + h) \left[\int_0^\infty \left(\bar{F}_{D_{r:n}}(y) (F(y) - q) + q(1 - F(y)) \right) dy \right] \\ &\stackrel{(c)}{=} (b + h) \left[\int_0^\infty ((1 - B_{r,n}(F(y)))(F(y) - q) + q(1 - F(y))) dy \right]. \end{aligned}$$

Here, (a) follows from Lemma A-1. Equality (b) follows from Fubini-Tonelli which holds because, $s \mapsto 1$ is a positive function and $(\mathbb{R}, dF_{D_{r:n}})$ and (\mathbb{R}, dx) are complete, σ -finite measure spaces. Moreover, (c) holds because the cumulative distribution function of $D_{r:n}$ satisfies

$$F_{D_{r:n}}(x) = B_{r,n}(F(x)).$$

Therefore, one can derive the desired expression by decomposing the performance of π^λ as follows.

$$\mathcal{C}(\pi^\lambda, F, n) = \mathbb{E}_{\mathbf{D}_1^n \sim F} \left[\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)} [c_F(x)] \right] \stackrel{(a)}{=} \sum_{i=1}^n \lambda_i \mathbb{E}_{\mathbf{D}_1^n \sim F} [c_F(D_{i:n})] = \sum_{i=1}^n \lambda_i \mathbb{E}_{x \sim F_{D_{i:n}}} [c_F(x)],$$

where (a) follows from law of total expectation conditioning on the value of $\pi^\lambda(\mathbf{D}_1^n)$.

Next, we analyze $\text{opt}(F)$. Using Lemma A-1 we can rewrite the optimal cost as

$$\begin{aligned}
\text{opt}(F) &= c_F(x_F^*) \\
&= b(\mathbb{E}_F[D] - x_F^*) + (b+h) \int_0^{x_F^*} F(y) dy \\
&= b \left(\int_0^{x_F^*} \bar{F}(y) dy + \int_{x_F^*}^{\infty} \bar{F}(y) dy - x_F^* \right) + (b+h) \int_0^{x_F^*} F(y) dy \\
&= b \left(\int_{x_F^*}^{\infty} \bar{F}(y) dy - \int_0^{x_F^*} F(y) dy \right) + (b+h) \int_0^{x_F^*} F(y) dy \\
&= b \int_{x_F^*}^{\infty} \bar{F}(y) dy + h \int_0^{x_F^*} F(y) dy \\
&= (b+h) \left[q \int_{x_F^*}^{\infty} \bar{F}(y) dy + (1-q) \int_0^{x_F^*} F(y) dy \right] \\
&= (b+h) \int_0^{\infty} (1-q)F(y) \mathbb{1}\{y < x_F^*\} + q(1-F(y)) \mathbb{1}\{y \geq x_F^*\} dy \\
&\stackrel{(a)}{=} (b+h) \int_0^{\infty} (1-q)F(y) \mathbb{1}\{F(y) < q\} + q(1-F(y)) \mathbb{1}\{F(y) \geq q\} dy \\
&= (b+h) \int_0^{\infty} \min\{(1-q)F(y), q(1-F(y))\} dy.
\end{aligned}$$

(a) holds by definition of x_F^* . □

Proof of Theorem 1. Step 1. For any mixture of order statistics policy π^λ , by plugging the simplified expressions of $\mathcal{C}(\pi^\lambda, F, n)$ and $\text{opt}(F)$ computed in Proposition 1 in the epigraph formulation derived in Lemma E-1, we obtain that problem (2) is equivalent to,

$$\inf_{z \in \mathbb{R}} z \tag{A-1a}$$

$$\text{s.t.} \quad \sup_{F \in \mathcal{F}} \int_0^{\infty} \Psi_z^\lambda(F(y)) dy \leq 0. \tag{A-1b}$$

where $\Psi_z^\lambda : [0, 1] \rightarrow \mathbb{R}$ is such that for every $x \in [0, 1]$,

$$\Psi_z^\lambda(x) = \sum_{i=1}^n \lambda_i [(1 - B_{i,n}(x))(x - q) + q(1 - x) - (z + 1) \min\{(1 - q)x, q(1 - x)\}].$$

Step 2. We next aim to further simplify Problem (A-1). To that end, we establish the following equivalence

$$\sup_{F \in \mathcal{F}} \int_0^{\infty} \Psi_z^\lambda(F(y)) dy \leq 0 \quad \text{if and only if} \quad \sup_{\alpha \in (0,1)} \Psi_z^\lambda(\alpha) \leq 0. \tag{A-2}$$

First assume that $\sup_{\alpha \in (0,1)} \Psi_z^\lambda(\alpha) \leq 0$. Noting that $\Psi_z^\lambda(\cdot)$ is continuous on $[0, 1]$, we also have $\sup_{\alpha \in [0,1]} \Psi_z^\lambda(\alpha) \leq 0$. In such a case, for all $F \in \mathcal{F}$, since $F(y) \in [0, 1]$, we have that $\Psi_z^\lambda(F(y)) \leq 0$

for all $y \geq 0$ and it directly follows that

$$\int_0^\infty \Psi_z^\lambda(F(y))dy \leq 0.$$

Conversely, suppose that $\sup_{F \in \mathcal{F}} \int_0^\infty \Psi_z^\lambda(F(y))dy \leq 0$. Note that for any $z \in \mathbb{R}$, $\Psi_z^\lambda(\cdot)$ is continuous on $[0, 1]$ and therefore, it achieves its maximum on $[0, 1]$. Let $\alpha^* \in \arg \max_{\alpha \in [0, 1]} \Psi_z^\lambda(\alpha)$. Let G be defined by,

$$G(x) = \begin{cases} 0 & \text{if } x < 0 \\ \alpha^* & \text{if } x \in [0, 1] \\ 1 & \text{if } x \geq 1 \end{cases}$$

In turn we have

$$\sup_{\alpha \in (0, 1)} \Psi_z^\lambda(\alpha) = \sup_{\alpha \in [0, 1]} \Psi_z^\lambda(\alpha) = \Psi_z^\lambda(\alpha^*) = \int_0^1 \Psi_z^\lambda(G(y))dy \stackrel{(a)}{\leq} \sup_{F \in \mathcal{F}} \int_0^\infty \Psi_z^\lambda(F(y))dy \leq 0,$$

where (a) holds because $G \in \mathcal{F}$. As a consequence, (A-2) holds.

Furthermore, note that (A-2) implies that problem (A-1) is equivalent to,

$$\inf_{z \in \mathbb{R}} z \tag{A-3a}$$

$$\text{s.t.} \quad \sup_{\alpha \in (0, 1)} \sum_{i=1}^n \lambda_i [(1 - B_{i,n}(\alpha))(\alpha - q) + q(1 - \alpha)] - (z + 1) \min \{(1 - q)\alpha, q(1 - \alpha)\} \leq 0.$$

$$\tag{A-3b}$$

Remark that (A-3) is the epigraph formulation of

$$\sup_{\alpha \in (0, 1)} \sum_{i=1}^n \lambda_i \left[\frac{(1 - B_{i,n}(\alpha))(\alpha - q) + q(1 - \alpha)}{\min \{(1 - q)\alpha, q(1 - \alpha)\}} - 1 \right].$$

Hence, by equivalence between (E-2) and (A-3) we conclude that,

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^\lambda, F) = \sup_{\alpha \in (0, 1)} \sum_{i=1}^n \lambda_i \left[\frac{(1 - B_{i,n}(\alpha))(\alpha - q) + q(1 - \alpha)}{\min \{(1 - q)\alpha, q(1 - \alpha)\}} - 1 \right].$$

For the last step of the proof, we use the following lemma (whose proof is deferred to Appendix D), which establishes that the worst-case computed above is achieved by a Bernoulli distribution.

Lemma A-2. For any $r \in \{1 \dots, n\}$ and $\alpha \in [0, 1]$,

$$\mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(1 - \alpha)) = \frac{(1 - B_{r,n}(\alpha))(\alpha - q) + q(1 - \alpha)}{\min \{(1 - q)\alpha, q(1 - \alpha)\}} - 1.$$

This completes the proof. □

B Proofs for Section 4

Proof of Theorem 3. Fix $n \geq 1$. It is easy to see that

$$\inf_{\pi \in \Pi_n} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) \leq \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^\lambda, F).$$

We now prove that

$$\inf_{\pi \in \Pi_n} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) \geq \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^\lambda, F).$$

To do so, we claim that we only need to show that mixture of order statistics policies are optimal when reducing the space of distributions to Bernoulli ones. Formally, we need to show that

$$\inf_{\pi \in \Pi_n} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi, \mathcal{B}(\mu)) = \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)). \quad (\text{B-1})$$

Indeed, assuming that (B-1) holds, one concludes the proof by remarking that,

$$\begin{aligned} \inf_{\pi \in \Pi_n} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) &\geq \inf_{\pi \in \Pi_n} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi, \mathcal{B}(\mu)) \\ &\stackrel{(a)}{=} \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) \\ &\stackrel{(b)}{=} \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^\lambda, F), \end{aligned}$$

where (a) would follow from (B-1) and (b) is a consequence of Theorem 1.

We now prove (B-1).

We first reduce the set of policies $\pi \in \Pi_n$ without loss of optimality for the following problem.

$$\inf_{\pi \in \Pi_n} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi, \mathcal{B}(\mu)) \quad (\text{B-2})$$

We show that one may restrict attention to policies such that the support of the distribution of inventory is included in the interval defined by the smallest observed demand and the largest one. Formally, the following result is proved in Appendix D.

Lemma B-1. *For any policy $\pi \in \Pi_n$ there exists a policy $\pi' \in \Pi_n$ with a lower cost such that for every $\mathbf{D}_1^n \in \{0, 1\}^n$, the support of $\pi'(\mathbf{D}_1^n)$ is a subset of $[D_{1:n}, D_{n:n}]$.*

Note that Lemma B-1 implies that $\pi'(\mathbf{0}_1^n) = 0$ and $\pi'(\mathbf{1}_1^n) = 1$ where $\mathbf{0}_1^n$ (resp. $\mathbf{1}_1^n$) is the sequence of historical data in which all demand observations are 0 (resp. 1). In turn, we leverage this result to further reduce the space of policies without loss of optimality to the $(n+1)$ -dimensional space of sum-based policies defined as follows.

Definition 2. (*Sum-based policies*) *Consider a sequence $\mathbf{e} = (e_i)_{i \in \{0, \dots, n\}} \in [0, 1]^{n+1}$. We say that a policy $\pi^{\Sigma \mathbf{e}}$ is a sum-based policy if for any $i \in \{0, \dots, n\}$ and any $\mathbf{D}_1^n \in \{0, 1\}^n$, such that $\sum_{j=1}^n D_j = i$, we have that,*

$$\pi^{\Sigma \mathbf{e}}(\mathbf{D}_1^n) = e_i.$$

Let $\pi \in \Pi_n$ be a policy which support is included in the interval defined by the smallest observed demand and the largest one. By Lemma B-1 this restriction is without loss of optimality. We construct a sum-based policy that ensures the same cost as π against any Bernoulli distribution. Define for every $i \in \{0, \dots, n\}$ the set \mathcal{D}_n^i as

$$\mathcal{D}_n^i := \left\{ \mathbf{D}_1^n \in \{0, 1\}^n \mid \sum_{j=1}^n D_j = i \right\}.$$

Moreover, consider the sequence $\mathbf{e} = (e_i)_{i \in \{0, \dots, n\}} \in [0, 1]^{n+1}$ such that for every $i \in \{0, \dots, n\}$

$$e_i = \frac{1}{|\mathcal{D}_n^i|} \sum_{\mathbf{D}_1^n \in \mathcal{D}_n^i} \mathbb{E}_{x \sim \pi(\mathbf{D}_1^n)} [x].$$

By Lemma B-1 we have that $e_i \in [0, 1]$ for all $i \in \{0, \dots, n\}$, $e_0 = 0$ and $e_n = 1$ which implies that $\pi^{\Sigma \mathbf{e}}$ is a well defined sum-based policy.

To ease notations, let S_j denote the event $\{\sum_{i=1}^n D_i = j\}$ for every $j \in \{0, \dots, n\}$. We note that for every $\mu \in [0, 1]$ the cost of the policy π satisfies

$$\begin{aligned} \frac{\mathcal{C}(\pi, \mathcal{B}(\mu), n)}{b+h} &\stackrel{(a)}{=} \frac{1}{b+h} \mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left[\mathbb{E}_{x \sim \pi(\mathbf{D}_1^n)} [\mu b(1-x) + (1-\mu)hx] \right] \\ &= \mu \cdot q + \sum_{i=0}^n (1-\mu-q) \cdot \mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left[\mathbb{E}_{x \sim \pi(\mathbf{D}_1^n)} [x] \mid S_i \right] \cdot \mathbb{P}(S_i) \\ &\stackrel{(b)}{=} \mu \cdot q + \sum_{i=0}^n (1-\mu-q) \cdot \frac{1}{|\mathcal{D}_n^i|} \sum_{\mathbf{D}_1^n \in \mathcal{D}_n^i} \mathbb{E}_{x \sim \pi(\mathbf{D}_1^n)} [x] \cdot \mathbb{P}(S_i) \\ &= \mu \cdot q + \sum_{i=0}^n (1-\mu-q) \cdot e_i \cdot \mathbb{P}(S_i) = \mathcal{C}(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu), n), \end{aligned} \quad (\text{B-3})$$

where (a) holds because the support of $\pi(\mathbf{D}_1^n)$ is a subset of $[D_{1:n}, D_{n:n}]$, which is included in $[0, 1]$ for Bernoulli distributions and (b) follows from the fact that, for Bernoulli distributions, the distribution of \mathbf{D}_1^n conditional on $\{\sum_{j=1}^n D_j = i\}$ is that of a uniform law on \mathcal{D}_n^i .

Lemma B-1 along with (B-3) imply that the minimax problem across the general set of data-driven policies is actually equivalent to a minimax problem in which the space of policies is parameterized by a $(n+1)$ dimensional space. Namely, we have showed that

$$\inf_{\pi \in \Pi_n} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi, \mathcal{B}(\mu)) = \inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu)). \quad (\text{B-4})$$

Recall that for a policy $\pi^{\Sigma \mathbf{e}}$, e_i represents the inventory prescribed by the policy after observing i ones. Natural candidate policies in this space are ones for which the inventory level prescribed is increasing as a function of the number of ones observed in historical data. Our next result formalizes this idea.

