Optimization of Utility-Based Shortfall Risk

A Project Report

submitted by

VISHWAJIT PRAKASH HEGDE

in partial fulfilment of the requirements
for the award of the degree of

BACHELOR OF TECHNOLOGY &
MASTER OF TECHNOLOGY

DEPARTMENT OF MECHANICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY, MADRAS.
June 2022

THESIS CERTIFICATE

This is to certify that the thesis entitled **Optimization of Utility-Based Shortfall Risk**, submitted by **Vishwajit Prakash Hegde** (**ME17B039**), to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelors of Technology** and **Master of Technology**, is a bona fide record of the research work carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. L. A. Prashanth
Research Guide
Assistant Professor
Dept. of Computer Science and Engineering
IIT-Madras, 600036

Place: Chennai

Date: June 16, 2022

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to Dr. L. A. Prashanth for providing an opportunity to work under his guidance, and for his constant support throughout the duration of the project.

I would like to thank my college, IIT Madras for offering Data Science as an interdisciplinary dual degree course and for providing all the necessary facilities and infrastructure needed for my work.

I would also like to thank the organizers and coordinators of Data Science course for allowing me to work in the field of my interest.

Lastly, I would like to thank my family and peers for providing necessary help when needed.

ABSTRACT

Utility-Based Shortfall Risk (UBSR) is a convex risk measure with many desirable properties which make it preferable over other commonly used risk measures such as Value-at-Risk. UBSR makes use of a utility function for penalizing large losses. We consider the task of optimization of UBSR within a parameterized class of random variables. Optimization of UBSR is tackled under two different settings. First, we consider the bandit feedback setting where the optimization problem is modeled as the problem of minimizing simple regret. We apply Stochastic Risk Optimistic Optimization (StoROO) algorithm developed for optimizing general risk measures. We derive upper and lower confidence bounds required by the algorithm using concentration inequalities derived for UBSR. We also obtain upper bound on the simple regret.

Next, we consider the optimization of UBSR using stochastic gradient descent (SGD) algorithm, and derive a non-asymptotic bound for the last iterate of SGD. We propose a certain step size sequence and batch size sequence to obtain an upper bound on the error of the order $O(1/\sqrt{n})$ where n is the number of iterations for which SGD is run. We also consider UBSR optimization when the samples follow a Markov chain and are not independent and identically distributed (i.i.d). The derived bound is of the order $O(\sqrt{\log n/n})$.

TABLE OF CONTENTS

ABBREVIATIONS

UBSR	Utility-Based Shortfall Risk
StoROO	Stochastic Risk Optimistic Optimization
SGD	Stochastic Gradient Descent
CVaR	Conditional Value-at-Risk
VaR	Value-at-Risk
CPT	Cumulative Prospect Theory
SAA	Sample Average Approximation
UCB	Upper Confidence Bound
LCB	Lower Confidence Bound
i.i.d.	independent and identically distributed

NOTATION

λ	risk level
ℓ	utility function
SR_λ	UBSR function
ξ_i	*i*-th sample
t_m	*m*-sample estimate of UBSR
a_k	step-size for *k*-th iteration
W_1	Wasserstein distance of order 1

CHAPTER 1

Introduction

Risk minimization is an important consideration in many financial applications such as portfolio optimization. A risk measure is usually defined as a property of the probability distribution of returns/losses. Some of the widely studied risk measures are Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR) [28], Cumulative Prospect Theory (CPT) [31]. VaR has been in use for financial applications since the 1990s. In recent years, CVaR has replaced VaR as a preferable risk measure because of several shortcomings of the latter risk measure. One such problem is that VaR does not take into account the size of large losses in case of severe default events. On the other hand, CVaR is a coherent risk measure with desirable properties such as homogeneity, sub-additivity, translational invariance and monotonicity, cf. [1]. In [11], the authors propose to replace the properties of homogeneity and sub-additivity with convexity and thereby define a convex risk measure as a risk measure with the following properties: (i) convexity; (ii) translational invariance; and (iii) monotonicity.

Utility-Based Shortfall Risk (UBSR) which was first introduced in [11] is a risk measure which belongs to the class of convex risk measures as well. For a given loss function (representing some risk attitude) and a threshold value (risk level), the UBSR of a financial position is the minimum amount of capital needed to be added to the position such that the new position's risk level is below the prescribed risk level. Apart from being a convex risk measure, UBSR also has the following properties: (i) UBSR is invariant under randomization i.e, if two random variables have acceptable levels of risks, then the risk of the diversified position does not exceed the weighted sum of risks, cf. [13]; and (ii) In UBSR, a convex utility function is used to penalize the loss, which increases with the increase in loss. This utility function sometimes also referred to as loss function could be used to encode investor's risk preferences. Combined with the fact that CVaR is calculated by only using the values of the loss distribution beyond a certain quantile, it can be argued that UBSR as a risk measure is more desirable than CVaR.

In this work, we consider optimization of UBSR within a parameterized class of random variables. A possible application of this optimization problem is as follows: suppose a portfolio manager desires to invest a fixed amount of capital in 2 different assets. She has to decide on a ratio in which she would like to distribute the capital between the assets. For a chosen ratio, say θ, the resulting portfolio would be exposed to a certain amount of risk which could be quantified using UBSR. The goal of minimization of the risk becomes the problem of optimizing UBSR. Refer [17], [5] and [10] for the usage of UBSR in the context of portfolio optimization.

UBSR optimization is studied under two different settings in this work. First, we consider the problem under bandit feedback where the optimization problem is modeled as the problem of minimizing the simple regret. We apply StoROO (Stochastic Risk Optimistic Optimization) algorithm proposed in [30]. We construct confidence intervals for UBSR using the concentration inequalities derived in [25]. Regret bounds are obtained based on the generic regret analysis described in [30].

Next, we propose a Stochastic Gradient Descent (SGD) algorithm and derive non-asymptotic bounds for the last iterate of SGD. Stochastic gradient methods are very useful tools in stochastic optimization problems. They have been studied widely in the past couple of decades owing to their applications in large scale machine learning, refer [7] for a survey on the non-asymptotic analysis of these methods. In order to estimate the gradient of UBSR, we make use of the sensitivity formula derived in [17]. UBSR derivative estimate requires one to estimate UBSR value using a batch of samples. Such an estimate is biased meaning the estimation error does not have zero expectation. This results in UBSR derivative estimate having a bias which depends on the batch size used to estimate UBSR derivative.

The non-asymptotic analysis provided here is based on the non-asymptotic analysis of last iterate SGD provided in [18]. Here, the authors assume that the gradient estimate is unbiased and propose a new step size sequence for obtaining information theoretically optimal bounds of the order $O(1/\sqrt{n})$ where n is the number of iterations for which SGD is run. In this work, we propose a modified batch size sequence for estimation purposes along with the modified step size sequence for obtaining non-asymptotic bounds of the order $O(1/\sqrt{n})$. We also derive non-asymptotic bounds in the case where the samples are not i.i.d. but follow a Markov chain. The resulting bounds are of the order of $O(\sqrt{logn/n})$.

