

Controlling Tail Risk in Online Ski-Rental*

Michael Dinitz[†] Sungjin Im[‡] Thomas Lavastida[§] Benjamin Moseley[¶]
Sergei Vassilvitskii^{||}

Abstract

The classical ski-rental problem admits a textbook 2-competitive deterministic algorithm, and a simple randomized algorithm that is $e/(e-1)$ -competitive in expectation. The randomized algorithm, while optimal in expectation, has a large variance in its performance: it has more than a 37% chance of competitive ratio exceeding 2, and the change of the competitive ratio exceeding n is $\Theta(1/n)$!

We ask what happens to the optimal solution if we insist that the *tail risk*, i.e., the chance of the competitive ratio exceeding a specific value, is bounded by some constant δ . We find that this additional modification significantly changes the structure of the optimal solution. The probability of purchasing skis on a given day becomes non-monotone, discontinuous, and arbitrarily large (for sufficiently small tail risk δ and large purchase cost n).

1 Introduction

Decision-making under uncertainty about the future is a central topic in algorithm design; online algorithms, studied through the metric of competitive analysis, have been successful in guaranteeing worst-case performance against adversarial inputs. Arguably, the most basic online problem is the *Ski Rental* problem, which captures a commonly faced sub-problem, usually known as “rent or buy”: we need to decide whether to stay in the current state, paying some cost per time unit, or switch to another state, which is expensive but requires no further payment. In the specific ski rental problem, every morning Alice must decide whether to rent skis for \$1 or buy them for \$ n , in which case she never needs to rent them again. Her choice is non-obvious because she does not know the number of days, denoted as x , she is going to ski—the weather may become too warm, she may get injured, or may just get tired of the sport.

Folklore analysis says that committing to buying skis on the morning of day n is (deterministically) optimal, as Alice never over-spends by more than a factor of 2 (i.e., this approach has a competitive ratio bounded by 2), no matter how many days she ends up skiing. If Alice is willing to randomize, she can do even better—she can commit to buying the skis on day $i \in [n]$ with probability proportional to $\exp(i/n)$. This method gives the best possible $e/(e-1) \approx 1.58$ competitive ratio in expectation [6].

The competitive ratio is for the worst-case action by the adversary (who decides on the number of skiing days), it only holds *in expectation* for Alice. An easy calculation shows that an adversary that ends the ski season on day $n/2$ ensures that Alice exceeds the competitive ratio of 2 with probability $(\sqrt{e} - 1)/(e - 1) \approx 37\%$.

Thus more than a third of the time, Alice is better off following the deterministic strategy. Furthermore, the competitive ratio is $\Omega(n)$ with probability $\Omega(1/n)$ ¹, illustrating that the “best” randomized algorithm has a considerable chance of returning a solution significantly worse than the deterministic alternative. Importantly, in

*A full version of the paper can be accessed at <https://arxiv.org/abs/2308.05067>

[†]Department of Computer Science, Johns Hopkins University, Baltimore, MD. mdinitz@cs.jhu.edu. Supported in part by NSF grants CCF-1909111 and CCF-2228995. Work partially done while a Visiting Researcher at Google Research New York, NY.

[‡]Electrical Engineering and Computer Science, University of California, 5200 N. Lake Road, Merced CA 95344. sim3@ucmerced.edu. Supported in part by NSF grants CCF-1844939 and CCF-2121745.

[§]Jindal School of Management, University of Texas at Dallas, Richardson, TX. thomas.lavastida@utdallas.edu.

[¶]Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA. moseleyb@andrew.cmu.edu. Work supported in part by a Google Research Award, an Infor Research Award, a Carnegie Bosch Junior Faculty Chair and NSF Grants CCF-2121744 and CCF-1845146.

^{||}Google Research New York, NY. sergeiv@google.com.

¹This can occur when $x = \Theta(1)$, yet Alice ends up buying skis.

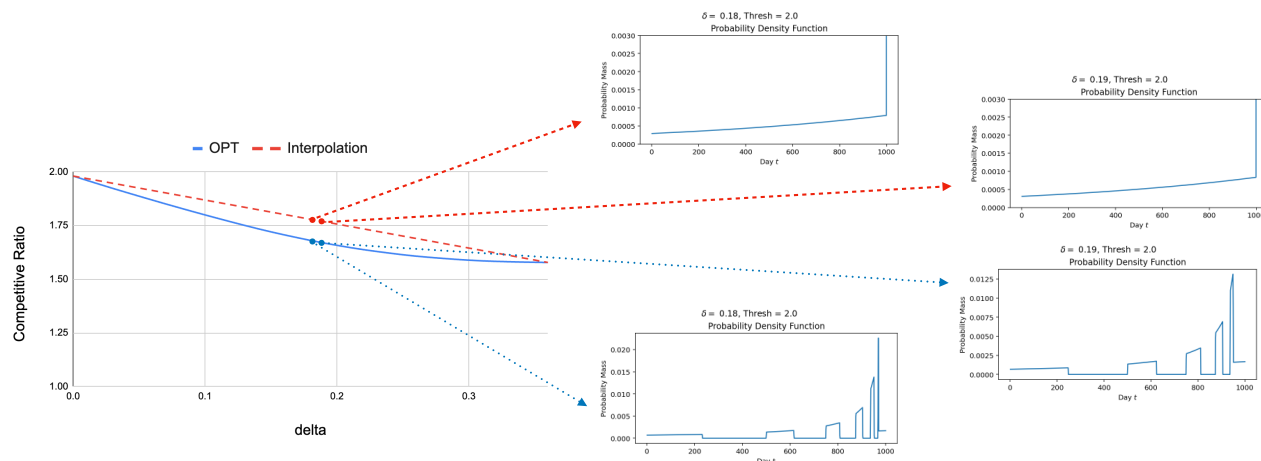


Figure 1: Figure of the competitive ratio as a function of δ for both the optimal strategy and the interpolation strategy. For fixed values of $\delta = \{0.18, 0.19\}$ we show the probability of purchasing skis on day $t \in [1000]$ for the two algorithms. Observe that the interpolation algorithm (red dashed line) puts a weight of $1 - \delta/\delta^* \approx 1/2$ on the last day in both cases. The optimal algorithm distributes the weight across the days, but does so in a non-continuous manner. For instance, observe that the figure for $\delta = 0.18$ has six non-zero intervals, whereas $\delta = 0.19$ has five non-zero intervals.

the online setting decisions are irrevocable, so results in expectation do not immediately lead to high-probability bounds. This is in contrast to an offline algorithm where bounding the approximation ratio in expectation often leads to giving the same bound with high probability by running the algorithm a logarithmic number of times independently and taking the best solution.

What if one desires upper-case bounds on the chance the randomized algorithm is worse than the deterministic algorithm? A natural direction is to find the optimal algorithm and study its competitive ratio as a function of the probability of the competitive ratio exceeding 2. We have the two endpoints—the deterministic algorithm with a ratio of 2, and the randomized algorithm with an *expected* ratio of $e/(e-1)$ and a probability of $(\sqrt{e}-1)/(e-1)$ of exceeding 2. The question we study in this work is what happens in between? What is possible if we put nontrivial constraints on the tail performance, and what is the structure of these optimal solutions?

To formalize the question we study, let a (γ, δ) -tail constraint denote the restriction that the probability (over the choices made by the algorithm) that the worst-case competitive ratio exceeds γ is at most δ . Let $\delta^* = (\sqrt{e}-1)/(e-1)$. For the ski rental problem, the deterministic algorithm is optimal and satisfies a $(2, 0)$ -tail constraint, whereas the randomized algorithm optimizes the competitive ratio among all of the algorithms satisfying a $(2, \delta^*)$ -tail constraint, or more generally, a $(2, \delta)$ -tail constraint for any $\delta \geq \delta^*$.

Given a collection of arbitrary (γ, δ) constraints, or even a single nontrivial tail constraint, what is the optimal algorithm? One may suspect that the solution is well-structured with behavior that is similar to the classical deterministic and randomized solutions. For example, suppose that we are given a single $(2, \delta)$ -tail constraint with $\delta < \delta^*$. It is not hard to see that one way of achieving this is to interpolate between the two classical solutions: with probability δ/δ^* we run the classical randomized algorithm, and with probability $1 - \delta/\delta^*$ we run the classical deterministic algorithm. This clearly satisfies the tail constraint, is a very simple algorithm, and inherits a number of nice properties (monotonically increasing probabilities which increase only exponentially, a single discontinuity at day n , non-zero probability on every day, etc.). Is this interpolation optimal? If not, does the optimal solution at least have these nice properties, or look “simple”?

1.1 Our Contribution. The answer to these questions is an emphatic *no*. Not only is the optimal solution *not* an interpolation between the classical algorithms, its structure is wildly different from any previously considered ski rental algorithm and does not satisfy any of the “nice” properties mentioned earlier. In particular, we show that the optimum solution has the following surprising properties:

- **Non-monotonicity.** The purchase probability does not grow monotonically in time t .
- **Arbitrarily many discontinuities.** Even for a single (γ, δ) tail constraint, the purchase probability becomes zero and then positive $\Omega(1/\delta)$ times for any sufficiently large n compared to any fixed sufficiently small δ .
- **Arbitrarily fast-growing purchase probability.** Again, even for a single (γ, δ) tail constraint, when the probability becomes positive, it grows much faster than before—the exponent doubles each time. This results in continuous sections of the purchase distribution with arbitrarily fast growth.