Lemma B-2. For any $n \geq 1$,

$$\inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu) \right) = \inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1 \\ (e_i) \text{ non-decreasing}}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu) \right).$$

The proof is deferred to Appendix D.

The last step of our proof consists in showing that the performance of any policy $\pi^{\Sigma \mathbf{e}} \in \Pi_n^{DS}$ such that $e_0 = 0$, $e_n = 1$ and $(e_i)_{i \in \{0, \dots, n\}}$ is non-decreasing, can be reproduced by a mixture of order statistics policy. Consider a sequence $(e_i)_{i \in \{0, \dots, n\}}$ satisfying these assumptions and define the vector of probabilities $\boldsymbol{\lambda}$ such that for all $i \in \{1, \dots, n\}$,

$$\lambda_i = e_{n-i+1} - e_{n-i}.$$

Note that for all $i \in \{1, \dots, n\}$, $\lambda_i \geq 0$ by monotonicity of $(e_i)_{i \in \{0, \dots, n\}}$ and $\sum_{i=1}^n \lambda_i = e_n - e_0 = 1$. Hence $\boldsymbol{\lambda}$ is a well defined probability vector. We now show that the mixture of order statistics policy π^λ incurs the same cost as $\pi^{\Sigma \mathbf{e}}$ against any Bernoulli distribution. Let $\mu \in [0, 1]$, then the cost of π^λ is

$$\begin{aligned} \frac{\mathcal{C} \left(\pi^\lambda, \mathcal{B}(\mu), n \right)}{b+h} &= \mu \cdot q + \sum_{i=0}^n (1 - \mu - q) \cdot \mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left[\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)} [x] \mid S_i \right] \cdot \mathbb{P}(S_i) \\ &\stackrel{(a)}{=} \mu \cdot q + \sum_{i=0}^n (1 - \mu - q) \cdot \sum_{k=1}^n \lambda_k \cdot \mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left[D_{k:n} \mid S_i \right] \cdot \mathbb{P}(S_i) \\ &\stackrel{(b)}{=} \mu \cdot q + \sum_{i=0}^n (1 - \mu - q) \cdot \mathbb{P}(S_i) \cdot \sum_{k=n-i+1}^n \lambda_k \\ &= \mu \cdot q + \sum_{i=0}^n (1 - \mu - q) \cdot \mathbb{P}(S_i) \cdot e_i = \mathcal{C} \left(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu), n \right), \end{aligned}$$

where (a) holds because π^λ prescribes $D_{k:n}$ with probability λ_k for any $k \in \{1, \dots, n\}$ and (b) follows from the fact that for every $k \in \{1, \dots, n\}$,

$$D_{k:n} = \begin{cases} 0 & \text{a.s. if } \sum_{j=1}^n D_j \leq n - k \\ 1 & \text{a.s. if } \sum_{j=1}^n D_j \geq n - k + 1. \end{cases}$$

As a consequence,

$$\inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1 \\ (e_i) \text{ non-decreasing}}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu) \right) \geq \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^\lambda, \mathcal{B}(\mu) \right). \quad (\text{B-5})$$

We finally conclude that,

$$\begin{aligned}
\inf_{\pi \in \Pi_n} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi, \mathcal{B}(\mu)) &\stackrel{(a)}{=} \inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu)) \\
&\stackrel{(b)}{=} \inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1 \\ (e_i) \text{ non-decreasing}}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu)) \\
&\stackrel{(c)}{\geq} \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)),
\end{aligned}$$

where (a) holds by (B-4), (b) follows from Lemma B-2 and (c) is a consequence of (B-5). This completes the proof. \square

Proof of Proposition 2. Assume that,

$$\sup_{\mu \in [0,1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) > \sup_{\mu \in [1-q,q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)). \quad (\text{B-6})$$

We show that π^{OS_1} is an optimal mixture of order statistics policy. Note that for every $r \in \{2, \dots, n\}$ we have

$$\begin{aligned}
\sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) &\stackrel{(a)}{=} \sup_{\mu \in [0,1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) \\
&\stackrel{(b)}{<} \sup_{\mu \in [0,1-q]} \mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(\mu)) \leq \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(\mu)),
\end{aligned}$$

where (a) follows from (B-6) and (b) holds by Lemma E-4 stated and proved in Appendix E.

In turn, for every $\pi^\lambda \in \Pi_n^{OS}$ such that $\lambda_1 < 1$, we have that,

$$\sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) = \sup_{\mu \in [0,1]} \sum_{i=1}^n \lambda_i \mathcal{R}_n(\pi^{OS_i}, \mathcal{B}(\mu)) > \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)).$$

As a consequence, π^{OS_1} is optimal and satisfies,

$$\mathcal{R}_n^* = \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) = \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)).$$

Similarly, assuming that (13) does not hold, we show by a similar argument that π^{OS_n} is optimal for Problem (3). \square

Proof of Proposition 3. Suppose first that

$$\sup_{\mu \in [0,1-q]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) > \sup_{\mu \in [1-q,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)). \quad (\text{B-7})$$

In such a case, we show that there exists an alternative policy with strictly lower worst-case per-

formance.

We first argue that λ must be such that $\lambda_1 < 1$. Indeed, note that, by assumption, we have

$$\sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) \leq \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)), \quad (\text{B-8})$$

The conjunction of (B-7) and (B-8) implies that $\lambda_1 < 1$.

Next we argue that the policy π^{OS_1} is strictly better than π^λ if $\mu \in [0, 1-q]$. We have

$$\begin{aligned} \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) &= \sup_{\mu \in [0, 1-q]} \sum_{i=1}^n \lambda_i \mathcal{R}_n(\pi^{OS_i}, \mathcal{B}(\mu)) \\ &\stackrel{(a)}{>} \sup_{\mu \in [0, 1-q]} \sum_{i=1}^n \lambda_i \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) \\ &= \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)). \end{aligned} \quad (\text{B-9})$$

where (a) follows from the fact that $\lambda_1 < 1$, together with Lemma E-4 stated and proved in Appendix E.

Next, we construct an explicit policy that improves upon π^λ .

For any $\nu \in [0, 1]$, consider the policy $\tilde{\pi}_\nu$ which chooses the policy π^{OS_1} with probability ν and the policy π^λ with probability $1 - \nu$. Remark that $\tilde{\pi}_\nu$ is a mixture of order statistics policy and for any $F \in \mathcal{F}$,

$$\mathcal{R}_n(\tilde{\pi}_\nu, F) = \nu \cdot \mathcal{R}_n(\pi^{OS_1}, F) + (1 - \nu) \cdot \mathcal{R}_n(\pi^\lambda, F).$$

Define the mapping L from $[0, 1]$ to \mathbb{R} such that,

$$L : \nu \mapsto \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\tilde{\pi}_\nu, \mathcal{B}(\mu)) - \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\tilde{\pi}_\nu, \mathcal{B}(\mu)).$$

We first show that L is continuous. Remark that it is sufficient to show that, the following mapping g is continuous.

$$g : \nu \mapsto \sup_{\mu \in [0, 1-q]} f(\nu, \mu),$$

where $f(\nu, \mu) := \mathcal{R}_n(\tilde{\pi}_\nu, \mathcal{B}(\mu))$. First remark that by Lemma A-2, the mapping $\mu \mapsto \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu))$ is continuous for every mixture of order statistic π^λ . Hence, f is continuous in its second component. Moreover, f is affine in its first component, and $f(\cdot, \mu)$ is M -Lipschitz for every $\mu \in [0, 1-q]$, where $M := \sup_{\mu \in [0, 1-q]} |\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) - \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu))|$. Remark that $M < \infty$ as it is the supremum of a continuous function on a compact. By continuity of $f(\nu, \cdot)$ on a compact set we also have that, for every $\nu_1, \nu_2 \in [0, 1]$, there exists μ_1 and μ_2 achieving the maximum for $f(\nu_1, \cdot)$ and $f(\nu_2, \cdot)$ and

$$\begin{aligned} g(\nu_1) - g(\nu_2) &= f(\nu_1, \mu_1) - f(\nu_2, \mu_2) \\ &= f(\nu_1, \mu_1) - f(\nu_1, \mu_2) + f(\nu_1, \mu_2) - f(\nu_2, \mu_2) \leq f(\nu_1, \mu_2) - f(\nu_2, \mu_2) \leq M|\nu_1 - \nu_2|. \end{aligned}$$

Which implies that g is M -Lipschitz on $[0, 1]$ and thus continuous.

Hence L is continuous. Moreover, it satisfies $L(0) > 0$ and $L(1) \leq 0$ so, by the intermediate value theorem, we conclude that there exists $\nu^* \in (0, 1]$ such that $L(\nu^*) = 0$.

We now show that $\tilde{\pi}_{\nu^*}$ strictly improves on π^λ . Indeed, we have

$$\sup_{\mu \in [0, 1]} \mathcal{R}_n(\tilde{\pi}_{\nu^*}, \mathcal{B}(\mu)) \stackrel{(a)}{=} \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\tilde{\pi}_{\nu^*}, \mathcal{B}(\mu)) < \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)),$$

where (a) holds because $L(\nu^*) = 0$ and (b) follows from (B-9) and from the fact that $\nu^* > 0$. This shows that, π^λ is suboptimal.

Suppose that

$$\sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) < \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)).$$

In this case, the same reasoning, but by increasing the weight on the n^{th} order statistic would lead to a strict improvement. Therefore, if an optimal policy π^λ exists for problem (11) it must satisfy,

$$\sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) = \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)).$$

□

Proof of Proposition 4. It follows from (12) and (13) that there exists a $k \in \{2, \dots, n\}$ such that

$$\begin{aligned} \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_{k-1}}, \mathcal{B}(\mu)) &\leq \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_{k-1}}, \mathcal{B}(\mu)) \\ \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_k}, \mathcal{B}(\mu)) &\geq \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_k}, \mathcal{B}(\mu)). \end{aligned}$$

Pick a k verifying these two relations. We now construct a policy $\pi^{k, \gamma}$ randomizing between $D_{k-1:n}$ and $D_{k:n}$ and which satisfies the necessary condition (14). Consider the family of policies $(\pi^{k, \lambda})_{\lambda \in [0, 1]}$ prescribing $D_{k:n}$ w.p λ and $D_{k-1:n}$ w.p $1 - \lambda$.

We consider the function L defined from $[0, 1]$ to \mathbb{R} as,

$$L : \lambda \mapsto \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{k, \lambda}, \mathcal{B}(\mu)) - \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{k, \lambda}, \mathcal{B}(\mu)),$$

and note that $L(0) \leq 0$ and $L(1) \geq 0$. Moreover, L is continuous on $[0, 1]$ (see proof of Proposition 3). Thus by the intermediate value theorem, $L(\gamma) = 0$ for some $\gamma \in [0, 1]$. We denote by $\pi^{k, \gamma}$ our candidate policy that prescribes the order statistic $D_{k:n}$ w.p γ and $D_{k-1:n}$ w.p $1 - \gamma$. We define $\mu^- \in \arg \max_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{k, \gamma}, \mathcal{B}(\mu))$ and $\mu^+ \in \arg \max_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{k, \gamma}, \mathcal{B}(\mu))$ which exists by continuity on a compact. By construction of μ^+ , μ^- and because $L(\gamma) = 0$, we conclude that,

$$\mathcal{R}_n(\pi^{k, \gamma}, \mathcal{B}(\mu^-)) = \mathcal{R}_n(\pi^{k, \gamma}, \mathcal{B}(\mu^+)) = \sup_{\mu \in [0, 1]} \mathcal{R}_n(\pi^{k, \gamma}, \mathcal{B}(\mu)).$$

□

Proof of Proposition 5. Consider the family of priors $(p_\delta)_{\delta \in [0,1]}$ supported on $\{\mu^-, \mu^+\}$ and such that for any $\delta \in [0, 1]$,

$$p_\delta(\mu) = \begin{cases} \delta & \text{if } \mu = \mu^+ \\ 1 - \delta & \text{if } \mu = \mu^-. \end{cases}$$

We now show that there exists δ such that $\pi^{k,\gamma}$ is optimal for the problem,

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p_\delta} [\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu))]. \quad (\text{B-10})$$

We first establish a sufficient condition for a policy π^λ to be optimal for problem (B-10).

Remark that policies in Π_n^{OS} observe samples prior to decision hence,

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p_\delta} [\mathcal{R}_n(\pi, \mathcal{B}(\mu))] = \sum_{j=0}^n \mathbb{P} \left(\sum_{i=1}^n D_i = j \right) \inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p_\delta} \left[\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) \mid \sum_{i=1}^n D_i = j \right],$$

where the equality holds because π^λ observes the historical samples and because the posterior distribution of p_δ only depends on the number of ones observed, i.e. $\sum_{i=1}^n D_i$ is a sufficient statistic. To ease notations, let S_j denote the event $\{\sum_{i=1}^n D_i = j\}$ for every $j \in \{0, \dots, n\}$. To solve the inner optimization problem, we first notice that for every $j \in \{0, \dots, n\}$ we have

$$\begin{aligned} \mathbb{E}_{\mu \sim p_\delta} [\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) | S_j] &= \mathbb{P}(\mu = \mu^+ | S_j) \frac{\mu^+ - (1-q) + (1 - \mu^+ - q) \mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu^+)} [\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)}[x] | S_j]}{(1 - \mu^+) (1 - q)} \\ &\quad + \mathbb{P}(\mu = \mu^- | S_j) \frac{(1 - \mu^- - q) \mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu^-)} [\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)}[x] | S_j]}{\mu^- q} \\ &\stackrel{(a)}{=} a_j \cdot \mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\frac{1}{2})} [\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)}[x] | S_j] + b_j, \end{aligned}$$

where (a) holds because $\mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} [\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)}[x] | S_j] = \mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu')} [\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)}[x] | S_j]$ for any $\mu, \mu' \in (0, 1)$. This follows from the fact that for Bernoulli distributions, the distribution of \mathbf{D}_1^n conditional on S_j is the same for every $\mu \in (0, 1)$. Moreover,

$$a_j := \mathbb{P}(\mu = \mu^+ | S_j) \frac{1 - q - \mu^+}{(1 - \mu^+)(1 - q)} + \mathbb{P}(\mu = \mu^- | S_j) \frac{1 - q - \mu^-}{\mu^- q}$$

and $b_j = \mathbb{P}(\mu = \mu^+ | S_j) \frac{\mu^+ - (1-q)}{(1 - \mu^+)(1 - q)}$.