The rest of the report is organized as follows: In Chapter 2, we discuss the literature related to our work. In Chapter 3, we provide formal definition of the risk measure UBSR followed by its estimation methods. In Chapter 4, we describe the optimization of UBSR using bandit feedback. In Chapter 5, we describe the optimization of UBSR using stochastic gradient descent. In Chapter 6, we provide concluding remarks.

CHAPTER 2

Literature Survey

2.1 Risk Estimation

The estimation and optimization of risk measures such as Mean-Variance, VaR, CVaR, CPT, UBSR have been explored in the past couple of decades. In [16], the authors provide a review of Monte Carlo methods for the estimation of VaR and CVaR. [28] deals with the optimization of Conditional Value-at-Risk. In [26], the authors derive concentration bounds for the estimation of CVaR for both light-tailed and heavy-tailed distributions. In [10], the authors make use of stochastic root finding and importance sampling schemes for the estimation of utility-based shortfall risk. [17] provide Monte Carlo techniques for estimating UBSR. They also provide framework for the optimization of UBSR. In [21], the authors explore the estimation and optimization of UBSR where the data arrives in an online fashion.

2.2 Risk-Aware Bandits

In [20], [15] and [12] the authors consider the CVaR optimization problem in a best-arm identification framework under a fixed budget. [2] and [29] focus on the optimization of empirical variance in the multi-armed bandits setting. Optimization of weighted bandits with the weight distortion function based on Cumulative Prospect Theory (CPT) has been dealt with in [14]. The StoROO algorithm developed in [30] is a modified version of Stochastic Optimistic Optimization (StoOO) proposed in [23]. StoOO deals with maximizing conditional expectation. In [30], the authors provide confidence bounds for optimizing conditional quantiles (VaR) and CVaR.

2.3 Risk Optimization using SGD

In [8], the author provides various results and theorems on convex optimization and gradient descent schemes. In [22], the authors derive non-asymptotic bounds for both last iterate SGD (Robins-Munro algorithm) and SGD where the iterates are averaged (Polyak-Rupert averaging) assuming both strongly and non strongly convex objectives. However, it is assumed that the gradient estimate is unbiased. [9], [3] consider finite-sample analysis of zeroth order stochastic approximation, but they assume zero-mean noise on the function measurements, which is not the case for UBSR optimization considered here. [4] and [24] consider stochastic approximation of an abstract objective function where the function measurements are biased, and the bias can be controlled through a batch size. In [25], the authors use the estimation scheme from [17] to establish concentration inequalities for UBSR estimation.

CHAPTER 3

Utility-Based Shortfall Risk

3.1 Definition

For the definition of the risk measure UBSR, a convex utility function $\ell(.)$ and a risk level λ need to be specified. An acceptance set $\mathcal{A}$ for a random variable X is defined as follows:

$$\mathcal{A} := \{X \in L^{\infty} : \mathbb{E}[\ell(-X)] \leq \lambda\} \tag{3.1}$$

where L^{∞} denotes the set of bounded random variables and the expectation is taken with respect to the distribution of the random variable X.

Using the definition of the acceptance set, for a given utility function and a risk level, the utility-based shortfall risk (UBSR) is defined as

$$SR_{\ell,\lambda}(X) := inf\{t \in \mathcal{R} : t + X \in \mathcal{A}\} \tag{3.2}$$

In financial terms, UBSR can be defined as the minimum cash needed to be added to the financial position to make the risk fall below the prescribed level i.e, make the new position acceptable.

An interesting thing which can be observed about UBSR is that by defining the utility function as $\ell(x) = 1_{\{x>0\}}$, Value-at-Risk can be written in terms of UBSR as follows:

$$VaR_{1-\lambda}(X) = inf\{t \in \mathcal{R} : \mathbb{E}[\ell(-t - X)] \leq \lambda\} \tag{3.3}$$

From the utility function used for VaR, it can be seen that VaR penalizes all positive losses equally whereas UBSR uses a non-decreasing convex function to penalize the loss resulting in large penalties for large losses. Two widely used utility functions are exponential function, $\ell(x) = exp(\beta x)$, $\beta > 0$ and the piecewise polynomial function, $\ell(x) = \eta^{-1}([x]^{+})^{\eta}$, $\eta > 1$.

3.2 UBSR Estimation using Sample Average Approximation

Define the function

$$g(t) := \mathbb{E}[\ell(-t - X)] - \lambda. \tag{3.4}$$

The following assumption on g is necessary for the next claim.

Assumption 1. *There exists t_l, t_u such that $g(t_l) > 0$ and $g(t_u) < 0$.*

Using the above assumption and convexity and monotonicity of $\ell(.)$, it is shown in [10] that $SR_{\ell,\lambda}(X)$ is the unique root of the equation $g(t^*) = 0$, i.e., $SR_{\ell,\lambda}(X) = t^*$. For the estimation of UBSR, [17] propose the following procedure based on Sample Average Approximation (SAA). Define $\xi = -X$ where ξ represents random loss. Given n i.i.d. samples $\xi_1, \xi_2, ..., \xi_m$ of ξ, the estimate of UBSR is the solution to the following optimization problem

$$\begin{aligned} \min_{t \in T} \quad & t \\ \text{subject to} \quad & \frac{1}{m}\sum_{j=1}^{m} \ell(\xi_j - t) \leq \lambda. \end{aligned}$$

If ℓ is increasing, The estimate t_m of UBSR is the unique root of the equation

$$\frac{1}{m}\sum_{j=1}^{m}\ell(\xi_j - t) = \lambda. \tag{3.5}$$

Bisection Search method as proposed in [17] can be used to solve (3.5).

Algorithm 1 Bisection Search

```
**Input:** risk level $\lambda$; $m$; Samples $\xi_1$, $\xi_2$,....,$\xi_m$;
**Define:** Utility function $\ell()$
**Initialization:** $t_l$ such that $g(t_l) > \lambda$; $t_u$ such that $g(t_u) < \lambda$; $e = 1$;
**while** $|e| > 10^{-6}$ **do**
    $t_i = (t_l + t_u)/2$
    $e = \lambda - \frac{1}{m}\sum_{j=1}^{m}\ell(\xi_j - t_i)$
    **if** $e < 0$ **then**
        $t_l = t_i$
    **else**
        $t_u = t_i$
    **end if**
**end while**
**Return** $t_i$
```

3.3 UBSR Estimation using Stochastic Approximation

Stochastic root-finding problems can be solved using stochastic approximation algorithms, see [6]. The following stochastic approximation update (Robins-Munro Algorithm) is proposed by [10] for estimating UBSR:

$$t_{k+1} = \Pi(t_k + a_k(\hat{g}(t_k))), \tag{3.6}$$

where a_k is the step-size, $\hat{g}(t_k) = \ell(\xi_k - t_k) - \lambda$ is an estimate of $g(t)$ obtained using the sample sequence $\{\xi_i\}$ and $\Pi : \mathbb{R} \to [t_l, t_u]$ is a projection operator.

