Mitigating Tail Catastrophe in Steered Database Query Optimization with Risk-Averse Contextual Bandits

Mónika Farsang[1] **Paul Mineiro**[2] **Wangda Zhang**[2]
[1]Vienna University of Technology, Vienna, Austria [2]Microsoft Research, New York, NY, USA
monika.farsang@tuwien.ac.at
{pmineiro,wangdazhang}@microsoft.com

Abstract

Contextual bandits with average-case statistical guarantees are inadequate in risk-averse situations because they might trade off degraded worst-case behaviour for better average performance. Designing a risk-averse contextual bandit is challenging because exploration is necessary but risk-aversion is sensitive to the entire distribution of rewards; nonetheless we exhibit the first risk-averse contextual bandit algorithm with an online regret guarantee. We apply the technique to a self-tuning software scenario in a production exascale data processing system, where worst-case outcomes should be avoided.

1 Introduction

Contextual bandits [Auer et al., 2002, Langford and Zhang, 2007] are a mature technology with numerous applications: however, adoption has been most aggressive in recommendation scenarios [Bouneffouf and Rish, 2019], where the worst-case outcome is user annoyance. At the other extreme are medical and defense scenarios where worst-case outcomes are literally fatal. In between are scenarios of interest where bad outcomes are tolerable but should be avoided, e.g., logistics; finance; and self-tuning software, where the term *tail catastrophe* highlights the inadequacy of average case performance guarantees in real-world applications [Marcus et al., 2021]. These scenarios demand risk-aversion, i.e., decisions should sacrifice average performance in order to avoid worst-case outcomes, and incorporating risk-aversion into contextual bandits would facilitate adoption.

This paper solves risk-averse decision making for contextual bandits via reduction to regression, resulting in the first risk-averse contextual bandit algorithm with an online regret guarantee. The regret guarantee applies over adversarially chosen context sequences and includes the exploration choices made by the algorithm. The approach utilizes arbitrary (online learnable) function classes to estimate a risk measure and introduces no computational overhead relative to the risk-neutral setting; introduces statistical overhead directly related to the desired level of risk-aversion, with no overhead in the risk-neutral limit; and composes with other innovations within the Decision-to-Estimation framework [Foster et al., 2021], e.g., linear representability [Zhu and Mineiro, 2022]. We apply the technique to a self-tuning software scenario in a production exascale data processing system [Zhang et al., 2022], where worst-case outcomes (e.g., performance regressions) should be avoided.

Related Work Prior work on risk-averse *contextual* bandits is limited. Sun et al. [2017] address the adversarial contextual setting by treating total risk as a constraint, but requires an additional risk value observed along with cost. Bouneffouf [2016] presents a contextual UCB algorithm which optimizes for mean reward, but which modulates the level of ϵ-greedy exploration based upon a risk estimate. Concurrent to our work, Saux and Maillard [2023] recently also use the UCB framework solving a convex problem under the assumption of linear bandits, which do not apply to any of our non-linear predictors in the experiments. Huang et al. [2021] study the finite sample behaviour of off-policy estimation for a broad class of risk measures.

Machine Learning for Systems Workshop at 37th NeurIPS Conference, 2023, New Orleans, LA, USA.

The inadequacy of average-case guarantees is a recurring theme in real-time systems applications. Jalaparti et al. [2013] improve tail latencies of request-response workflows by minimizing variance. Schad et al. [2010] use the same performance measure in cloud computing. CVaR optimization is also present in systems applications: Mena et al. [2014] propose a multi-objective optimization technique with CVaR as risk metric in a sizing and allocation problem of renewable generation, whereas Moreno and Strbac [2015] limit risk exposure to high impact low probability events in distribution substations through this metric. However, only a small number of related bandit studies tackle risk-aware optimization in systems applications. Marcus et al. [2021] present a bandit optimizer to improve the tail latency of queries. Sachidananda and Sivaraman [2021] design an autoscaler using a multi-armed bandit algorithm to optimize median or tail latency for microservice applications.