To illustrate these points, we show the interpolation and the optimal competitive ratios as a function of δ in Figure 1. To add, we show the purchase distributions for two nearby points (namely $\delta = 0.18$ and $\delta = 0.19$) for both cases. Note how the solution, while structured enough to reason about, is not in any way “smooth”. For example, even for these very close δ , the number of nonzero regions is distinct—there is a “discontinuity” in how the optimal solution behaves as δ changes from 0.19 to 0.18. This is despite the fact that the achieved competitive ratio *is* smoothly changing with δ , as is apparent from Figure 1.

Proving these properties of the optimal solution is our main technical contribution. The formal statements are quite complex, but can be found in Section 4, particularly, Lemma 4.6 and Corollary 4.1. As a side effect of our techniques we can also design an algorithm to actually compute the optimal solution (see Section 3.2).

We would like to emphasize that ski rental is traditionally considered to be an extraordinarily *simple* setting for online algorithms. The fact that adding a single tail constraint (which as discussed is something we naturally want in an online setting where we cannot run multiple times) results in such a complex and nonobvious solution structure is, in our opinion, extremely surprising. We hope that this opens up a new set of questions on tail bounds in online algorithms.

1.1.1 Technical Overview. In order to introduce our techniques, we first need to introduce some notation. Consider the following characterization of any randomized algorithm for ski rental. Prior to skiing on the first day, Alice flips a coin and commits to buying skis on the morning of day i with probability f_i . Note the f_i ’s form a distribution, i.e., $\sum_i f_i = 1$. We call $f = \{f_i\}_{i \in [n]}$, the *purchase distribution*.

The purchase distribution, along with the adversary’s choice of the last skiing day x , induces the competitive ratio, which itself is a random variable. We denote the *competitiveness* function α_f , where $\alpha_f(x)$ is the *expected* competitive ratio when Alice stops skiing at the end of day x and chooses her purchase day from the distribution f .

Let f^* be the randomized algorithm minimizing the expected competitive ratio, i.e., minimizing $\max_x \alpha_{f^*}(x)$. It is known that this algorithm sets $f_i^* \propto \exp(i)$. The optimal choice of f^* balances two competing objectives: buying early is good if the number of skiing days exceeds n , on the other hand, buying late is good if the adversary’s choice for the number of skiing days is small. The balance is done in the worst-case over all adversary’s choices, but in expectation over the random coins flipped by the algorithm. In the optimal solution the balance has the effect that the competitiveness function $\alpha_{f^*}(x) = e/e-1$, that is, it is constant over all $x \in [n]$. Thus, no matter what day the adversary picks as the last ski day, Alice will have the exact same expected competitive ratio of $e/e-1$.

Optimal Solution Characterization. In order to understand the structure of the optimal purchase distribution in the presence of tail bounds, we first need to give a characterization of this optimal distribution. It turns out that this characterization will naturally lead to an efficient algorithm to *construct* the optimal purchase distribution under any combination of tail constraints, although there are some technical complications which have to be overcome along the way (see Section 3.2 for details). While our characterization and algorithm holds for arbitrary combinations of tail constraints, for simplicity we will now assume that there is only a single (γ, δ) -tail constraint.

For every day x , if we condition on the adversary choosing x to be the final day of skiing, then there are essentially two constraints which the optimal purchase distribution f must obey: i) *competitiveness constraint*: the expected competitive ratio using f must be at most the optimal competitive ratio opt , i.e., $\alpha_f(x) \leq \text{opt}$; and ii) *tail constraint*: the probability when we choose a purchase day from f that we achieve competitive ratio worse than γ must be at most δ (for a formal version of this in terms of “bad intervals”, see Section 3).

These constraints are both inequalities, but our main characterization theorem (see Theorem 3.1 in Section 3.1) is that for every x , *at least one of these is tight*. That is, for every day x , when we condition

on the adversary picking day x as the final day, *either* the competitive ratio of f is equal to $\mathbf{opt}$ *or* the probability of f achieving competitive ratio worse than γ is equal to δ .

To gain some intuition for this, recall that if we do not have tail constraints, then in the optimal purchase distribution the competitive ratio is exactly the same, $(e/e-1)$, for every time x which the adversary might choose. This is no longer true in the presence of tail constraints, but if there is some day when the competitive ratio is strictly less than the optimal competitive ratio, then it can only be because the tail bound is tight on that day.

This structure theorem leads to solutions with surprisingly complex structure. But it also means that this complex structure is purely a function of which constraint is tight at which days. This is also (at a very high level) how our algorithm works: we guess $\mathbf{opt}$ (which is non-obvious, see the full version of the paper for more discussion), and then can iteratively set the probability for each day to be whatever makes one of the two constraints tight. Examples of this theorem “in action” can be found in the full version of the paper. The days in which the expected competitive ratio dips below $\mathbf{opt}$ are precisely the days where the tail constraint is tight, and these transitions also obviously correspond to transitions in the purchase distribution.

Notably, while this characterization allows us to find the optimal purchase distribution algorithmically, the exact value of the purchase distribution may not have a closed form with elementary functions. In the full version of the paper we show that the exact solution involves the Lambert W function.

Single Tail Constraint. With the characterization of the optimal solution in place, we consider the solution in the presence of a single (γ, δ) tail constraint. Here the pair of invariants we described above leads to a solution with a non-trivial shape, which we illustrate in the bottom panes of Figure 1. Observe that the purchase distribution, f , oscillates between periods of 0 probability of buying, together with periods of ever higher probability of buying.

Although we do not compute the exact closed form of the competitive ratio due to the difficulty mentioned above, we characterize the optimum purchase probability distribution quite precisely when t is small—including exactly the times when the probability distribution drops to 0 and the exact exponentially growing exponents.

As the optimum solution for the discrete version of ski-rental problem is technically complex, to make our presentation transparent, here we only present our result for the continuous setting. In the continuous setting, by scaling we can assume without loss of generality that Alice can buy skis for \$1 and rent for \$ dt per dt time; she will ski for x time where x is a fractional value in $(0, 1]$. Thus, the randomized algorithm can be described as a purchase probability distribution $f(t)$ over $(0, 1]$; see Section 2 for more details. The following theorem reveals the optimum solution’s structure.

THEOREM 1.1. (Corollary 4.2) *Let $\gamma \geq 2$ be an integer. Let $\{P_j\}_{j \geq 0}$ be a partition of $(0, 1/(\gamma - 1)]$, defined as $P_0 = (0, \frac{1}{\gamma}]$ and $P_j = (\frac{1}{\gamma-1} \cdot (1 - \frac{1}{\gamma^j}), \frac{1}{\gamma-1} \cdot (1 - \frac{1}{\gamma^{j+1}})]$ for all $j \geq 1$. Let ℓ_j be the start point of P_j . In the continuous version of the ski rental problem with a single tail constraint (γ, δ) , the optimum solution $f(t)$ has the following structure in every P_j such that $j < 1/(2\gamma\delta) - 1$:*

- $f(t) = c_j \cdot e^{\gamma^j t}$ for t in the interval $(\ell_j, \ell_j + \hat{t}_j]$, and
- $f(t) = 0$ for all $t \in (\ell_j + \hat{t}_j, \ell_{j+1}]$,

where c_j and $\hat{t}_j < \ell_{j+1} - \ell_j$ are some constants depending on δ and j .

Let’s assume $\gamma = 2$ to illustrate the theorem. The theorem shows that with a single tail constraint $(2, \delta)$, the optimum solution has certain recurring structures over time intervals, $P_0, P_1, \dots$, which are of exponentially decreasing lengths by a factor of 2. Here, note that we can only show the structure for early intervals. For example, if $\delta < 1/12$, we can analytically show the recurring structure over $P_0 = (0, 1/2]$, $P_1 = (1/2, 3/4]$ and $P_2 = (3/4, 7/8]$. In such an interval P_j , $f(t)$ initially grows in proportion to $e^{2^j t}$ and then drops to 0 and remains 0 until the next interval P_{j+1} starts.

We sketch the analysis of this result. We begin with a key observation that the probability mass of f on each P_j must be at most δ , which follows from taking a close look at the tail constraint. More precisely, the tail constraint can be shown to be equivalent to the probability mass on each interval $I(x)$ ending at x being at most δ . It turns out that $P_0, P_1, \dots$ are such intervals.

We show that in beginning of P_0 , the competitiveness constraint must be tight, thus f grows in proportion to e^t and then at some point to respect the tail constraint it drops to 0. Afterwards, we show that the tail constraint

must be tight. Since f must restart collecting probability mass from the beginning of each P_j , which is shorter than the previous interval P_{j-1} , it has to accelerate the collection.

The actual analysis is based on a careful induction argument, but this sketches the high-level proof idea. We note that in the discrete version we can show that the number of times f drops to 0 is at least $\Omega(1/\delta)$ (for large enough n). See Section 4 for more details.

Single Pure Tail Constraint. As a side effect of our techniques, we are actually able to give a precise and explicit description of the optimal solution when there is a single tail constraint of the form $(\gamma, 0)$, i.e., the competitive ratio is never allowed to exceed γ . We call these *pure* tail constraints. A simple calculation shows that such a constraint implies that $f_i = 0$ for $i < n/(\gamma - 1)$: we must always rent for the first $n/(\gamma - 1)$ days. Since this is the only tail constraint, the remaining question is how to allocate the probability mass such that maximum of the competitiveness function on the remaining interval is minimized. In this case we can precisely show the value of the optimal competitive ratio and the shape of the optimal solution as follows (see the theorems in Section 5 for more precise statements).