CHAPTER 4

UBSR Optimization in a $\mathcal{X}$-Armed Bandits Framework

4.1 Problem Formulation

The optimization problem under the bandit framework is the problem of choosing the best arm which minimizes the simple regret. In contrast to K-armed bandits problem, instead of choosing the best arm out of a finite set of arms, the best arm, x is chosen from a continuous input space $\mathcal{X} \subset [0, 1]^D$. In each time step t, arm x_t is chosen and we obtain the value $f(x_t, \omega_t)$ where f is an unknown function outputting the value of the financial position for the chosen x_t and $\omega_t \in \Omega$ where Ω denotes the probability space representing uncontrollable variables. The distribution corresponding to $f(x, .)$ is denoted by $\mathbb{P}_x$. A risk measure h can be defined as some function ψ of the probability distribution $\mathbb{P}_x$ as $\Gamma(x) = \psi(\mathbb{P}_x)$. It is assumed that there exists at least one $x^* \in \mathcal{X}$ such that $\Gamma(x^*) = sup_{x \in \mathcal{X}} \Gamma(x)$. The goal is to minimize the simple regret $r_T = \Gamma(x^*) - \Gamma(x_T)$ with x_T the value returned after using a budget T.

4.2 Hierarchical Partitioning

In bandit algorithms, an Upper Confidence Bound(UCB) is maintained for each arm and the algorithm chooses the arm with the highest UCB in each round. Since in $\mathcal{X}$-armed bandits setting, arm needs to be chosen from a continuous input space

$\mathcal{X}$, a technique known as Hierarchical partitioning is used to partition the input space in each round and the candidate arms are the centers of all the existing partitions. Depending on certain conditions, a partition(cell) can be expanded into K sub-regions.

Let $\mathcal{P}_{h,j}$ denote j-th cell at depth h. Then

$$\mathcal{P}_{0,1} = \mathcal{X},\ \mathcal{P}_{h,j} = \bigcup_{i=0}^{K-1} \mathcal{P}_{h+1,j-i} \tag{4.1}$$

The following assumptions are made while applying the hierarchical partitioning:

Assumption 2. *There exists a decreasing sequence $\delta(h)$, such that for any $h \geq 0$ and for any cell $\mathcal{P}_{h,j}$, $sup_{x \in \mathcal{P}_{h,j}} ||x - x_{h,j}||_\infty \leq \delta(h)$ with $x_{h,j}$ the center of $\mathcal{P}_{h,j}$.*

Assumption 3. *There exists $\nu > 0$ such that every cell of depth h contains a ball of radius $\nu\delta(h)$.*

$\mathcal{T}_t$ denotes the resulting tree after having expanded some cells till time step t. The nodes of the tree correspond to different cells. $\mathcal{L}_t$ denotes the set of leaf nodes of $\mathcal{T}_t$. The Upper Confidence Bound(UCB) is defined using a piecewise constant function U. For all $x \in \mathcal{P}_{h,j}$, define $\bar{U}_{h,j}$ such that $U(x) = \bar{U}_{h,j}$.

The following smoothness property in the neighborhood of the global maxima is assumed:

$$\forall x \in \mathcal{X},\ \Gamma(x^*) - \Gamma(x) \leq \beta ||x - x^*||^\gamma \ with\ \beta, \gamma > 0 \tag{4.2}$$

In order to create confidence bounds for each cell, the algorithm samples the nodes in $\mathcal{L}_t$ at their centers. Then using the deviation inequalities corresponding to the chosen risk measure, $U_{h,j}$ is calculated. However this value is the UCB corresponding to only the center of the cell (h, j). In order to create UCB for the entire cell $\bar{U}_{h,j}$, a bias term $B_{h,j}$ is added. The resulting UCB is given as:

$$\bar{U}_{h,j} = U_{h,j} + B_{h,j},\ B_{h,j} = \hat{\beta}\delta(h)^{\hat{\gamma}},\ \hat{\beta} \geq \beta,\ \hat{\gamma} \leq \gamma \tag{4.3}$$

For each cell, the algorithm also needs a lower confidence bound(LCB) in order to provide guarantees on the value of Γ. It is denoted by $L_{h,j}$.

4.3 Stochastic Risk Optimistic Optimization (StoROO) Algorithm

The StoROO algorithm requires the following inputs: (i) error probability η, (ii) number of children K, (iii) time horizon T, (iv) $\hat{\beta}$, (v) $\hat{\gamma}$ and (vi) functions to calculate UCB and LCB. Initially, the input space is expanded into K sub-regions and sampled once. $\mathcal{L}_t$ represents the set of leaf nodes after t rounds. For each cell (h, j) in $\mathcal{L}_t$, $\bar{U}_{h,j}(t)$ is computed. The cell with the maximum $\bar{U}_{h,j}(t)$ is selected. It is denoted by $\mathcal{P}_{h_t,j_t}$. There are two possibilities for the selected cell. The algorithm can either sample again from the selected cell which results in reduction in variance or expand this cell which causes reduction in bias. The algorithm decides to expand the cell when the following condition holds:

$$U_{h_t,j_t} - L_{h_t,j_t} \leq \hat{\beta}\delta(h)^{\hat{\gamma}} \tag{4.4}$$

Denote the set of nodes having the highest LCB among the expanded nodes by $\mathscr{L}_T$. When the budget T gets over, StoROO returns the node with the highest $\hat{\Gamma}$ among the deepest nodes of $\mathscr{L}_T$.

Algorithm 2 StoROO

```
Input: error probability $\eta > 0$; number of children $K$; time horizon $T$; $\hat{\beta} > 0$; $\hat{\gamma} > 0$;
Define: UCB and LCB
Initialization: $n = 1$; $t = 1$;
Expand into $K$ sub-regions the root node $(0, 0)$ and sample one time each child
while $n \leq T$ do
    for $(h, j) \in \mathcal{L}_t$ do
        compute $\bar{U}_{h,j}(t)$
    end for
    Select $(\tilde{h}, \tilde{j}) = argmax_{(h,j)\in\mathcal{L}_t} \bar{U}_{h,j}(t)$
    Compute the LCB $L_{\tilde{h},\tilde{j}}(t)$
    if $U_{\tilde{h},\tilde{j}} - L_{\tilde{h},\tilde{j}} \leq \hat{\beta}\delta(h)^{\hat{\gamma}}$ then
        expand the node, remove $\tilde{h}, \tilde{j}$ from $\mathcal{L}_t$, add to $\mathcal{L}_t$ the $K$ sub-cells of $\mathcal{P}_{\tilde{h},\tilde{j}}$ and sample each new node once, $n = n + K$, $t = t + 1$
    else
        sample the state $x_t = x_{\tilde{h},\tilde{j}}$ and collect the observation $Y_{x_{h_t,j_t}}$, $n = n + 1$, $t = t + 1$
    end if
end while
Return the node according to the returning rule
```

4.4 Generic Regret Bound

Define the event $\mathcal{A}_\eta$ in the following way:

$$\mathcal{A}_\eta = \bigcap_{T \geq t \geq 1} \bigcap_{\mathcal{P}_{h,j} \in \mathcal{T}_t} \{U^{\eta}_{h,j}(t) \geq \Gamma(x_{h,j}), L^{\eta}_{h,j}(t) \leq \Gamma(x_{h,j})\} \tag{4.5}$$

The regret bounds can be derived only if one can construct the confidence bounds such that the probability of event $\mathcal{A}_\eta$ is at least $1 - \eta$.