2 Problem Statement

Traditional database query optimizers are built for general scenarios at all scales, and as a result they often generate suboptimal query execution plans for a specific workload before finetuning by the database administrators Leis et al. [2015]. To improve the query plan generated by the optimizer, recent work has adopted a data-driven instance optimization approach that uses machine learning algorithms to learn from the past workload and steer a query optimizer towards better query plan choices. SCOPE [Power et al., 2021], the exascale cloud data processing system at Microsoft, has a highly configurable query optimizer that uses a contextual bandit framework to select optimizer flags on a per-query basis [Zhang et al., 2022]. Different optimizer flag leads to different query plan and thus different performance and execution cost for an input query. For this application, although the system has a set of default flag settings, there is no single optimal flag configuration working for all input queries, and the best configuration depends on the specific query.

In earlier work Zhang et al. [2022], the objective of this steered query optimizer is mainly to increase the overall average performance of queries, and thus risk-neutral contextual bandits were used in SCOPE. While it is valuable to improve the average performance, it is important to also avoid performance regressions (queries with worse performance than using the default configurations), which lead to customer frustration and extra investigation work. At the scale of SCOPE, even if a small percentage of regressions are introduced as a result of employing learned optimizer flags, the number of customer incidents would easily overwhelm the product team. This observation motivates a risk-averse approach as formulated below.

Contextual Bandits We describe the contextual bandit problem, which proceeds over T rounds. At each round $t \in [T]$, the learner receives a context $x_t \in \mathcal{X}$ (the context space), selects an action $a_t \in \mathcal{A}$ (the action space), and then observes a loss $l_t(a_t)$, where $l_t : \mathcal{A} \to [0, 1]$ is the underlying loss function. In the query optimization scenario, we assembled query information as context and assessed the performance of multiple optimizer flag configurations (as actions) per query relative to a default configuration, using the fractional change as the reward.

Reduction to Regression We attack the contextual bandit problem via reduction to regression, working with a user-specified class of regression functions $\mathcal{F} \subseteq (\mathcal{X} \times \mathcal{A} \to [0, 1])$ that aims to estimate a risk measure ρ of the conditional loss distribution. We make the following realizability assumption

$$\forall a \in \mathcal{A}, t \in [T] : \exists f^* \in \mathcal{F} : f^*(x_t, a) = \rho\left((l_t)_a\right),$$

i.e., our function class includes a function which correctly estimates the value of the risk measure arising from any action a in context x_t. This constrains the adversary's choices, as l_t must be consistent with realizability, but there are many random variables that achieve a particular risk value.

Regression Oracle We assume access to an online regression oracle $\mathbf{Alg}_{\mathsf{Reg}}$, which is an algorithm for sequential predication under strongly convex losses using $\mathcal{F}$ as a benchmark class. More specifically, the oracle operates in the following protocol: at each round $t \in [T]$, the algorithm receives a context $x_t \in \mathcal{X}$, makes a prediction $\hat{f}_t$, where $\hat{f}_t(x_t, a)$ is interpreted as the prediction for action a, and then observes an action $a_t \in \mathcal{A}$ and realized outcome $l_t(a_t) \in [0, 1]$. Typically, its performance is evaluated by the squared loss $g_t(\hat{f}_t) = (\hat{f}_t(x_t, a_t) - l_t(a_t))^2$ when measuring the estimation regret.

Expectile Loss When optimizing for a risk-aware quantity, we can replace a loss that elicits the mean with a loss that elicits the risk-aware quantity. The choice from the risk measures is limited, and one suitable choice is EVaR. EVaR is less familiar to the machine learning community but is a popular risk-measure in financial applications [Bellini and Di Bernardino, 2017], whose proponents champion the superior finite-sample guarantees induced by strong convexity [Rossello, 2022]. Ziegel [2016] shows the class of elicitable law-invariant coherent risk measures for real-valued random variables is precisely Entropic Value at Risk (EVaR_q) for $q \in \left(0, \frac{1}{2}\right]$, defined as

$$\mathsf{EVaR}_q(D) = \arg\min_{\hat{v}\in[0,1]} \mathbb{E}_{v\sim D}\left[(1-q)\left((v-\hat{v})_+\right)^2 + q\left((\hat{v}-v)_+\right)^2\right], \tag{1}$$

where $(x)_+ = \max(x, 0)$. This asymmetrical strongly convex loss encourages overprediction relative to the mean, implying infrequent large losses correspond to increased risk. A minimizer of equation (1) is called an *expectile*. Certain technical qualifications are necessary for the minimum to be achieved (bounded realization suffices). We refer to the elicitation loss function as *expectile loss*.