Note that, by definition of mixture of order statistics policies, we must have that, $\pi^\lambda(\mathbf{D}_1^n)$ has a support included in $[0, 1]$ since $\mathbf{D}_1^n \in \{0, 1\}^n$. Hence, we obtain that for any $x_0 \in [0, 1]$ if a policy π^λ satisfies the following property,

$$\mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\frac{1}{2})} [\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)}[x] | S_j] = \begin{cases} 0 & \text{if } a_j > 0 \\ x_0 & \text{if } a_j = 0 \\ 1 & \text{if } a_j < 0, \end{cases}$$

then it is optimal for problem (B-10). We now aim at proving that there exists a prior such that

$\pi^{k,\gamma}$ satisfies this sufficient condition. The challenge is that the sufficient condition involves the sign of the coefficients $(a_j)_{j \in \{0, \dots, n\}}$ depending on μ^- , μ^+ and δ . We simplify this dependence with our next lemma by showing that for any choice of $\mu^- < 1 - q \leq \mu^+$, we can construct a prior such that the sequence $(a_j)_{j \in \{0, \dots, n\}}$ is decreasing and hits 0 exactly once. Formally, we show the following.

Lemma B-3. *For any $\mu^- \in [0, 1 - q)$, $\mu^+ \in [1 - q, 1)$, and for any $j_0 \in \{1, \dots, n\}$, there exists $\delta' \in [0, 1]$ such that under prior $p_{\delta'}$, the sequence $(a_j)_{j \in \{0, \dots, n\}}$ is strictly decreasing and $a_{j_0} = 0$.*

The proof is deferred to Appendix D. Lemma B-3 implies that for any $j_0 \in \{1, \dots, n\}$ and $x_0 \in [0, 1]$, any policy $\pi^\lambda \in \Pi_n^{OS}$ that satisfies

$$\mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\frac{1}{2})} \left[\mathbb{E}_{x \sim \pi^\lambda(\mathbf{D}_1^n)} [x] \mid S_j \right] = \begin{cases} 0 & \text{if } j \leq j_0 - 1 \\ x_0 & \text{if } j = j_0 \\ 1 & \text{if } j \geq j_0 + 1, \end{cases}$$

is optimal for problem (B-10). We finally prove that $\pi^{k,\gamma}$ satisfies this simplified sufficient condition.

Note that by construction, for any $j \in \{1, \dots, n\}$,

$$\mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\frac{1}{2})} \left[\mathbb{E}_{x \sim \pi^{k,\gamma}(\mathbf{D}_1^n)} [x] \mid S_j \right] = \lambda \mathbb{E} [D_{k:n} \mid S_j] + (1 - \lambda) \mathbb{E} [D_{k-1:n} \mid S_j]$$

which implies that,

$$\mathbb{E}_{\mathbf{D}_1^n \sim \mathcal{B}(\frac{1}{2})} \left[\mathbb{E}_{x \sim \pi^{k,\gamma}(\mathbf{D}_1^n)} [x] \mid S_j \right] = \begin{cases} 0 & \text{if } j \leq n - k \\ \lambda & \text{if } j = n - k + 1 \\ 1 & \text{if } j \geq n - k + 2. \end{cases} \quad (\text{B-11})$$

Therefore, Lemma B-3 applied with $j_0 = n - k + 1$ implies that there exists δ_k such that, $\pi^{k,\gamma}$ is optimal for problem (B-10). Setting $p^* = p_{\delta_k}$, we showed that,

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p^*} \left[\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) \right] = \mathbb{E}_{\mu \sim p^*} \left[\mathcal{R}_n(\pi^{k,\gamma}, \mathcal{B}(\mu)) \right].$$

□

Proof of Theorem 4. First assume that, (12) and (13) hold and consider $k \in \{2, \dots, n\}$, $\gamma \in [0, 1]$, $\mu^- \in [0, 1 - q]$ and $\mu^+ \in [1 - q, 1]$ as defined in Proposition 4. We have that $\pi^{k,\gamma}$ satisfies the necessary condition (15).

We now show that $\pi^{k,\gamma}$ is optimal for the problem

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0, 1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)).$$

First, we remark that,

$$\mathcal{R}_n(\pi^{k,\lambda}, \mathcal{B}(\mu^-)) = \sup_{\mu \in [0, 1]} \mathcal{R}_n(\pi^{k,\gamma}, \mathcal{B}(\mu)) \geq \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0, 1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu))$$

because $\pi^{k,\gamma} \in \Pi_n^{OS}$. To prove the lower bound, we note that for these choices of k, γ, μ^- and μ^+ , Proposition 5 ensures that there exists a prior p^* supported on $\{\mu^-, \mu^+\}$ such that,

$$\inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p^*} [\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu))] = \mathbb{E}_{\mu \sim p^*} [\mathcal{R}_n(\pi^{k,\gamma}, \mathcal{B}(\mu))]. \quad (\text{B-12})$$

Therefore,

$$\begin{aligned} \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) &= \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{p \in \Delta([0,1])} \mathbb{E}_{\mu \sim p} [\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu))] \\ &\stackrel{(c)}{\geq} \sup_{p \in \Delta([0,1])} \inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p} [\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu))] \\ &\geq \inf_{\pi^\lambda \in \Pi_n^{OS}} \mathbb{E}_{\mu \sim p^*} [\mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu))] \\ &\stackrel{(d)}{=} \mathbb{E}_{\mu \sim p^*} [\mathcal{R}_n(\pi^{k,\gamma}, \mathcal{B}(\mu))] \stackrel{(e)}{=} \mathcal{R}_n(\pi^{k,\gamma}, \mathcal{B}(\mu^-)), \end{aligned}$$

where (c) holds by weak duality, (d) is a consequence of (B-12) and (e) follows from (15). The lower bound matches our upper bound. Thus all inequalities are equalities and we have exhibited a saddle point for (18). This implies that,

$$\mathcal{R}_n^* \stackrel{(a)}{=} \inf_{\pi^\lambda \in \Pi_n^{OS}} \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi^\lambda, \mathcal{B}(\mu)) = \mathcal{R}_n(\pi^{k,\gamma}, \mathcal{B}(\mu^-)) = \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{k,\gamma}, F),$$

where (a) follows from Theorem 3.

In other words, $\pi^{k,\gamma}$ is an optimal minimax data-driven algorithm and its performance can be explicitly computed by evaluating it against a specific Bernoulli distribution. \square

Proof of Corollary 1. Fix $n \geq 1$. We note that it is sufficient to show that, for every $F \in \mathcal{F}$,

$$\mathcal{R}_n(\pi^{\text{cvx}(k,\gamma)}, F) \leq \mathcal{R}_n(\pi^{k,\gamma}, F). \quad (\text{B-13})$$

We then conclude the proof by remarking that,

$$\mathcal{R}_n^* \leq \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{\text{cvx}(k,\gamma)}, F) \stackrel{(a)}{\leq} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi^{k,\gamma}, F) \stackrel{(b)}{=} \mathcal{R}_n^*,$$

where (a) follows from (B-13) and (b) is a consequence of Theorem 4.

We now prove (B-13). Fix $F \in \mathcal{F}$. We have that,

$$\begin{aligned} \mathcal{C}(\pi^{\text{cvx}(k,\gamma)}, F, n) &= \mathbb{E}_{\mathbf{D}_1^n \sim F} \left[\mathbb{E}_{x \sim \pi^{\text{cvx}(k,\gamma)}(\mathbf{D}_1^n)} [c_F(x)] \right] \\ &= \mathbb{E}_{\mathbf{D}_1^n \sim F} [c_F((1-\gamma)D_{k-1:n} + \gamma D_{k:n})] \\ &\stackrel{(a)}{\leq} \mathbb{E}_{\mathbf{D}_1^n \sim F} [(1-\gamma)c_F(D_{k-1:n}) + \gamma c_F(D_{k:n})] = \mathcal{C}(\pi^{k,\gamma}, F, n), \end{aligned}$$

where (a) holds because $x \mapsto c_F(x)$ is the expectation of a family of convex functions and is thus convex. \square

C Proofs for Section 5

Proof of Theorem 5. The proof of this theorem goes as follows. We first establish, in Lemma C-1, a characterization of the asymptotic behavior of single order statistic policy sequences for different regimes. As a corollary, we derive the asymptotic approximation of the worst-case performance of SAA.

Finally, we leverage the characterization of the optimal policy derived in Theorem 4 to reduce the understanding of the optimal performance to a problem involving mixture of order statistics. We finally, conclude by applying again Lemma C-1.

Step 1: We characterize the performance of a sequence of single order statistic policies. We first remark that a sequence of single order statistic policies can be characterized by a sequence $\mathbf{r} := (r_n)_{n \geq 1}$ where for each $n \geq 1$, $r_n \in \{1, \dots, n\}$. We denote by $\pi^{\mathbf{r}}$, the sequence of policies such that for any $n \geq 1$, $\pi_n^{\mathbf{r}} = \pi^{OS_{r_n}}$.

The next result characterizes the asymptotic behavior of the worst-case performance of $\pi^{\mathbf{r}}$.

Lemma C-1. *i) If \mathbf{r} is such that, $\lim_{n \rightarrow \infty} \frac{|r_n - qn|}{\sqrt{n}} = \ell < \infty$, then*

$$\lim_{n \rightarrow \infty} \sqrt{n} \cdot \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\mathbf{r}}, F) = \max \left[\max_{\delta \geq 0} H^+(\delta, \ell), \max_{\delta \geq 0} H^-(\delta, \ell) \right].$$

where, $H^+(\delta, \ell) := \frac{\delta}{q(1-q)} \left(1 - \Phi \left(\frac{\delta - \ell}{\sqrt{q(1-q)}} \right) \right)$, $H^-(\delta, \ell) := \frac{\delta}{q(1-q)} \left(1 - \Phi \left(\frac{\delta + \ell}{\sqrt{q(1-q)}} \right) \right)$ and Φ is the cdf of the standard gaussian distribution.

Furthermore,

- If the sequence μ_n is such that $\lim_{n \rightarrow \infty} \sqrt{n} (1 - q - \mu_n) = \delta > 0$, then

$$\lim_{n \rightarrow \infty} \sqrt{n} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n)) = H^+(\delta, \ell).$$

- If the sequence μ_n is such that $\lim_{n \rightarrow \infty} \sqrt{n} (\mu_n - (1 - q)) = \delta > 0$, then

$$\lim_{n \rightarrow \infty} \sqrt{n} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n)) = H^-(\delta, \ell).$$

ii) If \mathbf{r} is such that, $\lim_{n \rightarrow \infty} \frac{|r_n - qn|}{\sqrt{n}} = \infty$ then,

$$\lim_{n \rightarrow \infty} \sqrt{n} \cdot \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\mathbf{r}}, F) = \infty$$

and there exists a sequence of elements $\{\mu_n\}_{n \geq 1}$ in $[0, 1]$ such that

$$\lim_{n \rightarrow \infty} \sqrt{n} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n)) = \infty.$$

The proof is presented in Appendix D.

Lemma C-1 establishes that there are two notable regimes driving the asymptotic worst-case performance. In the first regime where the sequence of order statistics is asymptotically “close”

to $\lceil qn \rceil$, in the sense that $r_n = qn + \mathcal{O}(\sqrt{n})$, the worst-case relative regret decreases at a rate of $\Theta(1/\sqrt{n})$. We also establish a closed form expression of the exact limiting constant associated to the rate of convergence and characterize the family of near worst-case distributions, namely Bernoulli distributions whose means go to $1 - q$ at a rate $\Theta(1/\sqrt{n})$.

In the second regime for which the sequence of order statistics is asymptotically “far” from $\lceil qn \rceil$, we show that the worst-case relative regret decreases at a slower rate as it converges at a rate of $\omega(1/\sqrt{n})$. This naturally implies that this family of policy sequences is necessarily suboptimal and strictly dominated by sequences of order statistics asymptotically “close” to $\lceil qn \rceil$.

Step 2: We now characterize the asymptotic performance of SAA. Let $\mathbf{r}^{\text{SAA}} = (\lceil qn \rceil)_{n \in \mathbb{N}}$ and recall that for every $n \in \mathbb{N}$, we have $\pi_n^{\text{SAA}} = \pi_n^{\mathbf{r}^{\text{SAA}}}$. Note that $\lim_{n \rightarrow \infty} \frac{|r_n^{\text{SAA}} - qn|}{\sqrt{n}} = 0$. Therefore, by (i) in Lemma C-1 we conclude that,

$$\lim_{n \rightarrow \infty} \sqrt{n} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\text{SAA}}, F) = \max \left[\max_{\delta \geq 0} H^+(\delta, 0), \max_{\delta \geq 0} H^-(\delta, 0) \right] = C^*. \quad (\text{C-1})$$

Step 3: We finally derive an asymptotic approximation of the optimal performance. We use the characterization of the optimal policy derived in Theorem 4 and denote by $(k_n)_{n \geq 1}$ and $(\gamma_n)_{n \geq 1}$ the sequences of parameters that describe the optimal policy when facing n samples. For any $n \in \mathbb{N}$, we have $\pi_n^{\mathbf{k}, \gamma} = \pi_n^{k_n, \gamma_n}$. For every $n \in \mathbb{N}$, and every $\mu_n \in [0, 1]$ remark that,

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\text{SAA}}, F) \geq \mathcal{R}_n^* \stackrel{(a)}{=} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\mathbf{k}, \gamma}, F) \stackrel{(b)}{\geq} \gamma_n \mathcal{R}_n(\pi^{k_n}, \mathcal{B}(\mu_n)) + (1 - \gamma_n) \mathcal{R}_n(\pi^{k_n - 1}, \mathcal{B}(\mu_n)), \quad (\text{C-2})$$

where (a) holds by Theorem 4 and (b) is by definition of $\pi_n^{\mathbf{k}, \gamma}$.