The following definition of the event $\mathcal{B}_\eta$ and the vector of safe constants will be useful in deriving the regret bound based on the number of times a node has to be sampled before expansion.

Definition 1. *Let $m_{\eta,h}(\theta, \kappa, \alpha) = log(\theta T^2/\eta)\left(\frac{\kappa}{\hat{\beta}\delta(h)^{\hat{\gamma}}}\right)^{\alpha}$ and $N_{h,j}(t) = \sum_{s=1}^{t} \mathbf{1}_{X(s)\in\mathcal{P}_{h,j}}$, a vector of safe constants $\nu = (\theta, \kappa, \alpha)$ is composed of constants $\theta > 0$, $\kappa > 0$, and $\alpha > 0$ such that* *the event*

$$\mathcal{B}_\eta = \bigcap_{T \geq t \geq 1} \bigcap_{N_{h,j} \geq m_{\eta,h}(\theta,\kappa,\alpha)} \bigcap_{\mathcal{P}_{h,j} \in \mathcal{T}_t} \{U^\eta_{h,j}(t) - L^\eta_{h,j}(t) \leq \hat{\beta}\delta(h)^{\hat{\gamma}}\}$$

has a probability at least $1 - \eta$

The following definition of the ν-near optimality dimension is used to calculate the minimum depth reached by StoROO with a budget T.

Definition 2. *The* ν*-near optimality dimension is the smallest* $d \geq 0$ *such that for all* $\epsilon \geq 0$*, there exists* $C \geq 0$ *such that the maximal number of disjoint* $l_{\hat{\beta},\hat{\gamma}}$*-balls of radius* $\nu\epsilon$ *with center in* $\mathcal{X}_\epsilon$ *is less than* $C\epsilon^{-d}$*.*

In the above definition, $l_{\hat{\beta},\hat{\gamma}}$ is the Holderian semi-metric given by $l_{\hat{\beta},\hat{\gamma}}(x, x') = \hat{\beta}||x - x'||^{\hat{\gamma}}$. It is shown in [23] that the near optimality dimension is given by $d = D(1/\hat{\gamma} - 1/\gamma)$, where γ depends on the smoothness property of the function g under consideration as defined in (4.2) and D is the dimension of the parameter space $\mathcal{X}$. If one has the smoothness information about the function, i.e., if we know the value of γ, by setting $\hat{\gamma} = \gamma$, we can make $d = 0$.

The following theorem as proposed in [30] provides a generic regret bound for StoROO algorithm.

Theorem 1. *Assume that* $\delta(h) = c\rho^h$ *for some* $c \geq 0$ *and* $\rho < 1$*, and assume that* $\nu = (\theta, \kappa, \alpha)$*. Thus with probability* $\mathbb{P}(\mathcal{A}_\eta \cap \mathcal{B}_\eta)$*, the regret of StoROO is bounded as*

$$r_T \leq c_1 \left[\frac{log(\theta T^2/\eta)}{T}\right]^{\frac{1}{d+\alpha}} \quad with\ c_1 = 2\hat{\beta}\left[\frac{KC\kappa^\alpha[2\hat{\beta}]^{-d}}{(1 - \rho^{d\hat{\gamma}+\hat{\gamma}\alpha})}\right]^{\frac{1}{d+\alpha}} \tag{4.6}$$

The reader is advised to refer [30] for more details on the analysis of the algorithm.

4.5 Regret Bound for UBSR

In this section, we apply the StoROO algorithm for optimizing the risk measure UBSR. It involves using concentration inequalities derived for UBSR and to arrive at the upper and lower confidence bounds for StoROO algorithm. Using the UCB, LCB and the definition of the probability of event $\mathcal{B}_\eta$, we arrive at the vector of safe constants (θ, κ, α). Plugging the vector of safe constants into the generic regret bound expression, we get the regret bound for UBSR. Here we are dealing with the problem of minimizing UBSR. The algorithm is developed for maximizing $\Gamma(x)$. Hence we define $\Gamma(x) = -SR_\lambda(X)$.

The algorithmic procedure for optimizing UBSR involves running algorithm 2 with the UCB and LCB derived for UBSR. The pseudocode for the estimation of UBSR using Bisection Search is provided in algorithm 1.

For deriving confidence intervals, we use the deviation inequalities provided in [25].

Theorem 2. *Let Y be a r.v. satisfying the Bernstein's condition with parameters σ^2, b. Let the utility function in the definition of $SR_\lambda(Y)$ be L_1-Lipschitz. Let $\xi_{n,\lambda}$ be the solution to the constrained problem in* (3.5). *Then, for any $\frac{\epsilon}{L_1} > \frac{32\sigma^2}{\sqrt{n}}$*

$$\mathbb{P}(|\xi_{n,\lambda} - SR_\lambda(Y)| > \epsilon) \leq exp\left(-n\left(\frac{\epsilon}{L_1} - \frac{32\sigma^2}{\sqrt{n}}\right)^2\right) \tag{4.7}$$

The following proposition is the main contribution of our work.

Proposition 1. *For any $\eta > 0$, for all $h \geq 0$, for all $0 \leq j \leq K^h$ and for all $1 \leq t \leq T$,* *define*

$$U^{\eta}_{h,j}(t) = -\xi^t_{n,\lambda}(h,j) + L_1\left(\frac{\sqrt{log(T^2/\eta)} + 32\sigma^2}{\sqrt{N_{h,j}(t)}}\right) \tag{4.8}$$

$$L^{\eta}_{h,j}(t) = -\xi^t_{n,\lambda}(h,j) - L_1\left(\frac{\sqrt{log(T^2/\eta)} + 32\sigma^2}{\sqrt{N_{h,j}(t)}}\right) \tag{4.9}$$

with $\xi^t_{n,\lambda}(h,j)$ *as the solution to* (3.5)

Proof. Consider the event,

$$\xi_\eta = \{\forall h \geq 0, \forall 0 \leq j \leq K^h, 1 \leq t \leq T, |\xi^t_{n,\lambda}(h,j) - SR_\lambda(Y_{x_{h,j}})| \geq \epsilon^{\eta}_{N_{h,j}(t)}\}$$

$$\mathbb{P}(\xi_\eta) = \mathbb{P}(\forall h \geq 0, \forall 0 \leq j \leq K^h, 1 \leq t \leq T, |\xi^t_{n,\lambda}(h,j) - SR_\lambda(Y_{x_{h,j}})| \geq \epsilon^{\eta}_{N_{h,j}(t)})$$

Let $m \leq T$ be the total number of nodes present in $\mathcal{T}_t$ after the budget is exhausted. For some $1 \leq w \leq m$, ζ^s_w denote the s-th time when the cell w has been sampled. Let $Y_w(\zeta^s_w)$ denote the reward obtained by playing the arm x_w. Using this,

$$\begin{aligned}
&\mathbb{P}\left(|\xi^t_{n,\lambda}(h,j) - SR_\lambda(Y_{x_{h,j}})| \geq \epsilon^{\eta}_{N_{h,j}(t)}\right) \\
&\quad = \mathbb{P}\left(\frac{1}{N_{h,j}(t)}|inf\{z \in \mathbb{R} \mid \frac{1}{N_{h,j}(t)}\sum_{s=1}^{N_{h,j}(t)} l(Y_{h,j}(\zeta^s_{h,j}) - z) \leq \lambda\} - SR_\lambda(Y_{x_{h,j}})| \geq \epsilon^{\eta}_{N_{h,j}(t)}\right)
\end{aligned}$$

