

# Uniform Loss Algorithms for Online Stochastic Decision-Making With Applications to Bin Packing

Siddhartha Banerjee\*  
 sbanerjee@cornell.edu  
 Cornell University  
 Ithaca, NY

Daniel Freund  
 dfreund@mit.edu  
 MIT  
 Cambridge, MA

## ABSTRACT

We consider a general class of finite-horizon online decision-making problems, where in each period a controller is presented a stochastic arrival and must choose an action from a set of permissible actions, and the final objective depends only on the aggregate type-action counts. Such a framework encapsulates many online stochastic variants of common optimization problems including bin packing, generalized assignment, and network revenue management. In such settings, we study a natural model-predictive control algorithm that in each period, acts greedily based on an updated certainty-equivalent optimization problem. We introduce a simple, yet general, condition under which this algorithm obtains uniform additive loss (independent of the horizon) compared to an optimal solution with full knowledge of arrivals. Our condition is fulfilled by the above-mentioned problems, as well as more general settings involving piece-wise linear objectives and offline index policies, including an airline overbooking problem.

## CCS CONCEPTS

• Theory of computation → Design and analysis of algorithms.

## KEYWORDS

online stochastic decision-making, approximate dynamic programming, model-predictive control, online bin packing

## ACM Reference Format:

Siddhartha Banerjee and Daniel Freund. 2020. Uniform Loss Algorithms for Online Stochastic Decision-Making With Applications to Bin Packing. In *ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '20 Abstracts)*, June 8–12, 2020, Boston, MA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3393691.3394224>

In this work, we focus on a class of online decision-making problems characterized by three features:

- *Exchangeable Actions*: The system evolves over a finite horizon, with the objective depending only on the aggregate counts of different actions taken over the horizon (and not on the order or timing of the actions).

\*Both authors contributed equally to this research.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*SIGMETRICS '20 Abstracts, June 8–12, 2020, Boston, MA, USA*

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7985-4/20/06.

<https://doi.org/10.1145/3393691.3394224>

- *Exogenous Randomness*: System dynamics are subject to *exogenous* stochastic fluctuations, i.e., random arrivals and events which are conditionally independent of the system state.
- *Limited Distributional Knowledge*: The controller has limited distributional information (Eg. noisy mean estimates) about arrivals; these are obtained, for example, from historical traces.

These three features are present in many problems of practical interest, across various fields. For now, the following canonical examples serve to illustrate some settings of interest:

- Resource allocation and generalized assignment: A finite supply of resources is allocated to arriving demand, with varying requirements and valuations, so as to maximize overall rewards.
- Online bin-packing: Cloud-computing jobs with different requirements arrive over a finite horizon, and the controller needs to load them on existing VMs with sufficient free resources, or start-up new VMs, while minimizing operating costs.
- Online routing in batch-processing systems: Partitioning arriving compute jobs into batches in batch-processing systems, shipments among trucks in logistics systems, and online financial transactions among payment-processing firms with different contract structures.

In all the above settings, given distributional knowledge of arrivals, the underlying control problems can be formulated as *Markov Decision Processes* (MDP), and corresponding optimal solutions computed via dynamic programming. However, due to their combinatorial nature, these problems suffer from the ‘curse of dimensionality’, due to which computing optimal policies is typically intractable. Furthermore, such intractable optimal policies also do not provide easy benchmarks for evaluating the performance of simpler heuristics.

An alternate control paradigm in such settings is that of *model-predictive control*, wherein the controller makes decisions based on some offline benchmark. One natural class of benchmarks are the so-called *prophet benchmarks* – hindsight optimal policies with full information of arrivals. Controls based on such benchmarks are often simple and easy to interpret; moreover, the benchmark also provides a natural way of computing performance bounds for such policies. However, most policies are known to incur an additive loss compared to the prophet benchmark<sup>1</sup> which grows

<sup>1</sup>The additive loss vis-à-vis the hindsight-optimal is classically referred to as the *regret* of a policy; in particular, we focus on the *expected sample-path regret*. This term however is interpreted differently in different communities, and in lieu of this, we use additive loss throughout this work.

with the horizon-length  $T$ ; in particular, for the problems we study, existing algorithms typically incur either constant-factor multiplicative losses (i.e.,  $\Theta(T)$  additive losses), or  $\Theta(\sqrt{T})$  additive losses. Significantly, however, such performance guarantees often coincide with *minimax bounds* for these settings, and thus hold even under adversarial arrivals. This leaves open the question as to whether these algorithms are too pessimistic, and if better performance is achievable in more structured settings, and in particular, given distributional assumptions on the arrivals (and knowledge of the corresponding distribution).

In light of the above, our aim in this work is to characterize conditions under which we can obtain algorithms that have a *uniform additive loss* (i.e., independent of  $T$ ) compared to a hindsight optimal policy. Such a guarantee is particularly strong, as it immediately implies small additive loss against the optimal policy, despite this being infeasible to compute in the settings we consider. We build on a recent line of work, started by [1] and [4], that provides uniform-loss algorithms for online multidimensional knapsack and matching problems, and extend these to obtain a simple but far-reaching condition for obtaining uniform-loss algorithms for a wide class of online decision-making problems. From a theoretical viewpoint, our approach helps characterize how performance guarantees for such policies depend on a combination of regularity properties of the stochastic arrival process, and the underlying geometry of the control problems. Significantly, our theoretical insights lead to new algorithms for several widely-studied problems, including the generalized assignment and bin-packing problems; our algorithms are simple, and yet perform order-wise better, both in theory and in experiments, than existing algorithms.