Expectiles can be elicited by asymmetric squared loss or log loss, and thus are a natural generalization of the mean. This generalization is based on the choice of the q value leading to risk-averse ($q < 0.5$) and standard risk-neutral ($q = 0.5$) settings.

3 Algorithm

By using the Estimation-to-Decision framework, we derive the resulting algorithm, which is the first risk-averse contextual bandit with an online guarantee. It is the SquareCB algorithm [Foster and Rakhlin, 2020] instantiated with an expectile loss regression oracle.

The oracle's performance via the expectile loss is measured as the following:

$$g_t(\hat{f}_t) \doteq \left((1-q)\left((v-\hat{v})_+\right)^2 + q\ \left((\hat{v}-v)_+\right)^2\right)\Big|_{v=l_t(a_t),\hat{v}=\hat{f}_t(x_t,a_t)}.$$

We assume $\mathbf{Alg}_{\mathsf{Reg}}$ guarantees that for any (potentially adaptively chosen) sequence $(x_t, a_t, l_t)_{t=1}^T$, $\sum_{t=1}^T \left(g_t(\hat{f}_t) - g_t(f^*)\right) \leq \mathbf{Reg}_{\mathsf{EVaR}_q}(T)$ for some (non-data-dependent) function $\mathbf{Reg}_{\mathsf{EVaR}_q}(T)$.

Algorithm 1 Finite Action Set

```
for $t = 1, 2, \dots, T$ do
    Receive context $x_t$.
    $\hat{f}_t \leftarrow \mathbf{Alg}_{\mathsf{Reg}}.\mathsf{predict}(x_t)$.
    $\hat{a}_t \leftarrow \mathsf{argmin}_{a\in\mathcal{A}} \hat{f}_{t,a}$.
    Sample $a_t \sim \mathsf{AL}(\hat{f}_t, \hat{a}_t) = \begin{cases} \frac{1}{|\mathcal{A}|+4\theta\gamma\left(\hat{f}(a)-\hat{f}(\hat{a}_t)\right)} & a \neq \hat{a}_t \\ 1 - \sum_{a\neq\hat{a}_t} \frac{1}{|\mathcal{A}|+4\theta\gamma\left(\hat{f}(a)-\hat{f}(\hat{a}_t)\right)} & a = \hat{a}_t \end{cases}$
    Play $a_t$ and observe loss $l_t$.
    Call $\mathbf{Alg}_{\mathsf{Reg}}.\mathsf{update}(x_t, a_t, l_t)$.
```

Theorem 3.1. *Algorithm 1 guarantees* $\mathbf{Reg}_{\mathsf{CB}}(T) \leq O\left(\frac{1}{\theta}\sqrt{|\mathcal{A}|T\mathbf{Reg}_{\mathsf{EVaR}_q}(T)}\right)$, *where* $\theta = \min(q, 1-q)$.

Proofs and further details are in our full paper Farsang et al. [2022]. We emphasize that this regret is with respect to the risk measure of the best action for each context, and includes the exploration activity of the algorithm. The θ factor indicates the difficulty of competing with an extreme expectile. The result is intuitive as θ is the strong convexity parameter of the expectile loss.

4 Results

We conducted experiments with the SCOPE query optimizer [Zhang et al., 2022] to demonstrate a scenario where average-case guarantees are inadequate, and to exhibit a trade-off between maximizing average-case and minimizing worst-case outcomes.