With this, we get:

$$\begin{aligned}
\mathbb{P}(\xi_\eta) &\leq \mathbb{P}(\exists 1 \leq w \leq T, \exists 1 \leq u \leq T, \mid inf\{z \in \mathbb{R} \mid \frac{1}{u}\sum_{s=1}^{u} l(Y_w(\zeta^s_w) - z) \leq \lambda\} - SR_\lambda(Y_{x_w})| \geq \epsilon^{\eta}_u) \\
&\leq \sum_{w=1}^{T}\sum_{u=1}^{T} \mathbb{P}(inf\{z \in \mathbb{R} \mid \frac{1}{u}\sum_{s=1}^{u} l(Y_w(\zeta^s_w) - z) \leq \lambda\} - SR_\lambda(Y_{x_w})| \geq \epsilon^{\eta}_u) \\
&\leq \sum_{w=1}^{T}\sum_{u=1}^{T} exp\left(-u\left(\frac{\epsilon^{\eta}_u}{L_1} - \frac{32\sigma^2}{\sqrt{u}}\right)^2\right)
\end{aligned}$$

Substituting $\epsilon_u^\eta = L_1\left(\frac{\sqrt{log(T^2/\eta)}+32\sigma^2}{\sqrt{u}}\right)$ we get,

$$\mathbb{P}(\xi_\eta) \leq \sum_{w=1}^{T}\sum_{u=1}^{T}\frac{\eta}{T^2} = \eta$$

□

We now state the regret bound of StoROO for UBSR optimization.

Theorem 3. *Assume that $\delta(h) = c\rho^h$ for some $c \geq 0$ and $\rho < 1$, and assume that $\nu = (\theta, \kappa, \alpha)$. Thus with probability $\mathbb{P}(\mathcal{A}_\eta \cap \mathcal{B}_\eta)$, the regret of StoROO for the optimization of utility-based shortfall risk is bounded as*

$$r_T \leq c_1\left[\frac{log(T^2/\eta)}{T}\right]^{\frac{1}{d+2}} \text{ with } c_1 = 2\hat{\beta}\left[\frac{KC\kappa^2[2\hat{\beta}]^{-d}}{(1-\rho^{d\hat{\gamma}+2\hat{\gamma}})}\right]^{\frac{1}{d+2}}$$
$$\text{and } \kappa = 2L_1\left(1+\frac{32\sigma^2}{\sqrt{log(T^2/\eta)}}\right) \tag{4.10}$$

Proof. Using the UCB and LCB, we obtain the vector of safe constants which can be plugged into the generic regret bound to get the regret bound for UBSR. The node (h, j) is expanded when $U^\eta_{h,j}(t) - L^\eta_{h,j}(t) \leq \hat{\beta}\delta(h)^{\hat{\gamma}}$. Substituting the values,

$$\hat{\beta}\delta(h)^{\hat{\gamma}} \geq 2L_1\left(\frac{\sqrt{log(T^2/\eta)}+32\sigma^2}{\sqrt{N_{h,j}(t)}}\right)$$
$$N_{h,j}(t) \geq \left(\frac{2L_1}{\hat{\beta}\delta(h)^{\hat{\gamma}}}\right)^2\left(\sqrt{log(T^2/\eta)}+32\sigma^2\right)^2$$
$$N_{h,j}(t) \geq \left(\frac{2L_1}{\hat{\beta}\delta(h)^{\hat{\gamma}}}\right)^2 log(T^2/\eta)\left(1+\frac{32\sigma^2}{\sqrt{log(T^2/\eta)}}\right)^2 \tag{4.11}$$

The node expansion is done when $N_{h,j}(t) \geq m_{\eta,h}(\theta, \kappa, \alpha)$ where

$$m_{\eta,h}(\theta,\kappa,\alpha) = log(\theta T^2/\eta)\left(\frac{\kappa}{\hat{\beta}\delta(h)^{\hat{\gamma}}}\right)^\alpha \tag{4.12}$$

Comparing (4.11) and (4.12), we get the vector of safe constants

$$(\theta, \kappa, \alpha) = \left(1,\ 2L_1\left(1 + \frac{32\sigma^2}{\sqrt{log(T^2/\eta)}}\right),\ 2\right) \tag{4.13}$$

Using the vector of safe constants and Theorem 1, we obtain the regret bound for UBSR. □

Assuming the knowledge of the function smoothness near optimum, we can set $d = 0$ and the regret bound turns out to be of the order $O\left(\sqrt{\frac{log(T^2)}{T}}\right)$. The order of the regret bound for UBSR optimization is comparable to the optimization of quantile (VaR) and CVaR as derived in [30].

CHAPTER 5

UBSR optimization using Stochastic Gradient Descent

In this chapter, we consider the problem of optimization of UBSR assuming convexity of $SR_\lambda(X(\theta))$ as a function of θ. It is also assumed that we only have access to the values $\{\xi_i\}$ taken from the distribution of $-X(\theta)$ in each iteration given a parameter θ. The Stochastic Gradient Descent (SGD) update for the optimization of UBSR is given as follows:

$$\theta_{k+1} = \Pi_\Theta(\theta_k - a_k h'_m(\theta_k)) \tag{5.1}$$

where a_k is a step-size parameter, $h'_m(\theta_k)$ is an estimate of $\frac{dSR_\lambda(\theta)}{d\theta}$ using m samples and Π_Θ is the projection on to the set Θ. Note that we don't have direct access to the gradient of UBSR. One has to compute an estimate of $\frac{dSR_\lambda(\theta)}{d\theta}$ using the samples $\{\xi_i, \xi_2, ..., \xi_m\}$ before running the update in (5.1).

In the next section, we describe the scheme for the estimation of UBSR derivative followed by a lemma on the rate at which the derivative estimate converges to the UBSR derivative.

5.1 Estimation of UBSR Derivative

The expression for the derivative of $SR_\lambda(X(\theta))$ with respect to θ is derived by [17].

$$\frac{dSR_\lambda(\theta)}{d\theta} = \frac{A(\theta)}{B(\theta)}, \tag{5.2}$$

where $A(\theta) = \mathbb{E}[\ell'(\xi(\theta) - SR_\lambda(\theta)))\xi'(\theta)]$, and $B(\theta) = \mathbb{E}[(\ell'(\xi(\theta) - SR_\lambda(\theta))]$.

[17] also provide a scheme for estimating the UBSR derivative. The expression for the same is given as follows:

$$h'_m(\theta) = \frac{A_m}{B_m}, \tag{5.3}$$

where $A_m(\theta) = \frac{1}{m}\sum_{i=1}^{m} \ell'(\xi_i(\theta) - t_m(\theta))\xi'_i(\theta),\ B_m(\theta) = \frac{1}{m}\sum_{i=1}^{m} \ell'(\xi_i(\theta) - t_m(\theta))$.