## Our Contribution

Our main contributions fall into two categories. First, we propose a unified algorithmic framework for a wide class of finite-horizon online control problems, and provide conditions under which this gives a *uniform loss guarantee* - additive loss compared to a hindsight (or clairvoyant) optimal solution which is independent of the horizon length. Next, we apply the framework to several widely-studied problems, and show how it leads to algorithms with stronger guarantees compared to the state-of-the-art in these settings.

*Framework, Algorithm and Performance Guarantees.* Stated informally, we obtain uniform loss guarantees for the following kind of problem: over a horizon of  $T$  periods, the controller observes a random type arriving in each period, and needs to take a type-specific action in response. We assume that we have some limited distributional information about the type arrivals (e.g., a single sample of the arrivals over the full time horizon). For the random arrivals we require neither independence nor identical distributions, but instead impose two light regularity conditions. Finally, our terminal objective is based only on the *type-action counts* – the aggregate number of times each action was taken over the entire horizon.

For such settings, we provide a unified algorithm which we call *Online Clairvoyant Optimum Pursuit* (OCOP). In our main result we prove that under weak assumptions on the regularity of the arrival processes and the geometry of the objective function, OCOP obtains an additive loss guarantee (compared to the hindsight optimal solution with exact knowledge of arrivals) that is independent of  $T$ . Our conditions, in particular, are fulfilled by iid arrival processes and

Regime	Distribution-oblivious	Known Distribution	OCOP
BW	$O(\log(T))$ [2]	$O(1)$ [2]	$O(1)$
PP	$O(\sqrt{T})$ [2]	$O(\sqrt{T})$ [2]	$O(1)$
LW	$O(\sqrt{T})$ [3]	$O(\sqrt{T})$ [2]	$O(1)$

**Table 1: Improvement of OCOP over state-of-the art guarantees for known distributions.**

piece-wise linear objective functions subject to a linear polytope, which can be used to encode various combinatorial problems.

*Applications.* We demonstrate how our general algorithm and conditions translate to specific problems. In particular, we show that our conditions for uniform loss hold in two general settings: (i) where the underlying objective is piecewise-linear with a fixed number of pieces, and (ii) where the objective function admits a class of offline index policies, i.e., policies based on computing an index for each type-action pair that is independent of the other types, and then sorting the types based on these indices. The most striking application of our results is to *online stochastic bin packing*. In this well-studied problem we are able to reduce the suboptimality gap from  $\theta(\sqrt{T})$  to  $\theta(1)$ , while generalizing beyond the assumption of iid arrivals. Further, our results work with just the first moment of the distribution, i.e., though our approach is not distribution-oblivious, it also does not require full distributional information.

Traditionally, the bin packing literature has measured its objective as *waste*, i.e., the amount of empty bin space across all bins opened by an algorithm. It has long been known that in the case of discrete item-size distributions occurring iid, the amount of waste of an optimal packing (i.e., the waste the clairvoyant would incur) scales either as  $\theta(T)$ ,  $\theta(\sqrt{T})$  or  $\theta(1)$  over a long time horizon – these regimes are referred to as linear waste (LW), perfectly packable (PP), and bounded waste (BW). Noticeably, even with LW it is possible to have the additional waste of an algorithm relative to the optimal packing scale as  $\theta(1)$ ; this is exactly the guarantee we provide, and it holds regardless of regime. In contrast, state-of-the-art algorithms in the literature so far had only guaranteed  $\theta(1)$  loss bounds for the bounded waste regime. We summarize the known results in Table 1, highlighting the difference between *distribution-oblivious* algorithms and ones that require knowledge of the arrival distributions (such as ours).

*Full paper.* A full version of this paper can be found at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3479189](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3479189).

*Acknowledgments.* The first author gratefully acknowledges support through NSF grants ECCS-1847393 and DMS-1839346 as well as ARL grant W911NF-17-1-0094.

## REFERENCES

- [1] Alessandro Arlotto and Itai Gurvich. Uniformly bounded regret in the multisecretary problem. *Stochastic Systems*, 2019.
- [2] Janos Csirik, David S Johnson, Claire Kenyon, James B Orlin, Peter W Shor, and Richard R Weber. On the sum-of-squares algorithm for bin packing. *Journal of the ACM (JACM)*, 53(1):1–65, 2006.
- [3] Varun Gupta and Ana Radovanovic. Lagrangian-based online stochastic bin packing. In *ACM SIGMETRICS Performance Evaluation Review*, volume 43, pages 467–468. ACM, 2015.
- [4] Alberto Vera and Siddhartha Banerjee. The bayesian prophet: A low-regret framework for online decision making. In *Abstracts of the 2019 SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems*, pages 81–82. ACM, 2019.