THEOREM 1.2. *In the ski-rental problem with tail constraint $(\gamma, 0)$, the optimum solution is the following (assuming $\gamma - 1$ divides $n - 1$):*

$$f_t = \begin{cases} 0 & t < \frac{n-1}{\gamma-1} \\ \frac{\lambda-1}{\gamma} - 1 & t = \frac{n-1}{\gamma-1} \\ \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} \cdot \left(1 + \frac{1}{n-1}\right)^{t-t_1-1} & t > \frac{n-1}{\gamma-1} \end{cases}$$

where the optimum competitive ratio λ tends to $1 + \frac{\gamma-1}{1+\gamma\left(e^{\frac{1}{1-\frac{1}{\gamma-1}}}-1\right)}$ as $n \rightarrow \infty$.

Note that when $\gamma = 2$ we get that $\lim_{n \rightarrow \infty} \lambda = 2$, i.e., we recover the classical deterministic bound, and when $\gamma \rightarrow \infty$ we get that $\lim_{n \rightarrow \infty} \lambda = 1 + \frac{1}{e-1} = \frac{e}{e-1}$, i.e., we recover the classical randomized bound.

1.2 Related Work

Ski-Rental and its Variants. The classical ski-rental problem captures the fundamental “rent or buy” dilemma that exists at the heart of many online problems. The deterministic 2-competitive break-even strategy was first analyzed by Karlin et al. [6] as a special case of Snoopy Caching and later the optimal $e/(e-1)$ -competitive randomized strategy was described in [5]. We note that the online primal-dual approach was used to give an optimal $e/(e-1)$ -competitive algorithm [3]. In separate work, Karlin et al. [4] gave optimal $e/(e-1)$ -competitive algorithms for dynamic TCP acknowledgment and the Bahncard problem by exploiting a connection to the classic ski-rental problem. Here, we expand the classical ski-rental problem to include the tail of the competitive ratio as a new metric in addition to the expected competitive ratio. Even with this small change, we find that the optimal algorithm exhibits interesting behaviour that is not observed in the standard setting.

There are several lines of work concerned with generalizing and analyzing new variants of the classical ski-rental problem. Motivated by applications in cloud cost optimization, Khanafer et al. [7] considered a variant of ski-rental where the number of days is randomly chosen from a distribution with known first or second moments (but otherwise unspecified). Wang et al. [16] consider a variant of ski-rental with multiple commodities that can be rented, purchased individually, or purchased as a group. Recently there has been significant interest in incorporating predictions derived from machine learning into online algorithms [8] which has resulted in a sequence of work applying this to the ski-rental problem [1, 2, 11, 13–15, 18].

Online Algorithms Beyond Expectation. It is standard to analyze the expected performance of algorithms in several contexts, such as minimizing the competitive ratio of an online algorithm or the regret of a policy in a multi-arm bandit setting. As we point out for the ski-rental problem, there may be cases where tailoring an algorithm to the mean may not be sufficient and other metrics such as the tail may be important. As another example, in multi-arm bandits achieving low regret on average necessitates a certain amount of exploration which can increase the variability in the attained reward. Wu et al. [17] consider the conservative bandits setting in which the goal is to minimize the cumulative regret of the policy while constraining the total reward earned at each time to be above some baseline level. Another example is the survival bandit problem [10, 12] in which rewards can be both positive and negative and the objective is to maximize the total reward while keeping the “risk of ruin”—the probability of the current total reward going below a fixed threshold—small.

2 Preliminaries

In the (discrete version of) ski rental problem, also a.k.a. rent-or-buy, every morning Alice must decide whether to rent skis for \$1 or buy them for \$ n , in which case she never needs to rent them again. The number of days (times), denoted as x , she will ski is unknown a priori, and it is revealed to Alice only at the end of day x .

Although Alice is allowed to rent beyond day n , a moment's thought shows that she wouldn't benefit it. Thus, without loss of generality, we can describe any algorithm by its purchase distribution $f = \{f_t\}_{t \in [n]}$. For $x, t \in [n]$, let $\alpha(t, x)$ denote the competitive ratio of the algorithm when the adversary chooses for x to be the last skiing day and Alice buys skis on the morning of day t . We have that,

$$\alpha(t, x) = \begin{cases} 1 & \text{if } t > x \\ \frac{n+t-1}{x} & \text{if } t \leq x \end{cases}$$

Given a distribution f over $[n]$, let $\alpha_f(x)$ be the expectation of $\alpha(t, x)$ where t is drawn from f . So we have that

$$(2.1) \quad \alpha_f(x) = \sum_{t \in [n]} \alpha(t, x) f_t = \sum_{t \leq x} \frac{n+t-1}{x} f_t + \sum_{t > x} f_t = \sum_{t \leq x} \frac{n+t-1}{x} f_t + 1 - \sum_{t \leq x} f_t$$

The following observations are immediate from the definition.

OBSERVATION 2.1. *Given a purchase distribution f , $\alpha_f(t)$ only depends on $f_1, f_2, \dots, f_t$. If we move the probability mass from a later time to an earlier time, $\alpha_f(n)$ decreases.*

It is well known that the deterministic algorithm that buys on day n , i.e., $f_1 = \dots = f_{n-1} = 0$ and $f_n = 1$, is $2 - 1/n$ -competitive and it is the best possible competitive ratio achievable for any deterministic algorithms. Also, when $f_t = (1/n)(1 + 1/n)^{t-1} / ((1 + 1/n)^n - 1)$ we recover the celebrated randomized algorithm whose competitive ratio is $(1 + 1/n)^n / ((1 + 1/n)^n - 1)$, which tends to $e/(e - 1)$ as $n \rightarrow \infty$. An easy calculation shows $\alpha_f(x)$ is a constant which is exactly the claimed competitive ratio and in fact keeping $\alpha_f(x)$ constant for all x makes the purchase probability grows exponentially in t .

Continuous Case. The ski rental problem is often discussed in the continuous setting as it exhibits a cleaner representation of the optimum solution; for example, see [4, 9]. More precisely, after scaling, we can assume wlog that Alice can ski for x amount of time where $x \in (0, 1]$ and the rental price for dt amount of time is dt and the purchase price is 1. Then, the optimum randomized algorithm has pdf $f(t) = e^t/(e - 1)$ and the competitive ratio is exactly $e/(e - 1)$. In our problem with tail constraints, we will mostly consider the discrete version as it seems to resist a simple closed form for the competitive ratio unlike the problem without tail constraints.

A Useful Lemma. The following lemma (proof in the full version) will prove to be useful in a few different places in our analysis. To interpret it, first note that thanks to Equation (2.1), the competitive ratio at some time x can be calculated using only the probabilities for times $t \leq x$. So the following lemma lets us say that if we have built “part” of the distribution, then we can extend it so that the competitive ratio stays constant by increasing each successive probability by a multiplicative $(1 + \frac{1}{n-1})$ factor. In other words, an appropriate exponential function keeps the competitive ratio constant.

LEMMA 2.1. *Let $1 < a < n$, let $f : [a - 1] \rightarrow [0, 1]$ such that $\sum_{t=1}^{a-1} f_t < 1$, and let $a \leq x' \leq n$. Then $\alpha_f(x) = \alpha_f(a - 1)$ for all $a \leq x \leq x'$ if and only if a) $f_x = \left(1 + \frac{1}{n-1}\right)^{x-a} f_a$ for all $a \leq x \leq x'$, where $f_a = \frac{1}{(a-1)(n-1)} \sum_{t=1}^{a-1} (n+t-1) f_t$, and b) $\sum_{t=1}^{x'} f_t \leq 1$.*

3 Characterizing the Optimal Solution

In this section, we consider the general case where there can be multiple tail constraints, each with an arbitrary threshold at least 2. Our goal is to prove a characterization of the optimal purchase distribution that, while not necessarily allowing us to write it explicitly, will enable us to reason about its properties. Moreover, a side effect of our characterization will be an efficient algorithm to actually construct the optimal purchase distribution.

Suppose that we are given a collection $\{(\gamma_i, \delta_i)\}_{i \in [k]}$ where $\gamma_i \geq 2 - 1/n$ and $0 \leq \delta_i < 1$ for all $i \in [k]$ ². The goal is to find the randomized algorithm with minimum expected competitive ratio (worst case over times the adversary might choose) subject to the requirements that for all $i \in [k]$, the probability that our algorithm has a competitive ratio larger than γ_i is at most δ_i , where again this is the worst-case over times the adversary might choose. The optimal solution's properties we will characterize for general tail constraints will be useful for discovering some surprising structural properties of the optimal solution later for a single tail constraint.

DEFINITION 3.1. *The bad interval for threshold γ and time x , which we denote by $I_\gamma(x)$, consists of all $t \in [x]$ such that if we buy at time t and the adversary chooses time x , then our competitive ratio is larger than γ . By definition,*

$$I_\gamma(x) = \{t \in [n] : \alpha(t, x) > \gamma\} = \left\{t \in [x] : \frac{n+t-1}{x} > \gamma\right\} = \{t : \max(0, \gamma x - n + 1) < t \leq x\}$$

OBSERVATION 3.1. *For all $\gamma \geq 2 - 1/n$ and all x , it is the case that $n \notin I_\gamma(x)$.*

We say that a purchase distribution f is feasible if it satisfies all tail constraints. The following lemma shows the solution is feasible if and only if it satisfies a collection of packing constraints pertaining to the bad intervals. The proof, as well as all the other missing proofs, are deferred to the full version of the paper.