It is important to note that $A_m(\theta)$ and $B_m(\theta)$ are not unbiased estimates of $A(\theta)$ and $B(\theta)$, since the UBSR estimate $t_m(\theta)$ is biased. This results in $h'_m(\theta)$ being a biased estimate of $\frac{dSR_\lambda(\theta)}{d\theta}$.

The analysis on the consistency property of the UBSR derivative estimate is done by [21]. Here we provide the assumptions required for the analysis. Recall that $\xi = -X$ and $g(t) := \mathbb{E}[\ell(-t - X)] - \lambda$.

Assumption 4. $\sup_{\theta\in\Theta} E(\xi(\theta)^2) \leq M_1$.

Assumption 4 requires the second moment of ξ to be bounded for all $\theta \in \Theta$ which is necessary to ensure that the sample-based estimate is asymptotically consistent.

Assumption 5. *Assumptions 1 holds for every $\theta \in \Theta$.*

Assumption 5 is necessary so that the UBSR can be estimated by solving (3.5).

Assumption 6. *The partial derivatives $\partial\ell(\xi(\theta - t(\theta))))/\partial\theta, \partial\ell(\xi(\theta) - t(\theta))/\partial t$ exist w.p. 1, and there exists a $\beta_1 > 0$ such that*

$$E\left[(\ell'(\xi(\theta) - SR_\lambda(\theta)))^2\right] \leq \beta_1 < \infty, \forall\theta \in \Theta.$$

Assumption 6 is necessary so that A_m and B_m converge asymptotically to A and B respectively and the error is normally distributed, see [17].

Assumption 7. *The loss function* $\ell(\cdot)$ *satisfies w.p.* 1

$$|\ell'(\xi(\theta) - t)| \leq L_1, |\ell''(\xi(\theta) - t)| \leq L_2, \forall (\theta, t) \in \Theta \times [t_l, t_u].$$

Assumption 7 is required for deriving the expression in (5.2).

Assumption 8. *The loss function* $\ell(\cdot)$ *is twice differentiable, and for any* $\theta \in \Theta$, $\ell'(\xi(\theta) - SR_\lambda(\theta)) > \eta$ *w.p.* 1.

Assumption 9. $\sup_{\theta \in \Theta} |\xi'(\theta)| \leq M_2$, *and* ξ' *is* L_3*-Lipschitz for all* $\theta \in \Theta$ *w.p.* 1.

Assumptions 7–9 are necessary to ensure that $\ell'(\xi(\theta) - SR_\lambda(\theta))\xi'(\theta)$ is Lipschitz. We now state the lemma corresponding to the consistency property of UBSR derivative as provided in [21].

Lemma 1. *Under Assumptions 4–9, for all* $m \geq 1$, *the UBSR derivative estimator* (5.3) *satisfies*

$$E\left|h'_m(\theta) - \frac{dSR_\lambda(\theta)}{d\theta}\right| \leq \frac{C_1}{\sqrt{m}}, \quad \textit{and} \quad E\left|h'_m(\theta) - \frac{dSR_\lambda(\theta)}{d\theta}\right|^2 \leq C_2,$$

where $C_1 = \frac{\sqrt{\beta_1}\varsigma M_1(L_1L_3 + 2M_2L_2)}{\eta^2}$ *and* $C_2 = \frac{16\beta_1^2 M_2^2}{\eta^4}$. *Here the constants* $\beta_1, L_1, L_2, L_3, M_1, M_2$ *and* η *are as specified in assumptions 4–9 above and* ς *is a universal constant.*

5.2 Non-Asymptotic Bound for UBSR Optimization Assuming Convexity

We make the following additional assumptions regarding the compactness of set Θ and convexity of $SR_\lambda(\theta)$.

Assumption 10. *The set Θ satisfies $|\theta_1 - \theta_2| \leq D$, $\forall\ \theta_1, \theta_2 \in \Theta$, for some $D > 0$.*

Assumption 11. *For any $\theta \in \Theta$, the function $h(\theta) = SR_\lambda(\theta)$ satisfies $h''(\theta) \geq 0$.*

Let n be the number of iterations for which SGD is run. It is assumed that n is known beforehand. The horizon n is split into p phases.

Let $p := \inf\{i : n \cdot 2^{-i} \leq 1\}$,

$$n_i := n - \lceil n \cdot 2^{-i} \rceil,\ 0 \leq i \leq p, \text{ and } n_{p+1} := n. \tag{5.4}$$

For notational conveninece, let $h(\theta) = SR_\lambda(\theta)$ and $h'(\theta) = \frac{dSR_\lambda(\theta)}{d\theta}$.

Theorem 4. *Suppose Assumptions 4–11 hold and suppose the update in* (5.1) *is performed for n iterations with step-size a_k and batch size m_k set as follows:*

$$a_k := \frac{a_0 \cdot 2^{-i}}{\sqrt{n}},\ \textit{and}\ m_k := 2^i \cdot n, \tag{5.5}$$

for some constant a_0 when $n_i < k \leq n_{i+1}$, $0 \leq i \leq p$ with n_i, p as defined in (5.4). *Then for any $n \geq 4$,*

$$\mathbb{E}[h(\theta_n) - h(\theta^*)] \leq \frac{\mathcal{K}_1}{\sqrt{n}} + \frac{\mathcal{K}_2}{n}, \tag{5.6}$$

where $\mathcal{K}_1 = 4D^2/a_0 + 39DC_1 + (10C_2 + 11B^2)a_0$, $\mathcal{K}_2 = 16a_0BC_1$ and $B = L_1M_2/\eta$.

Proof. The proof technique is similar to the one used in [4]. However, in our case the expression for the gradient estimate of UBSR is known. The bias term in the gradient estimate depends on the batch size used to estimate UBSR (see Lemma 1).

Before proving the claim of Theorem 4, we state and prove the following lemmas.

Lemma 2. *Under Assumptions 4–9, for all* $m \geq 1$,

$$\mathbb{E}[h'_m(\theta)^2] \leq C_2 + \frac{2BC_1}{\sqrt{m}} + B^2. \tag{5.7}$$

Proof. The proof of this lemma follows directly from Lemma 1. Let $h'(\theta) = \frac{dSR_\lambda(\theta)}{d\theta}$. We first bound $|h'(\theta)|$ as follows:

$$\begin{aligned} |h'(\theta)| &= \frac{|\mathbb{E}[(l'(\xi(\theta) - h(\theta))\xi'(\theta)]|}{|\mathbb{E}[(l'(\xi(\theta) - h(\theta))]|} \leq \frac{|\mathbb{E}[(l'(\xi(\theta) - h(\theta))\xi'(\theta)]|}{\eta} \\ &\leq \frac{|\mathbb{E}[(l'(\xi(\theta) - h(\theta))|\xi'(\theta)|]|}{\eta} \leq \frac{L_1 M_2}{\eta} = B. \end{aligned} \tag{5.8}$$

Using the fact that $|x| - |y| \leq |x - y|$ for any $x, y \in \mathbb{R}$ followed by an application of Lemma 1, we obtain

$$\mathbb{E}[|h'_m(\theta)|] \leq \mathbb{E}[|h'_m(\theta) - h'(\theta)|] + \mathbb{E}[|h'(\theta)|] \leq \frac{C_1}{\sqrt{m}} + |h'(\theta)|. \tag{5.9}$$