Remark that, (C-1) together with the first inequality of (C-2) imply that,

$$\limsup_{n \rightarrow \infty} \sqrt{n} \mathcal{R}_n^* \leq C^*. \quad (\text{C-3})$$

We now compute a lower bound on the limit of $\sqrt{n} \mathcal{R}_n^*$ that matches the upper bound derived in (C-3). We only need to show that $\liminf_{n \rightarrow \infty} \sqrt{n} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\mathbf{k}, \gamma}, F) \geq C^*$. Consider an increasing function ψ such that $\sqrt{\psi(n)} \sup_{F \in \mathcal{F}} \mathcal{R}_{\psi(n)}(\pi_{\psi(n)}^{\mathbf{k}, \gamma}, F)$ converges. By inequality (b) in (C-2), one only need to show that there exists a sequence $(\mu_n)_{n \in \mathbb{N}}$ such that,

$$\limsup_{n \rightarrow \infty} \sqrt{\psi(n)} (\gamma_{\psi(n)} \cdot \mathcal{R}_{\psi(n)}(\pi^{k_{\psi(n)}}, \mathcal{B}(\mu_{\psi(n)})) + (1 - \gamma_{\psi(n)}) \cdot \mathcal{R}_{\psi(n)}(\pi^{k_{\psi(n)} - 1}, \mathcal{B}(\mu_{\psi(n)}))) \geq C^*.$$

We prove that this lower bound holds by considering different scenarios for the sequence \mathbf{k} . Consider an increasing function $\tilde{\psi}$ such that $\frac{|k_{\tilde{\psi}(\psi(n))} - q\tilde{\psi}(\psi(n))|}{\sqrt{\tilde{\psi}(\psi(n))}}$ converges to a limit ℓ in $\mathbb{R} \cup \{\infty\}$. To ease notations, we let $f := \tilde{\psi} \circ \psi$ and $\mathbf{k}_f := (k_{f(n)})_{n \in \mathbb{N}}$.

Case 1: $\ell = \infty$. Note that,

$$\lim_{n \rightarrow \infty} \frac{|k_{f(n)} - qf(n)|}{\sqrt{f(n)}} = \lim_{n \rightarrow \infty} \frac{|k_{f(n)} - 1 - qf(n)|}{\sqrt{f(n)}} = \infty.$$

Hence, by (ii) in Lemma C-1 we conclude that there exists a sequence μ_n such that

$$\lim_{n \rightarrow \infty} \sqrt{f(n)} \mathcal{R}_n (\pi^{k_{f(n)}}, \mathcal{B}(\mu_{f(n)})) = \lim_{n \rightarrow \infty} \sqrt{f(n)} \mathcal{R}_{f(n)} (\pi^{k_{f(n)}-1}, \mathcal{B}(\mu_{f(n)})) = \infty.$$

and so,

$$\lim_{n \rightarrow \infty} \sqrt{f(n)} (\gamma_{f(n)} \cdot \mathcal{R}_{f(n)} (\pi^{k_{f(n)}}, \mathcal{B}(\mu_{f(n)})) + (1 - \gamma_{f(n)}) \cdot \mathcal{R}_{f(n)} (\pi^{k_{f(n)}-1}, \mathcal{B}(\mu_{f(n)}))) = \infty.$$

Case 2: $\ell < \infty$. In this case, (i) in Lemma C-1 establishes the asymptotic behavior of the worst-case expected relative regret. We remark that the limit depends only on ℓ . Therefore,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sqrt{f(n)} \mathcal{R}_n (\pi^{k_{f(n)}}, \mathcal{B}(\mu_{f(n)})) &= \max \left[\max_{\delta \geq 0} H^+(\delta, \ell), \max_{\delta \geq 0} H^-(\delta, \ell) \right] \\ \lim_{n \rightarrow \infty} \sqrt{f(n)} \mathcal{R}_{f(n)} (\pi^{k_{f(n)}-1}, \mathcal{B}(\mu_{f(n)})) &= \max \left[\max_{\delta \geq 0} H^+(\delta, \ell), \max_{\delta \geq 0} H^-(\delta, \ell) \right]. \end{aligned}$$

Let $\delta^+ \in \arg \max_{\delta \geq 0} H^+(\delta, \ell)$ and $\delta^- \in \arg \max_{\delta \geq 0} H^-(\delta, \ell)$. Assume that $H^+(\delta^+, \ell) \geq H^-(\delta^-, \ell)$ (the other case is proved by a similar argument) and consider the sequence $(\mu_n)_{n \in \mathbb{N}}$ defined as, $\mu_n = 1 - q - \frac{\delta^+}{\sqrt{n}}$ for every $n \geq 1$. By Lemma C-1 we conclude that,

$$\lim_{n \rightarrow \infty} \sqrt{f(n)} \mathcal{R}_n (\pi^{k_{f(n)}}, \mathcal{B}(\mu_{f(n)})) = \lim_{n \rightarrow \infty} \sqrt{f(n)} \mathcal{R}_n (\pi^{k_{f(n)}-1}, \mathcal{B}(\mu_{f(n)})) = H^+(\delta^+, \ell).$$

Hence,

$$\lim_{n \rightarrow \infty} \sqrt{f(n)} (\gamma_{f(n)} \cdot \mathcal{R}_{f(n)} (\pi^{k_{f(n)}}, \mathcal{B}(\mu_{f(n)})) + (1 - \gamma_{f(n)}) \cdot \mathcal{R}_{f(n)} (\pi^{k_{f(n)}-1}, \mathcal{B}(\mu_{f(n)}))) = H^+(\delta^+, \ell).$$

To conclude the proof, we need to show that $H^+(\delta^+, \ell) \geq C^*$. This is a straightforward consequence of the definition of δ^+ together with the following lemma.

Lemma C-2. For any $\ell \in \mathbb{R}$,

$$\max \left[\max_{\delta \geq 0} H^+(\delta, \ell), \max_{\delta \geq 0} H^-(\delta, \ell) \right] \geq C^*.$$

The proof is deferred to Appendix D.

We hence conclude that

$$\limsup_{n \rightarrow \infty} \sqrt{\psi(n)} (\gamma_{\psi(n)} \cdot \mathcal{R}_{\psi(n)} (\pi^{k_{\psi(n)}}, \mathcal{B}(\mu_{\psi(n)})) + (1 - \gamma_{\psi(n)}) \cdot \mathcal{R}_{\psi(n)} (\pi^{k_{\psi(n)}-1}, \mathcal{B}(\mu_{\psi(n)}))) \geq C^*.$$

Therefore we showed that,

$$C^* = \lim_{n \rightarrow \infty} \sqrt{n} \sup_{F \in \mathcal{F}} \mathcal{R}_n (\pi_n^{\text{SAA}}, F) \geq \limsup_{n \rightarrow \infty} \mathcal{R}_n^* \geq \liminf_{n \rightarrow \infty} \mathcal{R}_n^* \geq C^*,$$

which concludes the proof. □

D Proofs of Auxiliary Results

Proof of Lemma A-1. Fix $F \in \mathcal{F}$. We next analyze $c_F(x)$ by decomposing the expected cost. We have

$$\begin{aligned}
c_F(x) &= \mathbb{E}_{D \sim F} \left[b(D - x)^+ + h(x - D)^+ \right] \\
&= h \int_{[0,x]} (x - y) dF(y) + b \int_{(x,\infty)} (y - x) dF(y) \\
&= (b + h) \int_{[0,x]} (x - y) dF(y) + b \int_{[0,\infty)} (y - x) dF(y) \\
&= (b + h) \left(xF(x) - \int_{[0,x]} y dF(y) \right) + b(\mathbb{E}_F[D] - x) \\
&\stackrel{(a)}{=} b(\mathbb{E}_F[D] - x) + (b + h) \int_{[0,x]} F(y) dy,
\end{aligned}$$

where equation (a) is a consequence of Riemann-Stieltjes integration by part. \square

Proof of Lemma A-2. For every $\alpha \in [0, 1]$, let $F_\alpha := \mathcal{B}(1 - \alpha)$. For every $r \in \{1, \dots, n\}$ we have,

$$\begin{aligned}
\mathcal{C}(\pi^{OS_r}, F_\alpha, n) &= \mathbb{E}_{\mathbf{D}_1^n \sim F_\alpha} \left[\mathbb{E}_{x \sim \pi^{OS_r}(\mathbf{D}_1^n)} [c_{F_\alpha}(x)] \right] \\
&= \mathbb{E}_{\mathbf{D}_1^n \sim F_\alpha} [c_{F_\alpha}(D_{r:n})] \\
&= \mathbb{E}_{\mathbf{D}_1^n \sim F_\alpha} [\alpha h D_{r:n} + (1 - \alpha)b(1 - D_{r:n})] \\
&= (b + h) \left[(\alpha - q) \mathbb{E}_{\mathbf{D}_1^n \sim F_\alpha} [D_{r:n}] + q(1 - \alpha) \right].
\end{aligned}$$

When $\alpha \leq q$, we observe that $x_{F_\alpha}^* = 1$ and $\text{opt}(F_\alpha) = \alpha h$. Therefore,

$$\mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(1 - \alpha)) = \frac{(q - \alpha) \left(1 - \mathbb{E}_{\mathbf{D}_1^n \sim F_\alpha} [D_{r:n}] \right)}{(1 - q)\alpha}.$$

We have

$$\mathbb{E}_{\mathbf{D}_1^n \sim F_\alpha} [D_{r:n}] = \mathbb{P}_{\mathbf{D}_1^n \sim F_\alpha} (D_{r:n} = 1) = 1 - \mathbb{P}_{\mathbf{D}_1^n \sim F_\alpha} (D_{r:n} = 0) = 1 - B_{r,n}(\alpha),$$

where the last equality follows from the definition of the Bernstein polynomial. In turn, we conclude that for $\alpha \in [0, q)$,

$$\mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(1 - \alpha)) = \frac{(q - \alpha) B_{r,n}(\alpha)}{(1 - q)\alpha}. \tag{D-1}$$

When $\alpha \in [q, 1]$, we observe that $x_{F_\alpha}^* = 0$, $\text{opt}(F_\alpha) = (1 - \alpha)b$ and we establish similarly that,

$$\mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(1 - \alpha)) = \frac{(\alpha - q) (1 - B_{r,n}(\alpha))}{q(1 - \alpha)}. \tag{D-2}$$

We conclude the proof by remarking that (D-1) and (D-2) imply that for every $\alpha \in [0, 1]$,

$$\mathcal{R}_n \left(\pi^{OSr}, \mathcal{B}(1 - \alpha) \right) = \frac{(1 - B_{r,n}(\alpha))(\alpha - q) + q(1 - \alpha)}{\min \{(1 - q)\alpha, q(1 - \alpha)\}} - 1.$$

□

Proof of Lemma B-1. Consider a policy $\pi \in \Pi_n$ such that for some $\hat{\mathbf{D}}_1^n \in \{0, 1\}^n$, the support of $\pi \left(\hat{\mathbf{D}}_1^n \right)$ is not a subset of $[\hat{\mathbf{D}}_{1:n}, \hat{\mathbf{D}}_{n:n}]$. We now construct a policy $\pi' \in \Pi_n$ such that the support of $\pi' \left(\hat{\mathbf{D}}_1^n \right)$ is a subset of $[\hat{\mathbf{D}}_{1:n}, \hat{\mathbf{D}}_{n:n}]$ and which ensures a cost at least as good as the one incurred by π against any Bernoulli distribution.

Assume first that $\hat{D}_{1:n} = 0$ and $\hat{D}_{n:n} = 1$.

Recall that $G_{\hat{\mathbf{D}}_1^n}^\pi$ is the cdf of the distribution $\pi \left(\hat{\mathbf{D}}_1^n \right)$. We define π' such that for all $\mathbf{D}_1^n \in \{0, 1\}^n$, if $\mathbf{D}_1^n \neq \hat{\mathbf{D}}_1^n$, we have $\pi' \left(\mathbf{D}_1^n \right) = \pi \left(\mathbf{D}_1^n \right)$ and we construct the cdf of $\pi' \left(\hat{\mathbf{D}}_1^n \right)$ in order to ensure that the support is $[0, 1]$ as follows.

$$G_{\hat{\mathbf{D}}_1^n}^{\pi'}(y) = \begin{cases} 0 & \text{if } y < 0 \\ G_{\hat{\mathbf{D}}_1^n}^\pi(y) & \text{if } y \in [0, 1) \\ 1 & \text{if } y \geq 1. \end{cases}$$

For any $\mu \in [0, 1]$, let F_μ be the cdf of the Bernoulli distribution $\mathcal{B}(\mu)$. We have for any $x < 0$,

$$c_{F_\mu}(x) = \mu \cdot b \cdot (1 - x)^+ + (1 - \mu) \cdot b \cdot (-x)^+ > \mu \cdot b = c_{F_\mu}(0).$$

Similarly, one can show that for any $x > 1$, $c_{F_\mu}(x) > c_{F_\mu}(1)$.

Therefore, for every $\mu \in [0, 1]$, the difference in costs between π and π' satisfies,

$$\begin{aligned} \mathcal{C}(\pi, \mathcal{B}(\mu), n) - \mathcal{C}(\pi', \mathcal{B}(\mu), n) &\stackrel{(a)}{=} \mathbb{P}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left(\mathbf{D}_1^n = \hat{\mathbf{D}}_1^n \right) \left(\mathbb{E}_{x \sim \pi(\hat{\mathbf{D}}_1^n)} [c_{F_\mu}(x)] - \mathbb{E}_{x \sim \pi'(\hat{\mathbf{D}}_1^n)} [c_{F_\mu}(x)] \right) \\ &= \mathbb{P}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left(\mathbf{D}_1^n = \hat{\mathbf{D}}_1^n \right) \left(\int_{\mathbb{R}} c_{F_\mu}(y) dG_{\hat{\mathbf{D}}_1^n}^\pi(y) - \int_{[0,1]} c_{F_\mu}(y) dG_{\hat{\mathbf{D}}_1^n}^{\pi'}(y) \right) \\ &\stackrel{(b)}{\geq} \mathbb{P}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left(\mathbf{D}_1^n = \hat{\mathbf{D}}_1^n \right) \left(\int_{(-\infty;0)} c_{F_\mu}(0) dG_{\hat{\mathbf{D}}_1^n}^\pi(y) \right. \\ &\quad \left. + \int_{[0,1]} c_{F_\mu}(y) dG_{\hat{\mathbf{D}}_1^n}^\pi(y) + \int_{[1,\infty)} c_{F_\mu}(1) dG_{\hat{\mathbf{D}}_1^n}^\pi(y) - \int_{[0,1]} c_{F_\mu}(y) dG_{\hat{\mathbf{D}}_1^n}^{\pi'}(y) \right) \\ &\stackrel{(c)}{=} 0, \end{aligned}$$

where (a) holds because for all $\mathbf{D}_1^n \in \{0, 1\}^n$, such that $\mathbf{D}_1^n \neq \hat{\mathbf{D}}_1^n$, we have $\pi' \left(\mathbf{D}_1^n \right) = \pi \left(\mathbf{D}_1^n \right)$, (b) follows from the fact that $c_{F_\mu}(x) < c_{F_\mu}(0)$ for $x < 0$, and $c_{F_\mu}(x) > c_{F_\mu}(1)$ for $x > 1$. (c) is a consequence of the constructions of $G_{\hat{\mathbf{D}}_1^n}^{\pi'}$. Hence, this shows that we weakly improve the cost of policy π with the policy π' .