LEMMA 3.1. *A distribution f is feasible if and only if $\sum_{t \in I_{\gamma_i}(x)} f_t \leq \delta_i$ for all $i \in [k]$ and for all $x \in [n]$.*

3.1 Main Characterization Theorem. Let f^* be the optimal solution. By Lemma 3.1, this means that f^* is the distribution f minimizing $\max_{x \in [n]} \alpha_f(x)$ subject to $\sum_{t \in I_{\gamma_i}(x)} f_t \leq \delta_i$ for all $i \in [k]$ and for all $x \in [n]$. Let $\text{opt} = \max_{x \in [n]} \alpha_{f^*}(x)$.

We now prove some useful properties about f^* and opt . First we show that $\text{opt} < 2 - 1/n$ whenever the set of tail bounds is $\{(\gamma_i, \delta_i)\}_{i \in [k]}$ with $\gamma_i \geq 2 - 1/n$. Since the $2 - 1/n$ -competitive deterministic algorithm satisfies all tails constraints, it must be the case that $\text{opt} \leq 2 - 1/n$. Intuitively, randomization should yield a better competitive ratio and the following lemma formally proves it.

LEMMA 3.2. *If $\gamma_i \geq 2 - 1/n$ for all $i \in [k]$ then $\text{opt} < 2 - 1/n$.*

By definition of opt , it is immediate that $\alpha_{f^*}(x) \leq \text{opt}$ for any adversarial choice of x . The following lemma shows that the competitiveness function α is maximized on the last day n for the optimum distribution f^* . The proof easily follows from Observations 2.1 and 3.1: If the lemma were false, we can move the probability mass from an earlier time to time n , which keeps the solution feasible while increasing $\alpha_{f^*}(n)$.

LEMMA 3.3. *Let f^* be an optimal solution. Then $\alpha_{f^*}(n) = \text{opt}$.*

Now we can prove the main structural theorem that shows that for any time x the optimum solution must have the maximum competitiveness function value or the bad time interval packing constraint at the time must be tight. We show the theorem by showing that if the theorem were false at time t_1 , then we can move a probability mass to t_1 from a later time, thereby improving the competitive ratio.

THEOREM 3.1. *Let f^* be an optimal solution. Then for every $x \in [n]$, at least one of the following is true:*

- (competitiveness constraint) $\alpha_{f^*}(x) = \text{opt}$, or
- (tail constraint) $\sum_{t \in I_{\gamma_i}(x)} f_t^* = \delta_i$ for some $i \in [k]$

²Under this assumption we have that f_n does not affect any tail constraints, see Definition 3.1 and Observation 3.1. We focus on the case that $\gamma_i \geq 2 - 1/n$ since $2 - 1/n$ is the competitive ratio that is achieved by a deterministic algorithm, but note that smaller values of γ_i can be handled so long as δ_i is large enough for there to exist a feasible solution.

Proof. Suppose for contradiction that this is false. Let t_1 be the smallest value for which both conditions are false, and observe that Lemma 3.3 implies that $t_1 < n$. Let $t_2 > t_1$ be the smallest value larger than t_1 such that at least one of the two conditions holds, and again observe that Lemma 3.3 implies that such a t_2 must exist. It is also easy to see that $f_{t_2}^* > 0$, since otherwise at least one of the conditions must hold at $t_2 - 1$. More formally, suppose $f_{t_2}^* = 0$ for the sake of contradiction. By definition of t_2 , we have $\alpha_{f^*}(t_2 - 1) < \text{opt}$. It is easy to verify $\alpha_{f^*}(t_2 - 1) > \alpha_{f^*}(t_2)$; thus we have $\alpha_{f^*}(t_2) < \text{opt}$. Again by definition of t_2 , we know $\sum_{t \in I_{\gamma_i}(t_2-1)} f_t^* < \delta_i$ for any $i \in [n]$. Using the facts that $I_{\gamma_i}(t_2) \subseteq I_{\gamma_i}(t_2 - 1) \cup \{t_2\}$ and $f_{t_2}^* = 0$, we have $\sum_{t \in I_{\gamma_i}(t_2)} f_t^* \leq \sum_{t \in I_{\gamma_i}(t_2-1)} f_t^* < \delta_i$. This implies none of the conditions hold true at time t_2 , which is a contradiction.

Therefore, we can define the distribution that we get by shifting some very small $\epsilon > 0$ mass from t_2 to t_1 , i.e., $f_t = f_t^* + \epsilon$ if $t = t_1$; $f_t = f_t^* - \epsilon$ if $t = t_2$; and $f_t = f_t^*$ otherwise. Obviously f is still a probability distribution on $[n]$. Moreover, we claim that it is still a feasible solution if we choose a small enough ϵ . To see this, let $B = \{1, 2, \dots, t_1 - 1\}$, let $M = \{t_1, t_1 + 1, \dots, t_2 - 1\}$, and let $A = \{t_2, t_2 + 1, \dots, n\}$. For all $x \in B$, we have that $\sum_{t \in I_{\gamma_i}(x)} f_t = \sum_{t \in I_{\gamma_i}(x)} f_t^* \leq \delta_i$ for all $i \in [k]$, as required.

For $x \in M$, the mass in $I_{\gamma_i}(x)$ could be larger in f than in f^* (since t_1 could be in their bad interval), but by definition of t_2 this was strictly less than δ in f^* , so by choosing a small enough ϵ we can keep it below δ . Slightly more formally, we have that $\sum_{t \in I_{\gamma_i}(x)} f_t \leq \sum_{t \in I_{\gamma_i}(x)} f_t^* + \epsilon \leq \delta$ for small enough ϵ .

For $x \in A$, note that it is impossible for $\{t_1, t_2\} \cap I_{\gamma_i}(x) = \{t_1\}$ for any $i \in [k]$; this is straightforward from the definition of $I_{\gamma_i}(x)$. So there are three cases.

1. If $\{t_1, t_2\} \cap I_{\gamma_i}(x) = \{t_1, t_2\}$, then $\sum_{t \in I_{\gamma_i}(x)} f_t = \sum_{t \in I_{\gamma_i}(x)} f_t^* \leq \delta$ since f^* is feasible.
2. If $\{t_1, t_2\} \cap I_{\gamma_i}(x) = \{t_2\}$, then $\sum_{t \in I_{\gamma_i}(x)} f_t < \sum_{t \in I_{\gamma_i}(x)} f_t^* \leq \delta$.
3. If $\{t_1, t_2\} \cap I_{\gamma_i}(x) = \emptyset$, then $\sum_{t \in I_{\gamma_i}(x)} f_t = \sum_{t \in I_{\gamma_i}(x)} f_t^* \leq \delta$.

Hence f is feasible by Lemma 3.1.

Now let's consider the competitive ratios $\alpha_f(x)$. We break into three cases for x .

1. If $x \in B$, then we have that $\alpha_f(x) = \alpha_{f^*}(x) \leq \text{opt}$ (the final inequality is due to the optimality of f^*).
2. If $x \in M$ then the competitive ratio is worse in f than in f^* , i.e., $\alpha_f(x) > \alpha_{f^*}(x)$. But by the definition of M (and t_2) we know that $\alpha_{f^*}(x) < \text{opt}$, so by choosing a small enough ϵ we still have that $\alpha_f(x) \leq \text{opt}$.
3. If $x \in A$, then it is not hard to see that the competitive ratio decreases, i.e., $\alpha_f(x) \leq \alpha_{f^*}(x) \leq \text{opt}$.

Thus, we have a feasible solution f with $\max_{x \in [n]} \alpha_f(x) \leq \text{opt}$, so f is actually optimal. But since $n \in A$, we know that $\alpha_f(n) < \alpha_{f^*}(n) \leq \text{opt}$. This contradicts Lemma 3.3. Hence no such t_1 can exist, which implies the theorem. $\square$

3.2 Algorithm and Analysis. With Theorem 3.1 in hand, we can now give an algorithm to compute the optimal solution. Intuitively, since Theorem 3.1 says that for every day i either the competitiveness constraint or the tail constraint is tight, we can just iterate through the days, increasing the probability for day j until some constraint becomes tight.

We now formalize this. We are given n and $\{(\gamma_i, \delta_i)\}_{i \in [k]}$. As before, we will be concerned with the case where $\delta_i < 1$ (since larger values of δ_i imply the tail constraint is trivially satisfied) and $\gamma_i \geq 2 - 1/n$ for all i . In what follows we will assume that we have a guess λ which is equal to opt ; we discuss how to remove this assumption in the full version of the paper. Our algorithm is the following.

- Set $f_1 = \min(\min_{i \in [k]} \delta_i, \frac{\lambda-1}{n-1})$.
- For $j = 2$ to n : Set

$$f_j = \min \left(\min_{i \in [k]: j \in I_{\gamma_i}(j)} \left(\delta_i - \sum_{t \in I_{\gamma_i}(j) \setminus \{j\}} f_t \right), \frac{j}{n-1}(\lambda-1) - \sum_{t=1}^{j-1} \left(1 - \frac{j-t}{n-1} \right) f_t \right)$$

We can now prove that this algorithm is optimal. The proof proceeds by showing that an optimum solution must coincide with f to make the competitiveness or tail constraint tight, as required by Theorem 3.1. The proof is deferred to the full version of the paper.