Using $(|x| - |y|)^2 \leq (x - y)^2$ for any $x, y \in \mathbb{R}$, we obtain

$$\begin{aligned} \mathbb{E}[h'_m(\theta)^2] &\leq \mathbb{E}[(h'_m(\theta) - h'(\theta))^2] + 2\mathbb{E}[|h'_m(\theta)|]|h'(\theta)| - \mathbb{E}[h'(\theta)^2] \\ &\leq C_2 + 2\left(\frac{C_1}{\sqrt{m}} + |h'(\theta)|\right)|h'(\theta)| - h'(\theta)^2 \\ &= C_2 + 2\frac{C_1}{\sqrt{m}}|h'(\theta)| + h'(\theta)^2 \leq C_2 + \frac{2BC_1}{\sqrt{m}} + B^2, \end{aligned}$$

where the second inequality follows from Lemma 1 and (5.9). The last inequality follows from (5.8). □

Lemma 3. *Suppose Assumptions 4–11 hold. Suppose that the update in* (5.1) *is performed* *for n steps with step-size sequence* $\{a_k\}_{k=1}^n$. *Then for any* $1 < k_0 < k_1 \leq n$,

$$\sum_{k=k_0}^{k_1} 2a_k \mathbb{E}[h(\theta_k) - h(\theta_{k_0})] \leq \sum_{k=k_0}^{k_1} (2a_k D\mathcal{A}_k + a_k^2 \mathcal{B}_k), \tag{5.10}$$

where $\mathcal{A}_k = \frac{C_1}{\sqrt{m_k}}$, $\mathcal{B}_k = C_2 + 2B\mathcal{A}_k + B^2$.

Proof. Let $\delta_k = h'_m(\theta_k) - h'(\theta_k)$ and $\zeta_k = |\theta_k - \theta_{k_0}|$. Using the expression for gradient update in (5.1),

$$\begin{aligned}
\zeta_{k+1}^2 &= (\Pi_\Theta(\theta_k - a_k h'_m(\theta_k)) - \theta_{k_0})^2 \\
&\leq (\theta_k - a_k h'_m(\theta_k) - \theta_{k_0})^2 \qquad (5.11) \\
&= \zeta_k^2 - 2a_k h'_m(\theta_k)(\theta_k - \theta_{k_0}) + a_k^2 h'_m(\theta_k)^2 \\
&= \zeta_k^2 - 2a_k(\delta_k + h'(\theta_k))(\theta_k - \theta_{k_0}) + a_k^2 h'_m(\theta_k)^2 \\
&= \zeta_k^2 - 2a_k\delta_k(\theta_k - \theta_{k_0}) - 2a_k h'(\theta_k)(\theta_k - \theta_{k_0}) + a_k^2 h'_m(\theta_k)^2.
\end{aligned}$$

The inequality in (5.11) holds because θ_{k_0} belongs to set Θ and the distance of any θ outside the set from θ_{k_0} would be less than or equal to the distance of its corresponding projection, $\Pi_\Theta(\theta)$ from θ_{k_0}. Taking expectation on both sides and using Lemma 2, we obtain

$$\begin{aligned}
\mathbb{E}[\zeta_{k+1}^2] &\leq \mathbb{E}[\zeta_k^2] - 2a_k\mathbb{E}[h'(\theta_k)(\theta_k - \theta_{k_0})] - 2a_k\mathbb{E}[\delta_k(\theta_k - \theta_{k_0})] + a_k^2[C_2 + \frac{2BC_1}{\sqrt{m_k}} + B^2] \\
&\leq \mathbb{E}[\zeta_k^2] - 2a_k\mathbb{E}[h'(\theta_k)(\theta_k - \theta_{k_0})] + 2a_k\frac{C_1}{\sqrt{m_k}}|\theta_k - \theta_{k_0}| + a_k^2[C_2 + \frac{2BC_1}{\sqrt{m_k}} + B^2] \\
&= \mathbb{E}[\zeta_k^2] - 2a_k\mathbb{E}[h'(\theta_k)(\theta_k - \theta_{k_0})] + 2a_k\mathcal{A}_k\zeta_k + a_k^2[C_2 + 2B\mathcal{A}_k + B^2] \\
&\leq \mathbb{E}[\zeta_k^2] - 2a_k\mathbb{E}[h(\theta_k) - h(\theta_{k_0})] + 2a_k\mathcal{A}_k\zeta_k + a_k^2\mathcal{B}_k,
\end{aligned}$$

where the second inequality follows from Lemma 1 where as the last inequality follows from the convexity assumption. Rearranging the terms,

$$2a_k\mathbb{E}[h(\theta_k) - h(\theta_{k_0})] \leq \mathbb{E}[\zeta_k^2] - \mathbb{E}[\zeta_{k+1}^2] + 2a_k\mathcal{A}_k\zeta_k + a_k^2\mathcal{B}_k.$$

By taking summation over $k = k_0$ to k_1 and using 10 to bound ζ_k with D, we get (5.10). □

Lemma 4. *Suppose Assumptions 4–11 hold. Then, with $a_k = a$ and $m_k = m$, $\forall k \geq 1$,*

$$\sum_{k=1}^{n} \mathbb{E}[h(\theta_k) - h(\theta^*)] \leq \frac{D^2}{2a} + 2nD\mathcal{A} + \frac{naB^2}{2}, \tag{5.12}$$

where $\mathcal{A} = \frac{C_1}{\sqrt{m}}$.

Proof. Let $\delta_k = h'_m(\theta_k) - h'(\theta_k)$ and $\rho_{k+1} = \theta_k - a_k(h'(\theta_k) + \delta_k)$. Using convexity of $h(\theta)$, we obtain

$$\begin{aligned}
h(\theta_k) - h(\theta^*) \leq h'(\theta_k)(\theta_k - \theta^*) &= \left(\frac{\theta_k - \rho_{k+1}}{a_k} - \delta_k\right)(\theta_k - \theta^*) \\
&= \frac{1}{a_k}(\theta_k - \rho_{k+1} - a_k\delta_k)(\theta_k - \theta^*) \\
&= \frac{1}{2a_k}\left((\theta_k - \theta^*)^2 + (\theta_k - \rho_{k+1} - a_k\delta_k)^2 - (\rho_{k+1} - \theta^* + a_k\delta_k)^2\right) \qquad (5.13) \\
&= \frac{1}{2a_k}\left((\theta_k - \theta^*)^2 - (\rho_{k+1} - \theta^* + a_k\delta_k)^2\right) + \frac{a_k}{2}h'(\theta_k)^2,
\end{aligned}$$

where the equality in (5.13) is obtained using $a \cdot b = \frac{1}{2}(a^2 + b^2 - (a-b)^2)$. Substituting $h'(\theta_k)^2 \leq B^2$, we obtain

$$\begin{aligned}
h(\theta_k) - h(\theta^*) &\leq \frac{1}{2a_k}\left((\theta_k - \theta^*)^2 - (\rho_{k+1} - \theta^*)^2 - a_k^2\delta_k^2 - 2a_k(\rho_{k+1} - \theta^*)\delta_k\right) + \frac{a_k}{2}B^2 \\
&\leq \frac{1}{2a_k}\left((\theta_k - \theta^*)^2 - (\rho_{k+1} - \theta^*)^2 - 2a_k(\rho_{k+1} - \theta^*)\delta_k\right) + \frac{a_k}{2}B^2.
\end{aligned}$$