We now consider the case where $\hat{D}_{1:n} = \hat{D}_{n:n} = 0$. In that case, we have that $\mathcal{C}(\pi, \mathcal{B}(0), n) > 0$. Remarking that $\text{opt}(\mathcal{B}(0)) = 0$, we conclude that $\mathcal{R}_n(\pi, \mathcal{B}(0)) = \infty$, which shows the strict suboptimality of π .

A similar reasoning holds for $\hat{D}_{1:n} = \hat{D}_{n:n} = 1$.

We conclude the proof by repeating this process for every value of $\hat{\mathbf{D}}_1^n$. \square

Proof of Lemma B-2. It is clear that,

$$\inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma_{\mathbf{e}}}, \mathcal{B}(\mu) \right) \leq \inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1 \\ (e_i) \text{ non-decreasing}}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma_{\mathbf{e}}}, \mathcal{B}(\mu) \right).$$

We now show that,

$$\inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma_{\mathbf{e}}}, \mathcal{B}(\mu) \right) \geq \inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1 \\ (e_i) \text{ non-decreasing}}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma_{\mathbf{e}}}, \mathcal{B}(\mu) \right).$$

Consider $\mathbf{e} = (e_i)_{i \in \{0, \dots, n+1\}}$ and assume that there exists $j \in \{0, \dots, n-1\}$ such that $e_j > e_{j+1}$. We consider the sequence $\mathbf{f} := (f_i)_{i \in \{0, \dots, n+1\}}$ such that,

$$f_i = \begin{cases} e_i & \text{if } i \in \{0, \dots, n\} \setminus \{j, j+1\} \\ \frac{1}{\frac{q}{n-j} + \frac{1-q}{j}} \left(\frac{q}{n-j} e_j + \frac{1-q}{j} e_{j+1} \right) & \text{if } i \in \{j, j+1\}. \end{cases}$$

We show that the cost of $\pi^{\Sigma_{\mathbf{f}}}$ is weakly lower than the one of $\pi^{\Sigma_{\mathbf{e}}}$. We have, for any $\mu \in [0, 1]$,

$$\begin{aligned} \mathcal{C} \left(\pi^{\Sigma_{\mathbf{e}}}, \mathcal{B}(\mu), n \right) - \mathcal{C} \left(\pi^{\Sigma_{\mathbf{f}}}, \mathcal{B}(\mu), n \right) &\stackrel{(a)}{=} \mathbb{P}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left(\sum_{i=1}^n D_i = j \right) (c_{F_\mu}(e_j) - c_{F_\mu}(f_j)) \\ &\quad + \mathbb{P}_{\mathbf{D}_1^n \sim \mathcal{B}(\mu)} \left(\sum_{i=1}^n D_i = j+1 \right) (c_{F_\mu}(e_{j+1}) - c_{F_\mu}(f_{j+1})) \\ &= C_\mu \cdot (1 - \mu - q) \cdot \left(\frac{1-\mu}{n-j} (e_j - f_j) + \frac{\mu}{j+1} (e_{j+1} - f_{j+1}) \right), \end{aligned}$$

where $C_\mu = (b+h) \cdot \mu^j \cdot (1-\mu)^{n-j-1} \frac{n!}{j!(n-j-1)!}$ and (a) holds because $e_i = f_i$ for any i different from j and $j+1$. Note that $C_\mu \geq 0$ for all $\mu \in [0, 1]$. Letting $\mathcal{L}(\mu) = \frac{1-\mu}{n-j} (e_j - f_j) + \frac{\mu}{j+1} (e_{j+1} - f_{j+1})$ for all $\mu \in [0, 1]$, we only need to show that

$$\begin{aligned} \mathcal{L}(\mu) &\geq 0 && \text{for } \mu \in [0, 1-q] \\ \mathcal{L}(\mu) &\leq 0 && \text{for } \mu \in [1-q, 1]. \end{aligned}$$

Note that \mathcal{L} is a linear function of μ , $\mathcal{L}(0) = \frac{e_j - f_j}{n-j} \geq 0$ and $\mathcal{L}(1) = \frac{e_{j+1} - f_{j+1}}{j} \leq 0$, therefore, one only need to check $\mathcal{L}(1-q) = 0$. The latter equality holds by definition of f_j and f_{j+1} .

We hence conclude that for any $\mu \in [0, 1]$

$$\mathcal{C} \left(\pi^{\Sigma_{\mathbf{e}}}, \mathcal{B}(\mu), n \right) - \mathcal{C} \left(\pi^{\Sigma_{\mathbf{f}}}, \mathcal{B}(\mu), n \right) \geq 0.$$

By iterating this process for any i such that $e_i > e_{i+1}$, we conclude that

$$\inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu) \right) \geq \inf_{\substack{\mathbf{e} \in [0,1]^{n+1} \\ e_0=0, e_n=1 \\ (e_i) \text{ non-decreasing}}} \sup_{\mu \in [0,1]} \mathcal{R}_n \left(\pi^{\Sigma \mathbf{e}}, \mathcal{B}(\mu) \right).$$

This completes the proof. \square

Proof of Lemma B-3. To ease notations, let $p_j^+(\delta) := \mathbb{P}(\mu = \mu^+ \mid \sum_{i=1}^n D_i = j)$ for $j \in \{1, \dots, n\}$, and remark that,

$$a_j = p_j^+(\delta) \frac{1 - q - \mu^+}{(1 - \mu^+)(1 - q)} + (1 - p_j^+(\delta)) \frac{1 - q - \mu^-}{\mu^- q}.$$

Step 1: We first show that the sequence $(a_j)_{j \in \{0, \dots, n\}}$ is decreasing. By assumption, $0 < \mu^- < 1 - q < \mu^+ < 1$, therefore, $\frac{1 - q - \mu^-}{\mu^- q} > 0$ and $\frac{1 - q - \mu^+}{(1 - \mu^+)(1 - q)} < 0$. Hence, to show that $(a_j)_{j \in \{0, \dots, n\}}$ is decreasing, it is sufficient to show that $(p_j^+(\delta))_{j \in \{0, \dots, n\}}$ is increasing. For every $j \in \{0, \dots, n\}$ we have,

$$\begin{aligned} p_j^+(\delta) &= \frac{\mathbb{P}(\sum_{i=1}^n D_i = j \mid \mu = \mu^+) \mathbb{P}(\mu = \mu^+)}{\mathbb{P}(\sum_{i=1}^n D_i = j \mid \mu = \mu^-) \mathbb{P}(\mu = \mu^-) + \mathbb{P}(\sum_{i=1}^n D_i = j \mid \mu = \mu^+) \mathbb{P}(\mu = \mu^+)} \\ &= \frac{\delta \binom{n}{j} (\mu^+)^j (1 - \mu^+)^{n-j}}{(1 - \delta) \binom{n}{j} (1 - \mu^-)^{n-j} (\mu^-)^j + \delta \binom{n}{j} (1 - \mu^+)^{n-j} (\mu^+)^j}. \end{aligned}$$

Therefore, for $j \in \{0, \dots, n-1\}$, we have

$$p_{j+1}^+(\delta) - p_j^+(\delta) = \delta (\mu^+)^j (1 - \mu^+)^{n-j-1} \frac{\mu^+ d_j - (1 - \mu^+) d_{j+1}}{d_j d_{j+1}},$$

where $d_j := \delta (1 - \mu^+)^{n-j} (\mu^+)^j + (1 - \delta) (1 - \mu^-)^{n-j} (\mu^-)^j \geq 0$. Furthermore,

$$\begin{aligned} \mu^+ d_j - (1 - \mu^+) d_{j+1} &= (1 - \delta) \mu^+ (1 - \mu^-)^{n-j} (\mu^-)^j - (1 - \delta) (1 - \mu^+) (1 - \mu^-)^{n-j-1} (\mu^-)^{j+1} \\ &= (1 - \delta) (1 - \mu^-)^{n-j-1} (\mu^-)^j (\mu^+ - \mu^-) > 0. \end{aligned}$$

Step 2: Let $j \in \{0, \dots, n\}$ and remark that $p_j^+(0) = 0$ whereas, $p_j^+(1) = 1$. Hence by making explicit the dependency of $a_j(\delta)$ in δ , we have that, $a_j(0) > 0$, $a_j(1) < 0$ and $\delta \mapsto a_j(\delta)$ is continuous. Hence, by the intermediate value theorem, there exists $\delta' \in [0, 1]$ such that $a_j(\delta') = 0$. \square

Proof of Lemma C-1. To prove this lemma we characterize two regimes for the sequence of order statistic policies $\pi^{\mathbf{r}}$. These two regimes depends on the value of the following limit

$$\lim_{n \rightarrow \infty} \frac{|r_n - qn|}{\sqrt{n}} := \ell.$$

When $\ell < \infty$, we show that the worst-case expected regret scales at a rate $\Theta\left(\frac{1}{\sqrt{n}}\right)$ and we derive an exact closed form characterization of the limiting constant, along with a family of candidate hard cases that nature may select.

We also show that for $\ell = \infty$, the worst-case expected regret scales at a rate $\omega\left(\frac{1}{\sqrt{n}}\right)$, which naturally implies sub-optimality of this class of order statistic policies.

First remark that Theorem 1 implies that,

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\mathbf{r}}, F) = \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu)).$$

Furthermore, we derive from Lemma A-2 a closed form expression of the relative-regret of an order statistic policy against a Bernoulli distribution. Namely, we have that

$$\mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu)) = \frac{\mu - (1-q)}{(1-q)(1-\mu)} \mathbb{P}(D_{r_n:n} = 0) \quad \text{if } \mu \geq 1-q \quad (\text{D-3})$$

$$\mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu)) = \frac{1-q-\mu}{q\mu} \mathbb{P}(D_{r_n:n} = 1) \quad \text{if } \mu \leq 1-q \quad (\text{D-4})$$

Step 1: $\ell < \infty$. We show that in this case,

$$\lim_{n \rightarrow \infty} \sqrt{n} \cdot \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\mathbf{r}}, F) = \max \left[\max_{\delta \geq 0} H^+(\delta, \ell), \max_{\delta \geq 0} H^-(\delta, \ell) \right],$$

where, $H^+(\delta, \ell) := \frac{\delta}{q(1-q)} \left(1 - \Phi \left(\frac{\delta + \ell}{\sqrt{q(1-q)}} \right) \right)$ and $H^-(\delta, \ell) := \frac{\delta}{q(1-q)} \left(1 - \Phi \left(\frac{\delta - \ell}{\sqrt{q(1-q)}} \right) \right)$.

It follows from Theorem 1 that it is sufficient to show that

$$\lim_{n \rightarrow \infty} \sqrt{n} \cdot \sup_{\mu \in [0,1]} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu)) = \max \left[\max_{\delta \geq 0} H^+(\delta, \ell), \max_{\delta \geq 0} H^-(\delta, \ell) \right]. \quad (\text{D-5})$$

We analyze the worst-case expected regret incurred by a sequence policy of order statistics policies against different regimes of Bernoulli distribution with means $(\mu_n)_{n \in \mathbb{N}}$. These regimes are defined by the limit of the following sequence,

$$\lim_{n \rightarrow \infty} \sqrt{n}(\mu_n - (1-q)) = \delta,$$

where $\delta \in \mathbb{R} \cup \{-\infty, \infty\}$. We show that if $\delta \in \{0, -\infty, \infty\}$, meaning that μ_n does not converge to $1-q$ at a rate of $\frac{1}{\sqrt{n}}$, the asymptotic performance decreases at a faster rate than $\frac{1}{\sqrt{n}}$.

Case (a): $\delta = 0$.

In this cases, we have that $\mu_n \rightarrow 1-q$ as $n \rightarrow \infty$. Let $\mathcal{N}_+ := \{n \mid \mu_n \geq 1-q\}$ and $\mathcal{N}_- := \{n \mid \mu_n < 1-q\}$ and consider the subsequences $(\mu'_n)_{n \in \mathcal{N}_+}$ and $(\mu''_n)_{n \in \mathcal{N}_-}$. Since $\mathbb{N} = \mathcal{N}_+ \cup \mathcal{N}_-$, one of these two sets must be infinite. If one of them is finite, one only need to compute the limit for the second subsequence. When both are infinite, we establish the existence of the limit of $\sqrt{n}\mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n))$ and compute its value, by deriving the limit of the relative regret against both subsequence $(\mu'_n)_{n \in \mathcal{N}_+}$ and $(\mu''_n)_{n \in \mathcal{N}_-}$ and prove that the limit coincides.

In particular, remark that (D-3) implies that, for every $n \in \mathcal{N}_+$,

$$\sqrt{n}\mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu'_n)) \leq \sqrt{n} \frac{\mu'_n - (1-q)}{(1-q)(1-\mu'_n)} \xrightarrow{(a)} 0 \quad \text{as } n \rightarrow \infty,$$

where (a) holds because the numerator goes to 0 as $\delta = 0$ and the denominator converges to a constant. Similarly, (D-4) implies that for every $n \in \mathcal{N}_-$,

$$\sqrt{n}\mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n'')) \leq \sqrt{n} \frac{1-q-\mu_n''}{q\mu_n''} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Therefore,

$$\lim_{n \rightarrow \infty} \sqrt{n}\mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n)) = 0.$$

Case (b): $\delta \in \{-\infty, \infty\}$.