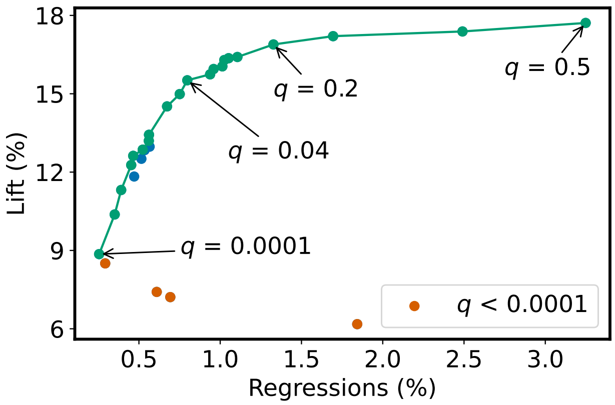

Figure 1: Query Optimization results. Varying the learning expectile (q) yields different realized lift and regression. Extreme expectiles are indicated in orange, the Pareto front is in green. With moderate q, reductions in regression are proportionally larger than reductions in lift.

The dataset of this experiment was collected from the SCOPE query engine. For each query, the dataset contains its feature embedding as well as the performance of multiple optimizer flag configurations relative to the default configuration. The number of actions per example varies depending upon constraints imposed by the optimizer: it ranges from 2 to 22, with a mean of 4.3 and a median of 3. We use this dataset to construct a query optimization simulator as follows: First, the algorithm is presented the query information and the configuration choices; Then the algorithm selects a configuration and receives the relative performance as the reward for that configuration. The experiment is implemented with VowpalWabbit [Langford et al., 2007]. Code to reproduce our results, along with the dataset, is available online [1].

Figure 1 summarizes the results, where the x-axis is the average performance regression for the regressed queries, and the y-axis shows the overall average performance lift. For some of those queries with large regressions, they will break the strict service level agreements, leading to customer dissatisfaction and extra manual investigation work. Therefore, it is necessary for the product team to adopt a risk-averse approach, mitigating those extreme outcomes in the tail distribution while trying to improve the overall performance.

As shown in the chart, varying the learning expectile (q) illuminates the trade-off between lift and regression. There is a moderate q regime where reductions in regression are proportionally larger than reductions in lift. In particular, our result demonstrates that the risk-averse algorithm effectively reduces the average magnitude of regressed queries by over 50%, from 3.2% (risk neutral q=0.5) to 1.3% (risk averse q=0.2), while maintaining the same level of overall performance lift. For $q \leq 0.0001$ every point is Pareto-dominated, as anticipated by the theoretical analysis (the regret bound degrades at extreme quantiles). Nevertheless, the relatively larger change in average performance regressions from risk-neutral to risk-averse with moderate q illustrates the usefulness of this technique.

5 Conclusion

This paper studies the application of contextual bandits to scenarios where average-case statistical guarantees are inadequate. We show that the composition of reduction to online regression and expectile loss is analytically tractable, computationally convenient, and empirically effective. Our experiment demonstrates the trade-off between maximizing average-case outcomes and minimizing worst-case performance in a steered database query optimizer. These results highlight the effectiveness of our risk-averse contextual bandit approach, which can be easily applied to other system problems that require risk aversion.

[1] https://github.com/zwd-ms/risk_averse_cb

References

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

Fabio Bellini and Elena Di Bernardino. Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506, 2017.

Djallel Bouneffouf. Contextual bandit algorithm for risk-aware recommender systems. *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 4667–4674, 2016.

Djallel Bouneffouf and Irina Rish. A survey on practical applications of multi-armed and contextual bandits. *CoRR*, abs/1904.10040, 2019. URL `http://arxiv.org/abs/1904.10040`.

Mónika Farsang, Paul Mineiro, and Wangda Zhang. Conditionally risk-averse contextual bandits. *arXiv preprint arXiv:2210.13573*, 2022.

Dylan Foster and Alexander Rakhlin. Beyond ucb: Optimal and efficient contextual bandits with regression oracles. In *International Conference on Machine Learning*, pages 3199–3210. PMLR, 2020.