THEOREM 3.2. *The function f returned by our algorithm is the unique optimal solution.*

4 Single Tail Constraint

In this section we consider a single tail constraint, (γ, δ) . To streamline our presentation, we assume γ is an integer, but our analysis should be easily generalizable to any $\gamma \geq 2 - 1/n$.

4.1 Analysis Overview. We start the analysis by defining disjoint intervals $P_0, P_1, P_2, \dots$, where the optimum purchase distribution f_t exhibits a recurring structure (Section 4.2). For $\gamma = 2$ which we assume for the illustration purpose here, the intervals are defined as $P_0 = [1, \frac{1}{2}(n-1)]$, $P_1 = [\frac{1}{2}(n-1) + 1, \frac{3}{4}(n-1)]$, $P_2 = [\frac{3}{4}(n-1) + 1, \frac{7}{8}(n-1)] \dots$. Note that the intervals have exponentially decreasing lengths. The key observation we make in Section 4.2 is that every P_j is a bad interval and therefore the probability mass of f_t on P_j is at most δ (Lemma 4.1).

We show how f_t accumulates δ probability mass in each interval P_j . We first consider P_0 in Section 4.3. Recall that f_t must make the competitive constraint or the tail constraint tight at every time (Theorem 3.1). Obviously, the tail constraint doesn't get tight until f_t accumulates δ mass in the beginning. The time, denoted as $\hat{t}$, is shown to be in the first interval P_0 (Lemma 4.2). Then using the fact that the competitiveness constraint must be satisfied for all $t \leq \hat{t}$, we can show that f_t is roughly proportional to $e^{t/n}$ for all $t \leq \hat{t}$, then drop to 0 because the total probability mass of f must be at most δ in P_0 .

The subsequent intervals P_j , $j \geq 1$, are considered in Section 4.4. Here, the key observation we make is that the tail constraint is satisfied for all $t > \hat{t}$ (Lemma 4.4). Thus we have a sequence of equations and by solving them, we obtain $f_t = f_{2t-(n-1)-1} + f_{2t-(n-1)}$. Here, when $t \in P_j$, the times appearing in the right-hand-side, $2t - (n-1) - 1$ and $2t - (n-1)$ are both in P_{j-1} . Intuitively, this implies that f_t grows twice faster in P_j than in P_{j-1} . In fact, by a careful induction we can precisely show that f_t 's probability mass at time t for $t \in P_j$ must be equal to that of some 2^j consecutive time steps in P_0 (Lemma 4.5). Thus, f_t grows in P_j with 2^j factor larger exponent than it does in P_0 (Corollary 4.1). It then accumulates δ probability mass in P_j and drops to 0 because the probability mass shouldn't exceed δ in P_j (Lemma 4.5).

Finally, we take $n \rightarrow \infty$ in Section 4.5 to obtain a more intuitively looking pdf in the continuous setting (Corollary 4.2). We note that we consider the discrete version for analysis because the recursive argument needs considerable care and it seems easier in the discrete setting.

4.2 Defining Disjoint Intervals. We will assume that $n-1$ is a sufficiently large power of γ , this will make the notation simpler since we will be interested in the optimum solution for sufficiently large values of n .

Let

$$\ell_j = \frac{n-1}{\gamma-1} \left(1 - \frac{1}{\gamma^j}\right) + 1 \quad \text{and} \quad r_j = \frac{n-1}{\gamma-1} \left(1 - \frac{1}{\gamma^{j+1}}\right)$$

We define an interval $P_j = [\ell_j, r_j]$ for $j \geq 0$. Note that $\ell_0 = 1$, and more generally:

$$(4.2) \quad \begin{aligned} P_0 &= \left\{ t : 1 \leq t \leq \frac{n-1}{\gamma} \right\} \\ P_j &= \left\{ t : \frac{(n-1)}{(\gamma-1)} \cdot \frac{(\gamma^j-1)}{\gamma^j} + 1 \leq t \leq \frac{(n-1)}{(\gamma-1)} \cdot \frac{(\gamma^{j+1}-1)}{\gamma^{j+1}} \right\} \end{aligned} \quad \text{for all } j \geq 1$$

We can check that $|P_j| = \frac{n-1}{\gamma^{j+1}}$. Hence these intervals are non-empty for all $j < \log_\gamma n$. Moreover they are disjoint. Simple algebra shows the following.

CLAIM 4.1. *For all $j \geq 1$, we have*

- $\ell_{j-1} = \gamma \ell_j - (n-1) - (\gamma-1)$; and
- $r_{j-1} = \gamma r_j - (n-1)$

Proof. By definition of r_j , we have,

$$\gamma \cdot r_j - n + 1 = \gamma \left(\frac{n-1}{\gamma-1} \left(1 - \frac{1}{\gamma^{j+1}} \right) \right) - (n-1) = \frac{n-1}{\gamma-1} \left(\gamma - \frac{1}{\gamma^j} - (\gamma-1) \right) = \frac{n-1}{\gamma-1} \left(1 - \frac{1}{\gamma^j} \right) = r_{j-1},$$

which proves the second claim. The first claim follows a similar algebra. $\square$

We will first show that the probability mass inside each of them is bounded.

LEMMA 4.1. *For any feasible solution f , we have that $\sum_{t \in P_j} f_t \leq \delta$ for every $j < \log_\gamma n - 1$*

Proof. From the definition of $I_\gamma(x)$, we have that

$$I_\gamma(r_j) = \{t : \gamma r_j - n + 1 < t \leq r_j\} = \{t : r_{j-1} < t \leq r_j\} = P_j$$

where the penultimate equality holds due to Claim 4.1, and the final equality is due to the strict inequality of the lower bound on t . So the tail constraint at r_j implies that $\sum_{t \in P_j} f_t = \sum_{t \in I_\gamma(r_j)} f_t \leq \delta$ due to Theorem 3.1, as claimed. $\square$

4.3 Understanding the Optimum Solution's Structure for $t \in P_0$. We are going to claim that for any fixed $i \geq 0$, by setting δ small enough (as a function of i , not n) the structure inside of every P_j for $j \leq i$ is both simple and surprising: there is a prefix which is a $\exp(\gamma^j)$ function, and then it drops to 0 (at least for n large enough that P_j is nontrivially large). So, in other words, the optimal solution can exhibit an *arbitrarily large number of drops to 0*, and can also exhibit *arbitrarily large growth*! This is in stark contrast to the classical solution, which is a simple exponential function everywhere.

Fix $i \geq 0$, and let $\delta < 1/(2\gamma(i+1))$. Assume n is sufficiently large for now; later we will require $n \geq 2\gamma^{i+1} + 1$. Let $\hat{t}$ be the largest integer such that

$$(4.3) \quad \sum_{t=1}^{\hat{t}} \frac{\lambda-1}{n-1} \left(1 + \frac{1}{n-1} \right)^{t-1} \leq \delta.$$

LEMMA 4.2. *For all sufficiently large n , it is the case that $\hat{t} < n/\gamma$, i.e., $\hat{t} \in P_0$*

Proof. Since the optimum competitive ratio is at least $\frac{e}{e-1}$ with no tail constraints, we have $\lambda \geq \frac{e}{e-1}$. Knowing that the left-hand-side is increasing in $\hat{t}$, it suffices to show that

$$\sum_{t=1}^{n/\gamma} \frac{e/(e-1)-1}{n-1} \left(1 + \frac{1}{n-1} \right)^{t-1} > \frac{1}{2\gamma} \geq \delta.$$

Indeed, for all sufficiently large n , we have,

$$\sum_{t=1}^{n/\gamma} \frac{e/(e-1)-1}{n-1} \left(1 + \frac{1}{n-1} \right)^{t-1} = \frac{1}{e-1} \left(\left(1 + \frac{1}{n-1} \right)^{n/\gamma} - 1 \right) \geq \frac{e^{1/\gamma} - 1}{e-1} \geq \frac{1/\gamma}{e-1} > \frac{1}{2\gamma},$$

which is no smaller than δ by definition. $\square$

Let $\{f_t\}_{t \in [n]}$ be the optimal solution, which we know by Theorem 3.2 is returned by our algorithm. We first show that P_0 consists of an exponentially increasing function, followed by the zero function.

LEMMA 4.3. *$f_t = \frac{\lambda-1}{n-1} \left(1 + \frac{1}{n-1} \right)^{t-1}$ for all $t \leq \hat{t}$, and $f_t = 0$ for all $\hat{t} + 2 \leq t \leq \frac{n-1}{\gamma}$.*

Proof. Let's start with the first part of the lemma, and focus on the $t \leq \hat{t}$ case. For sufficiently large n (as a function of δ), we know from the definition of our algorithm that $f_1 = \frac{\lambda-1}{n-1}$, as claimed, which makes the competitiveness constraint tight for time 1. By Theorem 3.1, the competitiveness constraint will stay tight until

there is δ mass in the bad interval for some t . Lemma 2.1 implies that to keep the competitiveness constraint tight, $f_t = \left(1 + \frac{1}{n-1}\right) f_{t-1}$. Hence $f_t = \frac{\lambda-1}{n-1} \left(1 + \frac{1}{n-1}\right)^{t-1}$ as claimed until we have accumulated δ mass total, which by definition occurs at time $\hat{t} + 1$, implying the first part of the lemma.