This case is handled by the following lemma proved at the end of Appendix D.

Lemma D-1. *If r_n is such that, $\lim_{n \rightarrow \infty} \frac{|r_n - qn|}{\sqrt{n}} = \ell < \infty$, then for any sequence μ_n such that*

$$\lim_{n \rightarrow \infty} \sqrt{n}|1-q-\mu_n| = \infty,$$

we have that

$$\lim_{n \rightarrow \infty} \sqrt{n}\mathcal{R}_n(\Pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n)) = 0.$$

Case (c): $\delta \in \mathbb{R} \setminus \{0\}$.

Assume first that $\delta > 0$. This implies that, $\mu_n - (1-q) \sim \frac{\delta}{\sqrt{n}}$.

Remark that for n large enough, $\mu_n \geq 1-q$ and by (D-3),

$$\sqrt{n}\mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n)) = \sqrt{n} \frac{\mu_n - (1-q)}{(1-q)(1-\mu_n)} \mathbb{P}(D_{r_n:n} = 0).$$

By assumption,

$$\sqrt{n} \frac{\mu_n - (1-q)}{(1-q)(1-\mu_n)} \rightarrow \frac{\delta}{(1-q)q} \quad \text{as } n \rightarrow \infty.$$

Moreover,

$$\begin{aligned} \mathbb{P}(D_{r_n:n} = 0) &= \mathbb{P}\left(\sum_{i=1}^n D_i \leq n - r_n\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{n\mu_n(1-\mu_n)}} \sum_{i=1}^n (D_i - \mu_n) - \frac{n(1-\mu_n) - r_n}{\sqrt{n\mu_n(1-\mu_n)}} \leq 0\right). \end{aligned} \quad (\text{D-6})$$

To conclude the proof, we show that

$$\frac{1}{\sqrt{n\mu_n(1-\mu_n)}} \sum_{i=1}^n (D_i - \mu_n) - \frac{n(1-\mu_n) - r_n}{\sqrt{n\mu_n(1-\mu_n)}} \Rightarrow \mathcal{N}\left(\frac{\delta + \ell}{\sqrt{q(1-q)}}, 1\right) \quad \text{as } n \rightarrow \infty. \quad (\text{D-7})$$

Remark that,

$$\frac{n(1-\mu_n) - r_n}{\sqrt{n\mu_n(1-\mu_n)}} = \frac{n(1-q-\mu_n) + qn - r_n}{\sqrt{n\mu_n(1-\mu_n)}} \rightarrow -\frac{\delta + \ell}{\sqrt{q(1-q)}} \quad \text{as } n \rightarrow \infty.$$

Hence, it also converges in distribution. Moreover, for all $n \geq 1$ and $m \leq n$, let

$$Z_{n,m} := \frac{1}{\sqrt{n\mu_n(1-\mu_n)}} (D_m - \mu_n).$$

We have that for every $1 \leq m \leq n$,

$$\mathbb{E} [Z_{n,m}^2] = \frac{\mu_n(1-\mu_n)^2 + (1-\mu_n)\mu_n^2}{n\mu_n(1-\mu_n)} = \frac{1}{n}.$$

Hence, for any $n \geq 1$,

$$\sum_{m=1}^n \mathbb{E} [Z_{n,m}^2] = \frac{\mu_n(1-\mu_n)}{(1-q)q} \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (\text{D-8})$$

Furthermore, let $\epsilon > 0$ then, for n large enough, $|Z_{n,m}| < \epsilon$ almost surely for every $m \leq n$. Hence,

$$\lim_{n \rightarrow \infty} \sum_{m=1}^n \mathbb{E} [Z_{n,m}^2; |Z_{n,m}| > \epsilon] = 0. \quad (\text{D-9})$$

By (D-8) and (D-9) and Theorem D-1 stated below, we conclude that,

$$\frac{1}{\sqrt{n\mu_n(1-\mu_n)}} \sum_{m=1}^n (D_m - \mu_n) = \sum_{m=1}^n Z_{n,m} \implies \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty. \quad (\text{D-10})$$

Theorem D-1 (Linderberg-Feller theorem (see Theorem 3.4.10 in Durrett (2019))). *For each n , let $X_{n,m}$, $1 \leq m \leq n$, be independent random variables with $\mathbb{E}[X_{n,m}] = 0$. Suppose,*

1. $\sum_{m=1}^n \mathbb{E} [X_{n,m}^2] \rightarrow \sigma^2 > 0$,
2. For all $\epsilon > 0$, $\lim_{n \rightarrow \infty} \sum_{m=1}^n \mathbb{E} [X_{n,m}^2; |X_{n,m}| > \epsilon] = 0$.

Then, $\sum_{m=1}^n X_{n,m} \implies \mathcal{N}(0, \sigma^2)$ as $n \rightarrow \infty$.

We thus conclude from Slutsky's theorem that (D-7) is satisfied and for any $\delta' > 0$ such that $\mu_n - (1-q) \sim \frac{\delta}{\sqrt{n}}$, we have,

$$\sqrt{n}\mathcal{R}_n(\pi_n^r, \mathcal{B}(\mu_n)) \rightarrow \frac{\delta}{q(1-q)} \left(1 - \Phi \left(\frac{\ell + \delta}{\sqrt{q(1-q)}} \right) \right) \quad \text{as } n \rightarrow \infty.$$

We now consider the case where there exists $\delta > 0$, such that $1 - q - \mu_n \sim \frac{\delta}{\sqrt{n}}$. Note that (D-4) together with (D-6) implies that

$$\sqrt{n}\mathcal{R}_n(\pi_n^r, \mathcal{B}(\mu_n)) = \sqrt{n} \frac{1 - q - \mu_n}{q\mu_n} \mathbb{P} \left(\frac{1}{\sqrt{nq(1-q)}} \sum_{i=1}^n (D_i - \mu_n) - \frac{n(1-\mu_n) - r_n}{\sqrt{nq(1-q)}} > 0 \right).$$

We hence conclude by a similar argument that

$$\mathbb{P} \left(\frac{1}{\sqrt{nq(1-q)}} \sum_{i=1}^n (D_i - \mu_n) - \frac{n(1-\mu_n) - r_n}{\sqrt{nq(1-q)}} > 0 \right) \rightarrow 1 - \Phi \left(\frac{\delta - \ell}{\sqrt{q(1-q)}} \right) \quad \text{as } n \rightarrow \infty$$

and

$$\sqrt{n}\mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n)) \rightarrow \frac{\delta}{q(1-q)} \left(1 - \Phi \left(\frac{\delta - \ell}{\sqrt{q(1-q)}} \right) \right) \quad \text{as } n \rightarrow \infty.$$

Conclusion of step 1. We now prove that (D-5) holds.

For any $n \in \mathbb{N}$, let $\mu_n^* := \arg \max_{\mu \in [0,1]} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu))$. Note that this sequence is well defined because $\mu \mapsto \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu))$ is continuous on a compact set. First remark that for any sequence $(\mu'_n)_{n \in \mathbb{N}}$, if there exists a sequence $(\mu''_n)_{n \in \mathbb{N}}$ such that

$$\lim_{n \rightarrow \infty} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu'_n)) < \lim_{n \rightarrow \infty} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu''_n)),$$

then, one cannot have $\mu'_n \in \arg \max_{\mu \in [0,1]} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu))$ for all $n \in \mathbb{N}$. We will use this property to characterize the asymptotic behavior of $(\mu_n^*)_{n \in \mathbb{N}}$.

We consider the sequence, $(\sqrt{n}|\mu_n^* - (1-q)|)_{n \in \mathbb{N}}$ and first assume that it is unbounded. Therefore, we can consider an increasing mapping ψ from \mathbb{N} to \mathbb{N} such that the induced subsequence $(\sqrt{\psi(n)}|\mu_{\psi(n)}^* - (1-q)|)_{n \in \mathbb{N}}$ converges to ∞ . Moreover, let \mathbf{r}_ψ denote the subsequence defined as $(r_{\psi(n)})_{n \in \mathbb{N}}$. By case (b) defined above, this implies that

$$\lim_{n \rightarrow \infty} \sqrt{n}\mathcal{R}_n(\Pi_n^{\mathbf{r}_\psi}, \mathcal{B}(\mu_{\psi(n)}^*)) = 0.$$

In turn, consider the sequence $(\hat{\mu}_\psi(n))_{n \in \mathbb{N}}$ such that, $\sqrt{\psi(n)}|\hat{\mu}_\psi(n) - (1-q)| \rightarrow \delta \in \mathbb{R}^*$. It follows from case (c) that

$$\lim_{n \rightarrow \infty} \sqrt{n}\mathcal{R}_n(\Pi_n^{\mathbf{r}_\psi}, \mathcal{B}(\hat{\mu}_\psi(n))) > 0,$$

which contradicts the optimality of $\mu_{\psi(n)}^*$ for n large enough.

Hence, $(\sqrt{n}|\mu_n^* - (1-q)|)_{n \in \mathbb{N}}$ is bounded. By Bolzano Weirestrass, we can consider a subsequence that converges. Remark that, if the limit of this subsequence is 0, we obtain a contradiction by the same argument using case (a) and case (c). We thus conclude that any subsequence has to satisfy,

$$\lim_{n \rightarrow \infty} \sqrt{\psi(n)} (\mu_{\psi(n)}^* - 1 - q) = \delta,$$

with $\delta \in \mathbb{R} \setminus \{0\}$. Case (c) implies that,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sqrt{n}\mathcal{R}_n(\Pi_n^{\mathbf{r}_\psi}, \mathcal{B}(\mu_{\psi(n)}^*)) &= H^+(\delta, \ell) & \text{if } \delta > 0 \\ \lim_{n \rightarrow \infty} \sqrt{n}\mathcal{R}_n(\Pi_n^{\mathbf{r}_\psi}, \mathcal{B}(\mu_{\psi(n)}^*)) &= H^-(\delta, \ell) & \text{if } \delta < 0, \end{aligned}$$

where $H^+(\delta, \ell) := \frac{\delta}{q(1-q)} \left(1 - \Phi \left(\frac{\delta + \ell}{\sqrt{q(1-q)}} \right) \right)$ and $H^-(\delta, \ell) := \frac{\delta}{q(1-q)} \left(1 - \Phi \left(\frac{\delta - \ell}{\sqrt{q(1-q)}} \right) \right)$. Furthermore, remark that if $\delta > 0$, we must have $H^+(\delta, \ell) = \max(\max_{\delta' \geq 0} H^+(\delta', \ell), \max_{\delta' \geq 0} H^-(\delta', \ell))$ otherwise, we can construct a sequence that strictly improves the limit. A similar result holds if $\delta < 0$.

Therefore, for any converging subsequence $(\sqrt{\psi(n)}|\mu_{\psi(n)}^* - (1-q)|)_{n \in \mathbb{N}}$ we have that

$$\lim_{n \rightarrow \infty} \sqrt{n} \mathcal{R}_n(\Pi_n^{\mathbf{r}^\psi}, \mathcal{B}(\mu_{\psi(n)}^*)) = \max \left[\max_{\delta \geq 0} H^+(\delta, \ell'), \max_{\delta \geq 0} H^-(\delta, \ell') \right],$$

which implies that,

$$\lim_{n \rightarrow \infty} \sup_{F \in \mathcal{F}} \sqrt{n} \mathcal{R}_n(\pi_n^{\mathbf{r}}, F) = \lim_{n \rightarrow \infty} \sqrt{n} \mathcal{R}_n(\Pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n^*)) = \max \left[\max_{\delta \geq 0} H^+(\delta, \ell'), \max_{\delta \geq 0} H^-(\delta, \ell') \right].$$

Step 2: $\ell = \infty$. We show that in this case,

$$\lim_{n \rightarrow \infty} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi_n^{\mathbf{r}}, F) = \infty.$$

Remark that, it is sufficient to construct a sequence μ_n such that,

$$\lim_{n \rightarrow \infty} \mathcal{R}_n(\pi_n^{\mathbf{r}}, \mathcal{B}(\mu_n)) = \infty$$

We assume that for n large enough, $r_n \leq qn$ (the case $r_n \geq qn$ is similar).

Let's consider the sequence of means such that for each $n \geq 1$, $\mu_n = 1 - \frac{r_n}{n}$. For n large enough we have that $r_n \leq qn$ which implies that $\mu_n \geq 1 - q$. Therefore, for n large enough the expected relative regret of $\pi_n^{\mathbf{r}}$ against $\mathcal{B}(\mu_n)$ is given by (D-3). Remark that,

$$\begin{aligned} \mathbb{P}(D_{r_n:n} = 0) &= \mathbb{P}\left(\sum_{i=1}^n D_i \leq n - r_n\right) \\ &\stackrel{(a)}{=} \mathbb{P}\left(\sum_{i=1}^n (D_i - \mu_n) \leq 0\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{n}\mu_n(1-\mu_n)} \sum_{i=1}^n (D_i - \mu_n) \leq 0\right) \stackrel{(b)}{\rightarrow} \frac{1}{2} \quad \text{as } n \rightarrow \infty, \end{aligned} \quad (\text{D-11})$$

where (a) holds as $\mu_n = 1 - \frac{r_n}{n}$ and (b) follows from (D-10). We further remark that,

$$\sqrt{n} \frac{\mu_n - (1-q)}{(1-q)(1-\mu_n)} \geq \sqrt{n} \frac{\mu_n - (1-q)}{(1-q)q} = \frac{(qn - r_n)}{(1-q)q\sqrt{n}} \rightarrow \infty \quad \text{as } n \rightarrow \infty, \quad (\text{D-12})$$

where the limit holds because $\ell = \infty$. Finally, (D-10) and (D-12) imply that

$$\lim_{n \rightarrow \infty} \sqrt{n} \mathcal{R}_n(\pi^{OS_{r_n}}, \mathcal{B}(\mu_n)) = \infty.$$

□

Proof of Lemma C-2. Define the function H from \mathbb{R} to \mathbb{R} as,

$$H : \ell \mapsto \max \left[\max_{\delta \geq 0} H^+(\delta, \ell), \max_{\delta \geq 0} H^-(\delta, \ell) \right].$$

We show that H is an even function increasing on $[0, \infty)$.