Dylan J Foster, Sham M Kakade, Jian Qian, and Alexander Rakhlin. The statistical complexity of interactive decision making. *arXiv preprint arXiv:2112.13487*, 2021.

Audrey Huang, Liu Leqi, Zachary Lipton, and Kamyar Azizzadenesheli. Off-policy risk assessment in contextual bandits. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 23714–23726. Curran Associates, Inc., 2021. URL `https://proceedings.neurips.cc/paper/2021/file/c7502c55f8db540625b59d9a42638520-Paper.pdf`.

Virajith Jalaparti, Peter Bodík, Srikanth Kandula, Ishai Menache, Mikhail Rybalkin, and Chenyun Yan. Speeding up distributed request-response workflows. *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, 2013.

John Langford and Tong Zhang. The epoch-greedy algorithm for contextual multi-armed bandits. *Advances in neural information processing systems*, 20(1):96–1, 2007.

John Langford, Lihong Li, and Alex Strehl. Vowpal wabbit online learning project, 2007.

Viktor Leis, Andrey Gubichev, Atanas Mirchev, Peter Boncz, Alfons Kemper, and Thomas Neumann. How good are query optimizers, really? *Proceedings of the VLDB Endowment*, 9(3):204–215, 2015.

Ryan Marcus, Parimarjan Negi, Hongzi Mao, Nesime Tatbul, Mohammad Alizadeh, and Tim Kraska. Bao: Making learned query optimization practical. In *Proceedings of the 2021 International Conference on Management of Data*, pages 1275–1288, 2021.

Rodrigo Mena, Martin Hennebel, Yan-Fu Li, Carlos Ruiz, and Enrico Zio. A risk-based simulation and multi-objective optimization framework for the integration of distributed renewable generation and storage. *Renewable and Sustainable Energy Reviews*, 37:778–793, 2014. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2014.05.046. URL `https://www.sciencedirect.com/science/article/pii/S1364032114003712`.

Rodrigo Moreno and Goran Strbac. Integrating high impact low probability events in smart distribution network security standards through cvar optimisation. 2015.

Conor Power, Hiren Patel, Alekh Jindal, Jyoti Leeka, Bob Jenkins, Michael Rys, Ed Triou, Dexin Zhu, Lucky Katahanas, Chakrapani Bhat Talapady, et al. The Cosmos big data platform at Microsoft: over a decade of progress and a decade to look forward. *Proceedings of the VLDB Endowment*, 14 (12):3148–3161, 2021.

Damiano Rossello. Performance measurement with expectiles. *Decisions in Economics and Finance*, pages 1–32, 2022.

Vighnesh Sachidananda and Anirudh Sivaraman. Learned autoscaling for cloud microservices with multi-armed bandits. *ArXiv*, abs/2112.14845, 2021.

Patrick Saux and Odalric Maillard. Risk-aware linear bandits with convex loss. In *International Conference on Artificial Intelligence and Statistics*, pages 7723–7754. PMLR, 2023.

Jörg Schad, Jens Dittrich, and Jorge-Arnulfo Quiané-Ruiz. Runtime measurements in the cloud. *Proceedings of the VLDB Endowment*, 3:460 – 471, 2010.

Wen Sun, Debadeepta Dey, and Ashish Kapoor. Safety-aware algorithms for adversarial contextual bandit. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3280–3288. PMLR, 06–11 Aug 2017. URL `https://proceedings.mlr.press/v70/sun17a.html`.

Wangda Zhang, Matteo Interlandi, Paul Mineiro, Shi Qiao, Nasim Ghazanfari, Karlen Lie, Marc Friedman, Rafah Hosn, Hiren Patel, and Alekh Jindal. Deploying a steered query optimizer in production at Microsoft. In *Proceedings of the 2022 International Conference on Management of Data*, 2022.

Yinglun Zhu and Paul Mineiro. Contextual bandits with smooth regret: Efficient learning in continuous action spaces. In *International Conference on Machine Learning*, pages 27574–27590. PMLR, 2022.

Johanna F Ziegel. Coherence and elicitability. *Mathematical Finance*, 26(4):901–918, 2016.