For the second part of the lemma, note that by definition of $\hat{t}$ the tail constraint first becomes tight at time $\hat{t} + 1$, i.e., $\sum_{t=1}^{\hat{t}+1} f_t = \delta$. But for all $t \in [\hat{t} + 2, n/\gamma]$, we know that $I_2(t) = [1, t]$. Hence for all $t \in [\hat{t} + 2, n/\gamma]$, we know that $\sum_{t' \in I_2(t) \setminus \{t\}} f_{t'} = \delta$, and hence $f_t = 0$ as claimed. $\square$

4.4 Understanding the Optimum Solution's Structure for $t \in P_j$, $j \geq 1$. We now prove that the tail constraint is tight for a large range of values.

LEMMA 4.4. *For all $t \in \cup_{j:0 \leq j \leq i} P_j \setminus [\hat{t}]$, the tail constraint is tight.*

Proof. For P_0 , this is implied by the proof of Lemma 4.3.

Now consider $1 \leq j \leq i$, and let $t \in P_j$. We know from Lemma 4.1 that there can be at most δ total mass inside of each P_j . This allows us to bound $\alpha_f(t)$ as follows.

$$\begin{aligned}
 \alpha_f(t) &= \sum_{t'=1}^t \frac{n+t'-1}{t} f_{t'} + 1 - \sum_{t'=1}^t f_{t'} = 1 + \sum_{t'=1}^t \frac{n+t'-1-t}{t} f_{t'} \\
 (\text{Disjointness of } P_0, P_1, \dots, P_j) \quad &= 1 + \sum_{j'=0}^{j-1} \sum_{t' \in P_{j'}} \frac{n+t'-1-t}{t} f_{t'} + \sum_{t' \in P_j: t' \leq t} \frac{n+t'-1-t}{t} f_{t'} \\
 (\text{Lemma 4.1}) \quad &\leq 1 + \sum_{j'=0}^{j-1} \frac{n+r_{j'}-1-t}{t} \delta + \frac{n-1}{t} \delta \\
 (r_{j'} \leq t \text{ for all } j') \quad &\leq 1 + j \frac{n-1}{t} \delta + \frac{n-1}{t} \delta \\
 &= 1 + (j+1) \frac{n-1}{t} \delta \\
 (\text{def of } \delta) \quad &< 1 + (j+1) \frac{n-1}{t} \frac{1}{2\gamma(i+1)} \\
 (j \leq i) \quad &\leq 1 + \frac{n-1}{2\gamma t} \\
 (t \geq n/\gamma) \quad &< 3/2
 \end{aligned}$$

Since $3/2 < \lambda$, the competitive ratio constraint cannot be tight anywhere, so the tail constraint must be tight everywhere by Theorem 3.1. $\square$

Now we can analyze the solution structure of each P_j for $j \leq i$. For each such j , let $\hat{t}_j = \lfloor \hat{t}/\gamma^j \rfloor$.

LEMMA 4.5. *Let $t \in P_j$ for $1 \leq j \leq i$.*

- *If $t \in [\ell_j, \ell_j + \hat{t}_j - 1]$, then*

$$f_t = \sum_{t'=1+\gamma^j(t-\ell_j)}^{1+\gamma^j(t-\ell_j)+(\gamma^j-1)} f_{t'}$$

- *If $t \in [\ell_j + \hat{t}_j + 1, r_j]$ then $f_t = 0$.*

Proof. We proceed by induction on j . We begin with the first part of the claim. Consider any $1 \leq j \leq i$. Lemma 4.4 implies that for every $t \in P_j$, the tail constraint is tight at both t and $t-1$. This implies that

$$(4.4) \quad f_t = \sum_{t' \in I_\gamma(t-1) \setminus I_\gamma(t)} f_{t'} = f_{\gamma \cdot t - (n-1) - (\gamma-1)} + \dots + f_{\gamma \cdot t - (n-1)}$$

First observe that if $t \in P_j$ the entries on the right hand side of Equation 4.4 are in P_{j-1} . Let $w_t = t - \ell_j$ then we have:

$$\gamma(\ell_j + w_t) - (n-1) - (\gamma-1) = \ell_{j-1} + \gamma w_t.$$

by Claim 4.1. Since $|P_{j-1}| = \gamma|P_j|$, we can conclude that $\ell_{j-1} + \gamma w_t \in P_{j-1}$. A similar calculation holds for the last point in the summation.

We will prove the Lemma by induction on j . Let $j = 1$ be the base case. Then, $\ell_0 = 1$ and $\ell_1 = (n-1)/\gamma + 1$, and by Claim 4.1, the summation in the Lemma statement begins at:

$$\ell_{j-1} + \gamma(t - \ell_j) = \gamma \cdot t - (n-1) - (\gamma-1),$$

which is equivalent to the first term in Equation 4.4. Combined with the fact that both sums carry on for $\gamma-1$ consecutive steps, this proves the base case.

For the inductive step, let $j > 1$, let $t \in P_j$ and let $t = \ell_j + w_t$. Then:

$$\begin{aligned} f_t &= \sum_{t' \in I_\gamma(t-1) \setminus I_\gamma(t)} f_{t'} \\ &= f_{\gamma \cdot t - (n-1) - (\gamma-1)} + \dots + f_{\gamma \cdot t - (n-1)} \\ &= f_{\ell_{j-1} + \gamma w_t} + \dots + f_{\ell_{j-1} + \gamma w_t + \gamma - 1} \\ &= \sum_{t'=1+\gamma^{j-1}\gamma w_t}^{1+\gamma^{j-1}\gamma w_t + \gamma^{j-1}-1} f_{t'} + \dots + \sum_{t'=1+\gamma^{j-1}(\gamma w_t + (\gamma-1)) + \gamma^{j-1}-1}^{1+\gamma^{j-1}(\gamma w_t + (\gamma-1)) + \gamma^{j-1}-1} f_{t'} \\ &= \sum_{t'=1+\gamma^j w_t}^{1+\gamma^j w_t + \gamma^j - 1} f_{t'}. \end{aligned}$$

Here the penultimate equality follows by induction and the last line follows because of the summands represent disjoint and consecutive intervals. This proves the first claim of the lemma.

We now consider the second claim of the lemma. The base case starts from $j = 0$, and the claim holds thanks to Lemma 4.3; recall that $r_0 = (n-1)/\gamma$. Now suppose that $t \in [\ell_j + \hat{t}_j + 1, r_j]$. Then $w_t = t - \ell_j \geq \hat{t}_j + 1$. So $\gamma w_t \geq \gamma(\hat{t}_j + 1) = \gamma \lfloor \hat{t}/\gamma^j \rfloor + \gamma \geq \lfloor \hat{t}/\gamma^{j-1} \rfloor + 1 = \hat{t}_{j-1} + 1$. Thus, we have

$$\begin{aligned} \gamma t - (n-1) - (\gamma-1) - (\ell_{j-1} + \hat{t}_{j-1} + 1) &= \gamma(\ell_j + w_t) - (n-1) - (\gamma-1) - (\ell_{j-1} + \hat{t}_{j-1} + 1) \\ &= \gamma w_t - (\hat{t}_{j-1} + 1) \geq 0, \end{aligned}$$

where the last equality follows from Claim 4.1. Further, we already showed above that $\gamma t - (n-1) - (\gamma-1), \dots, \gamma t - (n-1)$ are all in P_{j-1} . Therefore, they are all in $[\ell_{j-1} + \hat{t}_{j-1} + 1, r_{j-1}]$. Thus by induction we have that $f_t = f_{\gamma t - (n-1) - (\gamma-1)} + \dots + f_{\gamma t - (n-1)} = 0$ as claimed. $\square$

We can now combine this with the explicit formulas from Lemma 4.3 to give expressions for f_t for $t \in P_j$.

LEMMA 4.6. *Let $t \in P_j$.*

- *If $t = \ell_j$ then*

$$f_t = \sum_{t'=1}^{\gamma^j} \frac{\lambda-1}{n-1} \left(1 + \frac{1}{n-1}\right)^{t'-1}.$$

- *If $t \in [\ell_j + 1, \ell_j + \hat{t}_j - 1]$ then*

$$f_t = \sum_{t'=\gamma^j(t-\ell_j)+1}^{\gamma^j(t-\ell_j+1)} \frac{\lambda-1}{n-1} \left(1 + \frac{1}{n-1}\right)^{t'-1}.$$

- *If $t \in [\ell_j + \hat{t}_j + 1, r_j]$ then $f_t = 0$.*

Proof. The first and second cases follow directly from Lemmas 4.3 and 4.5, where we use the fact that since $t \leq \ell_j + \hat{t}_j - 1$, we know that $\gamma^j(t - \ell_j + 1) \leq \gamma^j \hat{t}_j = \gamma^j(\lfloor \hat{t}/\gamma^j \rfloor) \leq \hat{t}$, and hence the first case of Lemma 4.5 applies to all indices in the sum. The final case is directly from the second case of Lemma 4.5. $\square$

This clearly implies that the beginning of P_j is a γ^j exponential:

COROLLARY 4.1. *Let $t \in [\ell_j + 1, \ell_t + \hat{t}_j - 1]$. Then $f_t = \left(1 + \frac{1}{n-1}\right)^{\gamma^j} f_{t-1}$.*