We remark that for any $\ell \in \mathbb{R}$,

$$H^+(\delta, \ell) = H^-(\delta, -\ell),$$

which shows that H is even. To show that it is increasing on $[0, \infty)$, consider $\ell \geq 0$. As Φ is increasing, we have for any $\delta \geq 0$ that $\Phi\left(\frac{\delta+\ell}{\sqrt{q(1-q)}}\right) \geq \Phi\left(\frac{\delta-\ell}{\sqrt{q(1-q)}}\right)$ and, $H(\ell) = \max_{\delta \geq 0} H^-(\delta, \ell)$.

Let $\delta^* \in \arg \max_{\delta \geq 0} H^-(\delta, \ell)$. For any $\ell' \geq \ell$, we note that $\Phi\left(\frac{\delta^*-\ell}{\sqrt{q(1-q)}}\right) \geq \Phi\left(\frac{\delta^*-\ell'}{\sqrt{q(1-q)}}\right)$. Therefore, $H^-(\delta^*, \ell) \leq H^-(\delta^*, \ell')$. We conclude that,

$$H(\ell) = H^-(\delta^*, \ell) \leq H^-(\delta^*, \ell') \leq H(\ell').$$

Since H is an even increasing function, its minimum is achieved at 0 and we conclude that, for every $\ell \in \mathbb{R}$

$$\max \left[\max_{\delta \geq 0} H^+(\delta, \ell), \max_{\delta \geq 0} H^-(\delta, \ell) \right] = H(\ell) \geq H(0) = C^*.$$

□

Proof of Lemma D-1. For the sake of simple notations, consider the sequence $\alpha_n = 1 - \mu_n$ and consider a converging sequence $(\beta_n)_{n \in \mathbb{N}}$ that is a subsequence of $(\alpha_n)_{n \in \mathbb{N}}$. Let β denotes its limit. We assume that for n large enough $\beta_n \in [0, q]$ therefore, $\beta \in [0, q]$. (The case where $\beta_n \in [q, 1]$ is treated by similar arguments).

For every $r \in \{1, \dots, n\}$, and $\alpha \in [0, 1]$, let

$$\phi_r^n(\alpha) = \mathcal{R}_n \left(\pi^{OS_r}, \mathcal{B}(1 - \alpha) \right)$$

and remark that since $\beta_n \leq q$ for n large enough, we have by Lemma A-2 that

$$\phi_r^n(\beta_n) = \sqrt{n} \frac{(q - \beta_n)}{(1 - q)\beta_n} B_{r_n, n}(\beta_n).$$

Case 1: $\beta = 0$. By Taylor expansion, we obtain that for every $n \geq 1$, there exists $\xi_n \in [0, \beta_n]$ such that,

$$\phi_{r_n}^n(\beta_n) = \phi_{r_n}^n(0) + \phi'_{r_n}(\xi_n) \beta_n. \quad (\text{D-13})$$

Since $B'_{r_n, n}(\alpha) = n b_{r_n-1, n-1}(\alpha)$, we obtain that for $\alpha \in [0, q]$,

$$\phi'_{r_n}(\alpha) = \frac{n(q - \alpha) \alpha b_{r_n-1, n-1}(\alpha) - q B_{r_n, n}(\alpha)}{(1 - q)\alpha^2} \leq \frac{q n b_{r_n-1, n-1}(\alpha)}{(1 - q)\alpha}.$$

Furthermore,

$$\begin{aligned}\phi'_{r_n}(\xi_n)\beta_n &\leq \frac{qn}{1-q} \binom{n-1}{r_n-1} \xi_n^{r_n-2} (1-\xi_n)^{n-r_n} \beta_n \\ &\leq \frac{qn}{1-q} \binom{n-1}{r_n-1} \xi_n^{r_n-2} \beta_n \stackrel{(a)}{\leq} \frac{qn}{1-q} \binom{n-1}{r_n-1} \beta_n^{r_n-1},\end{aligned}\tag{D-14}$$

where (a) holds as $\xi_n \leq \beta_n$ and $r_n \geq 2$ for n large enough. The latter holds because $q > 0$ and $\lim_{n \rightarrow \infty} \frac{|qn-r_n|}{\sqrt{n}} < \infty$ which implies that $r_n \rightarrow \infty$ as n gets large.

Remarking that $\phi'_{r_n}(0) = 0$ and that $\binom{n-1}{r_n-1} \sim q \binom{n}{r_n}$, we conclude from (D-13) and (D-14) that it is sufficient to show that, $n^{3/2} \binom{n}{r_n} \beta_n^{r_n-1} \rightarrow 0$. Note that, $r_n = qn + O(\sqrt{n})$ therefore,

$$\binom{n}{r_n} \sim \sqrt{\frac{n}{2\pi r_n(n-r_n)}} \frac{n^n}{r_n^{r_n} (n-r_n)^{n-r_n}} \sim \frac{1}{\sqrt{2\pi q(1-q)n}} \frac{n^n}{r_n^{r_n} (n-r_n)^{n-r_n}}.$$

Furthermore, we have that, $\frac{r_n}{n} \rightarrow q$. Therefore,

$$\frac{n^n}{r_n^{r_n} (n-r_n)^{n-r_n}} = \exp\left(r_n \log\left(\frac{n}{r_n}\right) + (n-r_n) \log\left(\frac{n}{n-r_n}\right)\right) \stackrel{(a)}{=} \exp(O(n)),$$

where (a) holds because $\frac{n}{r_n} \rightarrow \frac{1}{q}$ and $\frac{n}{n-r_n} \rightarrow \frac{1}{1-q}$. Moreover, note that $n = o(r_n \log(\beta_n))$ because $\beta_n \rightarrow 0$. As a consequence,

$$\begin{aligned}n^{3/2} \binom{n}{r_n} \beta_n^{r_n-1} &\sim \frac{1}{\sqrt{2\pi q(1-q)}} \cdot n \cdot \exp(O(n)) \cdot \beta_n^{r_n-1} \\ &= \frac{1}{\sqrt{2\pi q(1-q)}} \cdot n \cdot \exp((r_n-1) \log(\beta_n) + o(r_n \log(\beta_n))) \rightarrow 0,\end{aligned}$$

where the limit holds because $(r_n-1) \log(\beta_n) \rightarrow -\infty$ and $n = \exp(o(r_n \log(\beta_n)))$.

Case 2: $\beta > 0$. Note that, by assumption on the sequence $(\alpha_n)_{n \in \mathbb{N}}$, we have that for n large enough, $\beta_n \leq \frac{r_n}{n}$. Indeed, if $\beta_n > \frac{r_n}{n}$, we would have that $\sqrt{n}(q - \beta_n) \leq \sqrt{n}(q - \frac{r_n}{n})$ which, in turn, would contradict that $\lim_{n \rightarrow \infty} \frac{|qn-r_n|}{\sqrt{n}} < \infty$ since $\lim_{n \rightarrow \infty} \sqrt{n}|q - \beta_n| = \infty$. Therefore, for n large enough, we have,

$$\begin{aligned}\mathbb{P}(D_{r_n:n} = 0) &= \mathbb{P}\left(\sum_{i=1}^n D_i < n(1-\beta_n) - r_n + n\beta_n\right) \\ &= \mathbb{P}\left(\sum_{i=1}^n D_i < n\mathbb{E}[D] - (r_n - n\beta_n)\right) \stackrel{(a)}{\leq} e^{-2(\beta_n - \frac{r_n}{n})^2 n} = e^{-2\left(\sqrt{n}(\beta_n - q) + \frac{qn-r_n}{\sqrt{n}}\right)^2},\end{aligned}$$

where (a) follows from Hoeffding inequality for bounded random variables (Theorem 2 in [Hoeffding \(1994\)](#)). Moreover, note that $\left(\sqrt{n}(\beta_n - q) + \frac{qn-r_n}{\sqrt{n}}\right)^2 = (\sqrt{n}(\beta_n - q))^2 (1 + o(1))$. Hence,

$$\sqrt{n}\mathcal{R}_n(\pi_n^*, \mathcal{B}(1-\beta_n)) \sim \sqrt{n} \frac{(q-\beta_n)}{(1-q)\beta} \mathbb{P}(D_{r_n:n} = 0) \leq \sqrt{n} \frac{(q-\beta_n)}{(1-q)\beta} e^{-2(\sqrt{n}(\beta_n - q))^2 (1+o(1))} \rightarrow 0,$$

where the limit holds because $\sqrt{n}|\beta_n - q| \rightarrow \infty$ as $n \rightarrow \infty$. Therefore, for any converging subsequence $(\beta_n)_{n \in \mathbb{N}}$ of the sequence $(\alpha_n)_{n \in \mathbb{N}}$, we have $\sqrt{n}\mathcal{R}_n(\pi_n^r, \mathcal{B}(1 - \beta_n)) \rightarrow 0$. This implies the same result on the sequence $(\alpha_n)_{n \in \mathbb{N}}$ itself. \square

E Additional Results and their Proofs

Lemma E-1. *Fix $n \geq 1$ and $\pi \in \Pi_n$. Then, problem (2) of finding the worst case performance for the policy π is equivalent to*

$$\begin{aligned} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) &= \inf_{z \in \mathbb{R}_+} z \\ \text{s.t.} \quad \mathcal{C}(\pi, F, n) &\leq (z + 1)\text{opt}(F) \quad \forall F \in \mathcal{F}, \end{aligned}$$

in the sense that both problems admit the same value.

Proof of Lemma E-1. Fix $\pi \in \Pi_n$. We have that $\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) \geq 0$ and

$$\sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) = \inf_{z \in \mathbb{R}_+} z \tag{E-2a}$$

$$\text{s.t.} \quad \mathcal{R}_n(\pi, F) \leq z \quad \forall F \in \mathcal{F}. \tag{E-2b}$$

We first show that for every $F \in \mathcal{F}$ and $z \in \mathbb{R}_+$, the following equivalence holds.

$$\mathcal{R}_n(\pi, F) \leq z \quad \text{if and only if} \quad \mathcal{C}(\pi, F, n) \leq (z + 1)\text{opt}(F).$$

Recall that, for every $F \in \mathcal{F}$, we have $\mathcal{R}_n(\pi, F) = \frac{\mathcal{C}(\pi, F, n)}{\text{opt}(F)} - 1$. Moreover, $\text{opt}(F) \geq 0$ for all $F \in \mathcal{F}$. When $\text{opt}(F) > 0$, the equivalence holds trivially. If $\text{opt}(F) = 0$, we consider two cases. If $\mathcal{C}(\pi, F, n) > 0$, we remark that both inequalities do not hold for any $z \in \mathbb{R}_+$. If $\mathcal{C}(\pi, F, n) = 0$, recall that by convention we set $\mathcal{R}_n(\pi, F) = 0$. Hence, both inequalities are satisfied for all $z \in \mathbb{R}_+$.

As a consequence, problem (E-2) is equivalent to

$$\begin{aligned} \sup_{F \in \mathcal{F}} \mathcal{R}_n(\pi, F) &= \inf_{z \in \mathbb{R}_+} z \\ \text{s.t.} \quad \mathcal{C}(\pi, F, n) &\leq (z + 1)\text{opt}(F) \quad \forall F \in \mathcal{F}. \end{aligned}$$

\square

Lemma E-2. *For every n , the following two conditions cannot hold simultaneously.*

$$\begin{aligned} \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) &> \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) \\ \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)) &< \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)). \end{aligned}$$

Proof of Lemma E-2. Assume by contradiction that there exists n such that,

$$\begin{aligned} \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) &> \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) \\ \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)) &< \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)). \end{aligned}$$

We have that,

$$\begin{aligned} \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) &\stackrel{(a)}{\leq} \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)) \\ &\stackrel{(b)}{<} \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)) \stackrel{(c)}{\leq} \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)), \end{aligned}$$

where (a) and (c) follow from Lemma E-4, and (b) holds by assumption. We therefore obtain a contradiction. \square

Lemma E-3. For $n \geq \frac{2}{\min(q, (1-q))^2}$,

$$\begin{aligned} \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) &\leq \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_1}, \mathcal{B}(\mu)) \\ \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)) &\geq \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_n}, \mathcal{B}(\mu)). \end{aligned}$$

Proof of Lemma E-3. For each $r \in \{1, \dots, n\}$ we define ϕ_r^n such that for every $\alpha \in [0, 1]$,

$$\phi_r^n(\alpha) = \mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(1-\alpha)).$$

We first show that,

$$\sup_{\alpha \leq q} \phi_1^n(\alpha) \geq \sup_{\alpha \geq q} \phi_1^n(\alpha). \quad (\text{E-4})$$

Note that by Lemma A-2 we have for all $r \in \{1, \dots, n\}$,

$$\phi_r^n(\alpha) = \begin{cases} \frac{(q-\alpha)B_{r,n}(\alpha)}{(1-q)\alpha} & \text{if } \alpha \in [0, q], \\ \frac{(\alpha-q)(1-B_{r,n}(\alpha))}{(1-\alpha)q} & \text{if } \alpha \in [q, 1]. \end{cases}$$

Observe that by definition, $B_{1,n}(\alpha) = \sum_{j=1}^n \binom{n}{j} \alpha^j (1-\alpha)^{n-j} = 1 - (1-\alpha)^n$. Hence, for $\alpha \geq q$

$$\phi_1^n(\alpha) = \frac{(\alpha-q)(1-\alpha)^n}{q(1-\alpha)} \leq \frac{(1-q)^n}{q}.$$

Moreover, $n \geq \frac{2}{\min(q, (1-q))^2}$ implies that $\frac{1}{n} \leq \min(q, (1-q))^2 \leq q^2 \leq q$. Therefore, we have that,

$$\sup_{\alpha \leq q} \phi_1^n(\alpha) \geq \phi_1^n\left(\frac{1}{n}\right) = \frac{(q-n^{-1})(1-(1-n^{-1})^n)}{(1-q)n^{-1}} \stackrel{(a)}{\geq} \frac{q(1-q)(1-e^{-1})}{(1-q)q^2} = \frac{1-e^{-1}}{q}$$

where inequality (a) holds because $n^{-1} \leq q^2$ and $\left(1 - \frac{1}{n}\right)^n \leq e^{-1}$. When $q \geq \frac{1}{2}$, we remark that since $n \geq 4$, (E-4) must hold because,

$$1 - e^{-1} \geq \frac{1}{2^n} \geq (1 - q)^n.$$

To prove equation (E-4) for $q \leq \frac{1}{2}$, as $n \geq 2q^{-2}$, we remark that, $(1 - q)^n \leq (1 - q)^{q^{-2}}$, therefore it is sufficient to show that

$$(1 - e^{-1}) \geq (1 - q)^{q^{-2}}.$$

This holds because $q \mapsto (1 - q)^{q^{-2}}$ is non-decreasing on $[0, \frac{1}{2}]$ and the inequality is true at $q = \frac{1}{2}$.