Proof. By Lemma 4.6, we know that

$$\begin{aligned} f_t &= \sum_{t'=\gamma^j(t-\ell_j)+1}^{\gamma^j(t+1-\ell_j)} \frac{\lambda-1}{n-1} \left(1 + \frac{1}{n-1}\right)^{t'-1} \\ &= \left(1 + \frac{1}{n-1}\right)^{\gamma^j} \sum_{t'=\gamma^j(t-\ell_j)+1}^{\gamma^j(t+1-\ell_j)} \frac{\lambda-1}{n-1} \left(1 + \frac{1}{n-1}\right)^{t'-1-\gamma^j} \\ &= \left(1 + \frac{1}{n-1}\right)^{\gamma^j} \sum_{t'=\gamma^j(t-\ell_j-1)+1}^{\gamma^j(t-\ell_j)} \frac{\lambda-1}{n-1} \left(1 + \frac{1}{n-1}\right)^{t'-1} \\ &= \left(1 + \frac{1}{n-1}\right)^{\gamma^j} f_{t-1} \end{aligned}$$

as claimed. $\square$

We finally quantify how large n should be compared to i . Note that Lemma 4.6 is well defined if and only if the third case interval $[\ell_j + \hat{t}_j + 1, r_j]$ is non-empty, i.e., $\ell_j + \hat{t}_j + 1 \leq r_j$. Thus, we need the following:

$$(4.5) \quad \frac{(n-1)}{(\gamma-1)} \cdot \frac{(\gamma^j-1)}{\gamma^j} + 1 + \lfloor \hat{t}/\gamma^j \rfloor + 1 \leq \frac{n-1}{\gamma-1} (1 - 1/\gamma^{j+1})$$

Since $\hat{t} < \frac{n}{\gamma}$ (Lemma 4.2) and $n-1$ is a power of γ , we have $\hat{t} \leq \frac{n-1}{\gamma}$. Thus, equation (4.5) is satisfied if

$$\frac{n-1}{\gamma-1} \cdot \left(1 - \frac{1}{\gamma^j}\right) + 1 + \frac{n-1}{\gamma^{j+1}} + 1 \leq \frac{n-1}{\gamma-1} \left(1 - \frac{1}{\gamma^{j+1}}\right),$$

when $\frac{n-1}{\gamma^{j+1}} \geq 1$. By simplifying the inequality we obtain, $n \geq \frac{2\gamma^{j+1}}{\gamma-1} + 1$. Both inequalities hold when $n \geq 2\gamma^{j+1} + 1$.

4.5 Continuous Case. Recall that in the continuous setting all times are between 0 and 1, and it costs 1 to buy. This is equal to the limit of the discrete case if we first reparameterize so all times are between 0 and 1 (i.e., time t becomes time t/n) and then take the limit as n goes to infinity. That is, we can first just “rename” each time t to t/n , write the resulting expressions that arise from renaming, and then take the limit as n goes to infinity.

When we do this renaming for the above results, we first get that the interval P_j is equal to

$$P_j = \left\{ t : \frac{1}{n} \cdot \frac{(n-1)}{(\gamma-1)} \cdot \frac{(\gamma^j-1)}{\gamma^j} + \frac{1}{n} \leq t \leq \frac{1}{n} \cdot \frac{(n-1)}{(\gamma-1)} \cdot \frac{(\gamma^{j+1}-1)}{\gamma^{j+1}} \right\} \quad \text{for all } j \geq 0$$

Similarly, ℓ_j is defined to be $\frac{1}{n} \cdot \frac{(n-1)}{(\gamma-1)} \cdot \frac{(\gamma^j-1)}{\gamma^j} + \frac{1}{n}$, and we define $\hat{t}$ as before (except now it is a number between 0 and 1) and $\hat{t}_j$ essentially as before, to $\lfloor \hat{t}n/\gamma^j \rfloor/n$. Then with this new parameterization, Corollary 4.1 turns into the following lemma.

LEMMA 4.7. *Let t be an integer multiple of $1/n$ in the interval $[\ell_j + (1/n), \ell_t + \hat{t}_j - (1/n)]$. Then $f_t = \left(1 + \frac{1}{n-1}\right)^{\gamma^j} f_{t-1}$.*

A corollary of this lemma is that, if we consider two times $t' > t$ that are both integer multiples of $1/n$ in $[\ell_j + (1/n), \ell_t + \hat{t}_j - (1/n)]$, then $f_{t'}/f_t = \left(1 + \frac{1}{n-1}\right)^{(t'-t)n\gamma^j}$. This is just because there are precisely $(t' - t)n$ integer multiples of $1/n$ between t' and t .

So in the continuous case, where we now take the limit as $n \rightarrow \infty$, we get that

$$\frac{f_{t'}}{f_t} = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n-1}\right)^{(t'-t)n\gamma^j} = e^{\gamma^j(t'-t)},$$

simply because $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n-1}\right)^n = e$. Moreover, we now have a simple formula for $\hat{t}$, since after we reparameterize and take the limit as n goes to ∞ in (4.3) we get that $\hat{t}$ is the value where $(\lambda - 1) \int_0^{\hat{t}} e^t dt = \delta$, which is precisely equal to $\ln\left(1 + \frac{\delta}{\lambda-1}\right)$. So this is our value of $\hat{t}$, and $\hat{t}_j = \hat{t}/\gamma^j$.

Finally, we note that when we reparameterize and take the limit as n goes to infinity of the final case of Lemma 4.6, we get that in the continuous setting $f(t) = 0$ for $t \in (\ell_j + \hat{t}_j, \ell_{j+1})$. Hence in the continuous setting we get the following corollary.

COROLLARY 4.2. *Let $\gamma \geq 2$ be an integer. Let $\{P_j\}_{j \geq 0}$ be a partition of $(0, 1/(\gamma-1)]$, defined as $P_0 = (0, \frac{1}{\gamma})$ and $P_j = (\frac{1}{\gamma-1} \cdot (1 - \frac{1}{\gamma^j}), \frac{1}{\gamma-1} \cdot (1 - \frac{1}{\gamma^{j+1}})]$ for all $j \geq 1$. Let ℓ_j be the start point of P_j . In the continuous version of the ski rental problem with a single tail constraint (γ, δ) , the optimum solution $f(t)$ has the following structure in every P_j such that $j < 1/(6\delta) - 1$:*

- $f(t) = c_j \cdot e^{\gamma^j t}$ for t in the interval $(\ell_j, \ell_j + \hat{t}_j]$, and
- $f(t) = 0$ for all $t \in (\ell_j + \hat{t}_j, \ell_{j+1}]$,

where c_j and $\hat{t}_j < \ell_{j+1} - \ell_j$ are some constants depending on δ and j .

5 Special Case of Single Tail Constraint: $\delta = 0$

A particularly simple setting is where there is only a single tail constraint, and it is *pure*: $\delta = 0$. In other words, we must have probability 0 of having competitive ratio worse than γ . In this setting we will not only be able to explicitly write the optimal solution, but will additionally be able to give opt as an explicit function of γ . The exact expressions are given in the following two theorems. We assume for simplicity that $\gamma - 1$ divides $n - 1$, and to simplify notation we will let $\lambda = \text{opt}$.

THEOREM 5.1. *Suppose that there is a single tail constraint $(\gamma, 0)$. Then the best competitive ratio achievable under that tail constraint is exactly*

$$\lambda = \text{opt} = 1 + \frac{\gamma - 1}{1 + \frac{\gamma}{n-1} \left(n \left(\left(\frac{n-1}{n-1} \right)^{(n-1)(1-\frac{1}{\gamma-1})} - 1 \right) + 1 \right)}$$

THEOREM 5.2. *Let f be the optimal purchase distribution under a single pure tail constraint $(\gamma, 0)$. Then*

$$f_t = \begin{cases} 0 & t < \frac{n-1}{\gamma-1} \\ \frac{\lambda-1}{\gamma} & t = \frac{n-1}{\gamma-1} \\ \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} \cdot \left(1 + \frac{1}{n-1}\right)^{t-t_1-1} & t > \frac{n-1}{\gamma-1} \end{cases}$$

It turns out to be easiest to first prove Theorem 5.2 and then prove Theorem 5.1.

Proof. [Proof of Theorem 5.2] Note that in the rest of this proof, $\lambda = \text{opt}$ but since we have not yet proved Theorem 5.1 we will not instantiate this to any particular value; it is simply whatever the optimal expected competitive ratio is.

Since $\delta = 0$, if $t \in I_\gamma(x)$ for some $x \in [n]$ then we must have $f_t = 0$. It is easy to see that $I_\gamma(x)$ is nonempty if and only if $x \in I_\gamma(x)$, which happens if and only if $x < \frac{n-1}{\gamma-1}$. So we know that $f_t = 0$ for all $t < \frac{n-1}{\gamma-1}$, and that $I_\gamma(t) = \emptyset$ for all $t \geq \frac{n-1}{\gamma-1}$. Thus Theorem 3.1 implies that $\alpha_{f^*}(t) = \lambda$ for all $t \geq \frac{n-1}{\gamma-1}$, where f^* is the optimal solution.

Let $t_1 = \frac{n-1}{\gamma-1}$ be the first time where $I_\gamma(t_1) = \emptyset$, and so $\alpha_{f^*}(t_1) = \lambda$ and $f_t = 0$ for all $t < t_1$. So we have

$$\lambda = \alpha_{f^*}(t_1) = \sum_{t \leq t_1} \frac{n+t-1}{t_1} f_t + 1 - \sum_{t \leq t_1} f_t = \frac{n+t_1-1}{t_1} f_{t_1} + 1 - f_{t_1} = \frac{n-1}{t_1} f_{t_1} + 1 = (\gamma-1)f_{t_1} + 1$$

Rearranging, we get that $f_{t_1} = \frac{\lambda-1}{\gamma-1}$.

Since we know that $f_t = \lambda = \alpha_{f^*}(t_1)$ for all $t \geq t_1$, we can apply Lemma 2.1 with $a-1 = t_1$ to obtain all of the other probabilities. In particular, Lemma 2.1 implies that

$$f_{t_1+1} = \frac{1}{t_1(n-1)}(n+t_1-1)f_{t_1} = \frac{\gamma-1}{(n-1)^2} \left(n + \frac{n-1}{\gamma-1} - 1 \right) \frac{\lambda-1}{\gamma-1} = \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} = \frac{\gamma}{n-1} f_{t_1}$$

Lemma 2.1 further implies that $f_t = \left(1 + \frac{1}{n-1}\right) f_{t-1}$ for all $t > t_1 + 1$. So we have that for all $t \geq t_1 + 1$,

$$f_t = \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} \cdot \left(1 + \frac{1}{n-1}\right)^{t-t_1-1}.$$

□

We can now use Theorem 5.2 to prove Theorem 5.1.

Proof. [Proof of Theorem 5.1] We can use that fact that f must be a distribution to give an exact characterization of λ in terms of n and γ . We have that

$$\begin{aligned} 1 &= \sum_{t=1}^n f_t = f_{t_1} + \sum_{t=t_1+1}^n f_t = \frac{\lambda-1}{\gamma-1} + \sum_{t=t_1+1}^n \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} \left(1 + \frac{1}{n-1}\right)^{t-t_1-1} \\ &= \frac{\lambda-1}{\gamma-1} + \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} \sum_{t=t_1+1}^n \left(1 + \frac{1}{n-1}\right)^{t-t_1-1} \\ &= \frac{\lambda-1}{\gamma-1} + \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} \sum_{i=0}^{n-t_1-1} \left(1 + \frac{1}{n-1}\right)^i \\ &= \frac{\lambda-1}{\gamma-1} + \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} \cdot \left(n \left(\left(\frac{n}{n-1} \right)^{n-t_1-1} - 1 \right) + 1 \right) \\ &= \frac{\lambda-1}{\gamma-1} + \frac{\gamma(\lambda-1)}{(n-1)(\gamma-1)} \cdot \left(n \left(\left(\frac{n}{n-1} \right)^{(n-1)(1-\frac{1}{\gamma-1})} - 1 \right) + 1 \right) \end{aligned}$$

Multiplying both sides by $\gamma-1$ gives us

$$\begin{aligned} \gamma-1 &= \lambda-1 + \frac{\gamma(\lambda-1)}{n-1} \left(n \left(\left(\frac{n}{n-1} \right)^{(n-1)(1-\frac{1}{\gamma-1})} - 1 \right) + 1 \right) \\ &= (\lambda-1) \left(1 + \frac{\gamma}{n-1} \left(n \left(\left(\frac{n}{n-1} \right)^{(n-1)(1-\frac{1}{\gamma-1})} - 1 \right) + 1 \right) \right). \end{aligned}$$

We can now solve for λ , giving us

$$\lambda = 1 + \frac{\gamma-1}{1 + \frac{\gamma}{n-1} \left(n \left(\left(\frac{n}{n-1} \right)^{(n-1)(1-\frac{1}{\gamma-1})} - 1 \right) + 1 \right)}$$

as claimed. □

Note that if we take the limit as $n \rightarrow \infty$ and use the fact that $\lim_{x \rightarrow \infty} (1 + 1/x)^x = e$, we get that λ approaches

$$1 + \frac{\gamma - 1}{1 + \gamma \left(e^{1 - \frac{1}{\gamma-1}} - 1 \right)}$$

In particular, when $\gamma = 2$ we get that $\lim_{n \rightarrow \infty} \lambda = 2$, i.e., we recover the classical deterministic bound, and when $\gamma \rightarrow \infty$ we get that $\lim_{n \rightarrow \infty} \lambda = 1 + \frac{1}{e-1} = \frac{e}{e-1}$, i.e., we recover the classical randomized bound.

6 Conclusion

In this work we extended the classic ski rental problem by taking the tail risk into account. While the problem has been studied for decades, and is perhaps the simplest online problem, there had been no previous investigation of how to balance expected performance and the risk of the randomized online algorithms performing worse than the deterministic option. We gave a characterization theorem of the optimal purchasing distribution, and used this theorem to prove several surprising properties of the optimal distribution even under a single tail constraint. At a high level, we showed that the optimal distribution has almost *none* of the nice properties that occur without tail constraints: it is non-monotone, can alternate between regions of zero probability and non-zero probability arbitrarily many times, and can grow arbitrarily quickly even when continuous. We also gave an explicit description of the special case of pure tail constraints, and developed an algorithm to compute the optimal purchase distribution efficiently.

We hope that our work inspires further investigation of tail bounds in online algorithms. As discussed, in online settings we cannot simply repeat an algorithm multiple times in order to convert a bound on the expectation into a bound on the tail, as we usually do in offline computational settings. Yet surprisingly, there seems to be little work on tail bounds in online settings: even the ski rental problem, the most basic online setting of all, had not been investigated prior to this work! What about other online settings where there are well-known and well-understood randomized algorithms, e.g., online matching, TCP aggregation, or the many variants of prophet inequalities? Can we characterize the optimal randomized algorithms for those problems in the presence of tail bounds? Do those optimal algorithms differ significantly from the pure expectation setting, as they do for ski rental?

Acknowledgements

We would like to thank Robert Kleinberg for fruitful discussions and comments on an earlier draft of this work.

References

- [1] Antonios Antoniadis, Christian Coester, Marek Eliás, Adam Polak, and Bertrand Simon. Learning-augmented dynamic power management with multiple states via new ski rental bounds. In Marc’Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 16714–16726, 2021.
- [2] Soumya Banerjee. Improving online rent-or-buy algorithms with sequential decision making and ML predictions. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- [3] Niv Buchbinder, Joseph Seffi Naor, et al. The design of competitive online algorithms via a primal–dual approach. *Foundations and Trends® in Theoretical Computer Science*, 3(2–3):93–263, 2009.
- [4] Anna R. Karlin, Claire Kenyon, and Dana Randall. Dynamic TCP acknowledgment and other stories about $e/(e-1)$. *Algorithmica*, 36(3):209–224, 2003.
- [5] Anna R. Karlin, Mark S. Manasse, Lyle A. McGeoch, and Susan S. Owicki. Competitive randomized algorithms for nonuniform problems. *Algorithmica*, 11(6):542–571, 1994.
- [6] Anna R. Karlin, Mark S. Manasse, Larry Rudolph, and Daniel Dominic Sleator. Competitive snoopy caching. *Algorithmica*, 3:77–119, 1988.
- [7] Ali Khanafer, Murali Kodialam, and Krishna P. N. Puttaswamy. The constrained ski-rental problem and its application to online cloud cost optimization. In *2013 Proceedings IEEE INFOCOM*, pages 1492–1500, 2013.
- [8] Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. *CoRR*, abs/2006.09123, 2020.

- [9] Debmalya Panigrahi. Lecture notes on online algorithms. https://courses.cs.duke.edu/fall115/compsci532/scribe_notes/lec19.pdf.
- [10] Filipo Studzinski Perotto, Mathieu Bourgain, Bruno C. Silva, and Laurent Vercouter. Open problem: Risk of ruin in multiarmed bandits. In Alina Beygelzimer and Daniel Hsu, editors, *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pages 3194–3197. PMLR, 2019.
- [11] Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ML predictions. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 9684–9693, 2018.
- [12] Charles Riou, Junya Honda, and Masashi Sugiyama. The survival bandit problem. *CoRR*, abs/2206.03019, 2022.
- [13] Anant Shah and Arun Rajkumar. Sequential ski rental problem. In Frank Dignum, Alessio Lomuscio, Ulle Endriss, and Ann Nowé, editors, *AAMAS '21: 20th International Conference on Autonomous Agents and Multiagent Systems, Virtual Event, United Kingdom, May 3-7, 2021*, pages 1173–1181. ACM, 2021.
- [14] Shufan Wang, Jian Li, and Shiqiang Wang. Online algorithms for multi-shop ski rental with machine learned advice. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- [15] Alexander Wei and Fred Zhang. Optimal robustness-consistency trade-offs for learning-augmented online algorithms. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- [16] Bingham Wu, Wei Bao, and Dong Yuan. Competitive analysis for two-level ski-rental problem. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 12034–12041. AAAI Press, 2021.
- [17] Yifan Wu, Roshan Shariff, Tor Lattimore, and Csaba Szepesvári. Conservative bandits. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 1254–1262. JMLR.org, 2016.
- [18] Ali Zeynali, Bo Sun, Mohammad Hassan Hajiesmaili, and Adam Wierman. Data-driven competitive algorithms for online knapsack and set cover. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 10833–10841. AAAI Press, 2021.