We similarly show that,

$$\sup_{\alpha \leq q} \phi_n^n(\alpha) \leq \sup_{\alpha \geq q} \phi_n^n(\alpha).$$

To do so, we first remark that, $B_{n,n}(\alpha) = \alpha^n$ and, for any $\alpha \leq q$, $\phi_n^n(\alpha) \leq \frac{q^n}{1-q}$ and,

$$\sup_{\alpha \geq q} \phi_n^n(\alpha) \geq \frac{1 - e^{-1}}{1 - q}.$$

We then conclude in a similar way. □

Lemma E-4. For any $\mu \leq 1 - q$ and any $r \in \{1, \dots, n - 1\}$,

$$\mathcal{R}_n\left(\pi^{OS_r}, \mathcal{B}(\mu)\right) \leq \mathcal{R}_n\left(\pi^{OS_{r+1}}, \mathcal{B}(\mu)\right).$$

Inequality is strict for $\mu \notin \{0, 1 - q\}$. Furthermore,

$$\sup_{\mu \in [0, 1 - q]} \mathcal{R}_n\left(\pi^{OS_r}, \mathcal{B}(\mu)\right) < \sup_{\mu \in [0, 1 - q]} \mathcal{R}_n\left(\pi^{OS_{r+1}}, \mathcal{B}(\mu)\right).$$

Similarly, for any $\mu \geq 1 - q$ and any $r \in \{1, \dots, n - 1\}$,

$$\mathcal{R}_n\left(\pi^{OS_r}, \mathcal{B}(\mu)\right) \geq \mathcal{R}_n\left(\pi^{OS_{r+1}}, \mathcal{B}(\mu)\right).$$

Inequality is strict for $\mu \notin \{1 - q, 1\}$. Furthermore

$$\sup_{\mu \in [1 - q, 1]} \mathcal{R}_n\left(\pi^{OS_r}, \mathcal{B}(\mu)\right) > \sup_{\mu \in [1 - q, 1]} \mathcal{R}_n\left(\pi^{OS_{r+1}}, \mathcal{B}(\mu)\right).$$

Proof of Lemma E-4. For each $r \in \{1, \dots, n\}$ we define ϕ_r^n such that for every $\alpha \in [0, 1]$,

$$\phi_r^n(\alpha) = \mathcal{R}_n\left(\pi^{OS_r}, \mathcal{B}(1 - \alpha)\right).$$

Remark that for all $r \in \{1, \dots, n-1\}$ and $\alpha \leq q$ we obtain from Lemma A-2 that,

$$\begin{aligned}\phi_r^n(\alpha) &= \frac{(q-\alpha)B_{r,n}(\alpha)}{(1-q)\alpha} \\ &= \frac{(q-\alpha)B_{r+1,n}(\alpha)}{(1-q)\alpha} + \frac{(q-\alpha)b_{r+1,n}(\alpha)}{(1-q)\alpha} \\ &= \phi_{r+1}^n(\alpha) + \frac{(q-\alpha)b_{r+1,n}(\alpha)}{(1-q)\alpha} \geq \phi_{r+1}^n(\alpha).\end{aligned}$$

Note that, $b_{r+1,n}(\alpha) > 0$ for $\alpha \in (0, 1)$ thus the above inequality is an equality only for $\alpha \in \{0, q\}$. Moreover, let $\alpha^* \in \arg \max_{\alpha \in [0, q]} \phi_r^n(\alpha)$ (which exists by continuity of ϕ_r^n). Remark that $\alpha^* \in (0, q)$ because, $\phi_r^n(0) = \phi_r^n(q) = 0$ and $\phi_r^n(q/2) > 0$. Therefore, we have

$$\sup_{\alpha \in [0, q]} \phi_r^n(\alpha) = \phi_r^n(\alpha^*) > \phi_{r+1}^n(\alpha^*) \geq \sup_{\alpha \in [0, q]} \phi_{r+1}^n(\alpha).$$

Similarly, we have that for $\alpha \geq q$,

$$\begin{aligned}\phi_r^n(\alpha) &= \frac{(\alpha-q)(1-B_{r,n}(\alpha))}{q(1-\alpha)} \\ &= \frac{(\alpha-q)(1-B_{r+1,n}(\alpha))}{q(1-\alpha)} - \frac{(\alpha-q)b_{r+1,n}(\alpha)}{q(1-\alpha)} \\ &= \phi_{r+1}^n(\alpha) - \frac{(\alpha-q)b_{r+1,n}(\alpha)}{q(1-\alpha)} \leq \phi_{r+1}^n(\alpha).\end{aligned}$$

We conclude by an argument similar to the one derived in the case where $\alpha \in [0, q]$ that equality holds only for $\alpha \in \{q, 1\}$ and that, $\sup_{\alpha \in [q, 1]} \phi_r^n(\alpha) < \sup_{\alpha \in [q, 1]} \phi_{r+1}^n(\alpha)$. \square

Lemma E-5. Let $U(n) = \int_0^\infty 2 \exp\left(-\frac{n\epsilon^2}{18+8\epsilon} \min(q, 1-q)\right) d\epsilon$. Then, $U(n) = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

Proof of Lemma E-5. Remark that

$$\begin{aligned}U(n) &= \int_0^\infty 2 \exp\left(-\frac{n\epsilon^2}{18+8\epsilon} \min(q, 1-q)\right) d\epsilon \\ &\leq 2 \left(\int_0^{\frac{18}{8}} \exp\left(-\frac{n\epsilon^2}{18} \min(q, 1-q)\right) d\epsilon + \int_{\frac{18}{8}}^\infty \exp\left(-\frac{n\epsilon}{8} \min(q, 1-q)\right) d\epsilon \right) \\ &= 2 \int_0^{\frac{18}{8}} \exp\left(-\frac{n\epsilon^2}{18} \min(q, 1-q)\right) d\epsilon + \mathcal{O}\left(\frac{1}{n}\right) \\ &\leq \frac{2}{\sqrt{n}} + 2 \int_{\frac{1}{\sqrt{n}}}^{\frac{18}{8}} \frac{\epsilon}{\epsilon} \exp\left(-\frac{n\epsilon^2}{18} \min(q, 1-q)\right) d\epsilon + \mathcal{O}\left(\frac{1}{n}\right) \\ &\leq \frac{2}{\sqrt{n}} + 2\sqrt{n} \int_{\frac{1}{\sqrt{n}}}^{\frac{18}{8}} \epsilon \exp\left(-\frac{n\epsilon^2}{18} \min(q, 1-q)\right) d\epsilon + \mathcal{O}\left(\frac{1}{n}\right) \\ &= \frac{1}{\sqrt{n}} \left(2 + \frac{18}{\min(q, 1-q)} \exp\left(-\frac{\min(q, 1-q)}{18}\right) \right) + o\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

\square

F Discussion: Algorithms and Performance

F.1 Suboptimality of SAA

Proposition 3 provides a necessary condition for a policy to be optimal. We present in Figure 3 a counter-example showing that SAA is not always achieving this necessary condition. Figure 3 presents the performance of SAA against Bernoulli distributions with different means with a value of $q = .9$ and $n = 20$.

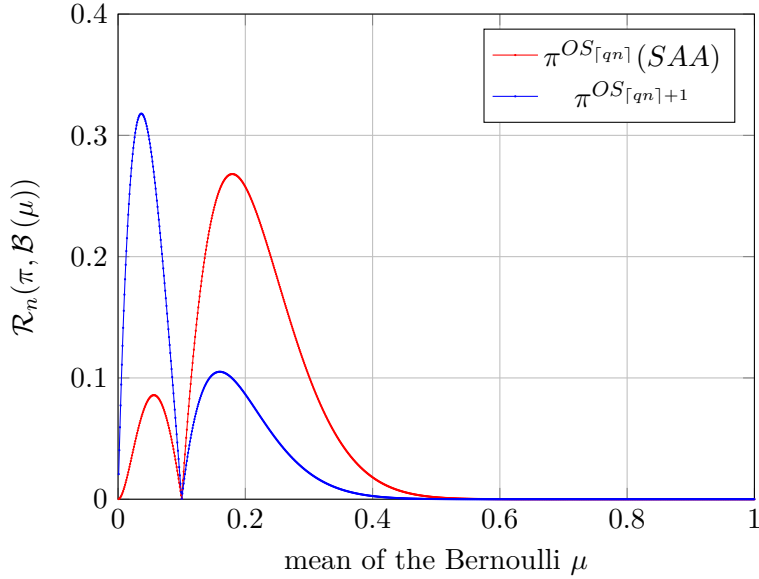


Figure 3: **Performance of the $[qn]^{\text{th}}$ (SAA) and $([qn] + 1)^{\text{th}}$ order statistic policies against Bernoulli distributions.** The figure depicts the performance of two order statistic policies against Bernoulli distribution as the mean μ varies ($q = .9$, $n = 20$).

We observe in that case that

$$\sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{\text{SAA}}, \mathcal{B}(\mu)) < \sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{\text{SAA}}, \mathcal{B}(\mu)).$$

This implies the suboptimality of SAA according to Proposition 3. We note from Figure 3 that, in this example, SAA suffers from a larger regret in the mode associated with large values of μ compared to the mode associated with smaller values of μ . In contrast, the $([qn] + 1)^{\text{th}}$ order statistic policy suffers from a larger regret than SAA on the mode associated with the small values of μ and from a smaller one in the regime where μ is large. This observation implies that a carefully chosen randomization of both policies would perform strictly better than SAA.

F.2 Insight on the Minimax Optimal Policy

The algorithm derived in Theorem 4 is defined as a randomization of the $(k - 1)^{\text{th}}$ and k^{th} order statistics for some $k \in \{2, \dots, n\}$. However, Theorem 4 does not provide any quantification of

the value of k . We present in Table 5 the values of k for different critical quantiles obtained by computing the minimax optimal policy for sample size smaller than 200. Recall that SAA uses the

	$k = \lceil qn \rceil$	$k = \lceil qn \rceil + 1$
$q = .7$	40.5%	59.5%
$q = .8$	41%	59%
$q = .9$	42.5%	57.5%

Table 5: **Parameter of the minimax optimal policy.** The table presents the proportion of time the parameter k (defined in Theorem 4) is respectively equal to $\lceil qn \rceil$ and $\lceil qn \rceil + 1$ for different values of q . This proportion is derived by computing the parameters of the optimal policy for any data size smaller than 200.

$\lceil qn \rceil^{th}$ order statistic. Therefore, Table 5 shows that for the first 200 samples, the minimax optimal policy always has in its support SAA and a neighboring order statistic. The relation between k and $\lceil qn \rceil$ is further discussed in the proof of Theorem 5. We show in the proof that k has to scale as $\lceil qn \rceil + o(\sqrt{n})$.

F.3 Algorithmic Implementation of the Optimal Policy

Theorem 4 presents the structure of the optimal data-driven policy. We next detail how to find the optimal tuning parameters k and γ for an optimal policy. To that end, we establish an additional structural result on single order statistic policies. We show that for any $r \in \{1, \dots, n-1\}$,

$$\mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(\mu)) \leq \mathcal{R}_n(\pi^{OS_{r+1}}, \mathcal{B}(\mu)) \quad \text{for all } \mu \leq 1 - q \quad (\text{F-5})$$

$$\mathcal{R}_n(\pi^{OS_r}, \mathcal{B}(\mu)) \geq \mathcal{R}_n(\pi^{OS_{r+1}}, \mathcal{B}(\mu)) \quad \text{for all } \mu \geq 1 - q. \quad (\text{F-6})$$

This is formally stated by Lemma E-4 presented in Appendix E.

Equations (F-5) and (F-6) formalize the fact that, against Bernoulli distributions, the performance of smaller order statistics is worse than larger ones when the mean is large, as they tend to underestimate the optimal inventory quantity. On the contrary, they perform better than larger order statistics when the mean is smaller than $1 - q$ since underestimating is valuable in that case.

Given equations (F-5) and (F-6), we now present an efficient algorithm to compute the parameters of the optimal policy. Algorithm 1 only needs to perform $\mathcal{O}(\log(n))$ line searches in order to find an order statistic k such that (16) and (17) are satisfied.

Data: critical fractile q , number of samples n

Result: Order statistic ranking k , weight γ and optimal value \mathcal{R}_n^*

if (12) and (13) hold **then**

Set $j = 1$ and $k = n$;

while $j < k$ **do**

$m = (j + k)/2$;

if $\sup_{\mu \in [1-q, 1]} \mathcal{R}_n(\pi^{OS_m}, \mathcal{B}(\mu)) - \sup_{\mu \in [0, 1-q]} \mathcal{R}_n(\pi^{OS_m}, \mathcal{B}(\mu)) \geq 0$ **then**

$j = m + 1$;

else

$k = m$;

end

end

Find the solution γ of the following equation by performing a line search to solve

$$\sup_{\mu \in [1-q, 1]} \gamma \mathcal{R}_n(\pi^{OS_k}, \mathcal{B}(\mu)) + (1 - \gamma) \mathcal{R}_n(\pi^{OS_{k-1}}, \mathcal{B}(\mu)) =$$

$$\sup_{\mu \in [0, 1-q]} \gamma \mathcal{R}_n(\pi^{OS_k}, \mathcal{B}(\mu)) + (1 - \gamma) \mathcal{R}_n(\pi^{OS_{k-1}}, \mathcal{B}(\mu));$$

else

$\gamma = 1$;

 If (12) does not hold, $k = 1$, whereas if (13) does not hold, $k = n$;

end

Set $\mathcal{R}_n^* = \sup_{\mu \in [1-q, 1]} \gamma \mathcal{R}_n(\pi^{OS_k}, \mathcal{B}(\mu)) + (1 - \gamma) \mathcal{R}_n(\pi^{OS_{k-1}}, \mathcal{B}(\mu))$;

return $k, \gamma, \mathcal{R}_n^*$;

Algorithm 1: Optimal data-driven policy