Taking expectations, and using $(\rho_{k+1} - \theta^*)^2 \geq (\theta_{k+1} - \theta^*)^2$, we obtain

$$\begin{aligned}\mathbb{E}[h(\theta_k) - h(\theta^*)] &\leq \frac{1}{2a_k}\Big(\mathbb{E}[(\theta_k - \theta^*)^2] - \mathbb{E}[(\theta_{k+1} - \theta^*)^2] - 2a_k\mathbb{E}[|\theta_{k+1} - \theta^*||\delta_k|]\Big) + \frac{a_k}{2}B^2 \\ &\leq \frac{1}{2a_k}\Big(\mathbb{E}[(\theta_k - \theta^*)^2] - \mathbb{E}[(\theta_{k+1} - \theta^*)^2] + 2a_k\mathcal{A}_k\mathbb{E}[|\theta_{k+1} - \theta^*|]\Big) + \frac{a_k}{2}B^2.\end{aligned} \tag{5.14}$$

By summing (5.14) over k, and using $a_k = a$ and $m_k = m$ along with the inequality $|\theta_k - \theta^*| \leq D, \ \forall k \geq 1$, we obtain (5.12). □

Proof of Theorem 4:

For $0 \leq i \leq p+1$, define ν_i as follows:

$$\nu_i = \arg\inf_{n_i < k \leq n_{i+1}} \mathbb{E}[h(\theta_k)], \ i \in [p+1], \text{ and } \nu_0 = \arg\inf_{\lceil \frac{n}{4} \rceil < k \leq n_1} \mathbb{E}[h(\theta_k)]. \tag{5.15}$$

The horizon n is split into p phases with each phase having a constant step-size and batch-size. We need to show that the final iterate θ_n is close to optima θ^*. Using $\nu_{p+1} = n$, we obtain

$$\mathbb{E}[h(\theta_n)] = \mathbb{E}[h(\theta_{\nu_0})] + \sum_{i=0}^{p} \mathbb{E}[h(\theta_{\nu_{i+1}}) - h(\theta_{\nu_i})]. \tag{5.16}$$

In order to bound $\mathbb{E}[h(\theta_{\nu_{i+1}}) - h(\theta_{\nu_i})]$, consider the case when $i \geq 1$. Using Lemma 3 with $k_0 = \nu_i$ and $k_1 = n_{i+2}$, we obtain

$$\begin{aligned}\frac{\sum_{k=\nu_i}^{n_{i+2}} 2a_k\mathbb{E}[h(\theta_k) - h(\theta_{\nu_i})]}{n_{i+2} - \nu_i + 1} &\leq \frac{\sum_{k=\nu_i}^{n_{i+2}}(2a_kD\mathcal{A}_k + a_k^2\mathcal{B}_k)}{n_{i+2} - \nu_i + 1} \\ &\leq 2a_{n_i+1}D\mathcal{A}_{n_i+1} + a_{n_i+1}^2\mathcal{B}_{n_i+1}.\end{aligned} \tag{5.17}$$

The inequality in (5.17) follows from the fact that a_k is a non increasing sequence and m_k is a non decreasing sequence resulting in $\mathcal{A}_k$ and $\mathcal{B}_k$ being non increasing sequences as well. Also note that $\nu_i \geq n_i + 1$. Now we define the step-size a_k and the batch size m_k as some polynomial function of n as follows:

$$a_k = a_0 \frac{2^{-i}}{n^{\alpha_1}}, \text{ and } m_k = 2^i \cdot n^{\alpha_2}, \tag{5.18}$$

for some constant a_0 and some positive constants α_1 and α_2 when $n_i < k \leq n_{i+1}$, $0 \leq i \leq p$. Substituting a_k and m_k in (5.17), we get

$$\frac{\sum_{k=\nu_i}^{n_{i+2}} 2a_k \mathbb{E}[h(\theta_k) - h(\theta_{\nu_i})]}{n_{i+2} - \nu_i + 1} \leq \frac{2DC_1 a_0 2^{-3i/2}}{n^{\alpha_1 + \alpha_2/2}} + \frac{a_0^2 2^{-2i}}{n^{2\alpha_1}} \left[C_2 + \frac{2BC_1}{2^{i/2} n^{\alpha_2/2}} + B^2 \right]. \tag{5.19}$$

Now we provide a lower bound for the expression on the left hand side of (5.19). Using $\mathbb{E}[h(\theta_k) - h(\theta_{\nu_i})] \geq 0$ whenever $n_i < k \leq n_{i+1}$. Therefore

$$\begin{aligned}
\frac{\sum_{k=\nu_i}^{n_{i+2}} 2a_k \mathbb{E}[h(\theta_k) - h(\theta_{\nu_i})]}{n_{i+2} - \nu_i + 1} &\geq \frac{\sum_{k=n_{i+1}+1}^{n_{i+2}} 2a_k \mathbb{E}[h(\theta_k) - h(\theta_{\nu_i})]}{n_{i+2} - \nu_i + 1} \\
&\geq 2a_{n_{i+2}} \frac{n_{i+2} - n_{i+1}}{n_{i+2} - n_i} \mathbb{E}[h(\theta_{\nu_{i+1}}) - h(\theta_{\nu_i})] \\
&\geq \frac{2a_{n_{i+2}}}{5} \mathbb{E}[h(\theta_{\nu_{i+1}}) - h(\theta_{\nu_i})] \\
&= \frac{2^{-i} a_0}{5n^{\alpha_1}} \mathbb{E}[h(\theta_{\nu_{i+1}}) - h(\theta_{\nu_i})],
\end{aligned} \tag{5.20}$$

where the second inequality follows from the assumption $\mathbb{E}[h(\theta_{\nu_{i+1}}) - h(\theta_{\nu_i})] \geq 0$, and the fact that $n_{i+2} - n_{i+1} \geq n_{i+2} - \nu_i + 1$. The last inequality follows from Lemma 4 of [4]. Combining the inequalities in (5.19) and (5.20), we obtain

$$\mathbb{E}[h(\theta_{\nu_{i+1}}) - h(\theta_{\nu_i})] \leq \frac{10DC_1 2^{-i/2}}{n^{\alpha_2/2}} + \frac{5a_0 2^{-i}}{n^{\alpha_1}} \left[C_2 + \frac{2BC_1}{2^{i/2} n^{\alpha_2/2}} + B^2 \right]. \tag{5.21}$$

The proof for the case when $i = 0$ is similar to the above proof resulting in the same inequality except that $i = 0$. Substituting the inequality (5.21) in (5.16), we obtain

$$\begin{aligned}
\mathbb{E}[h(\theta_n)] &\leq \mathbb{E}[h(\theta_{v_0})] + \sum_{i=0}^{p} \left(\frac{10DC_1 2^{-i/2}}{n^{\alpha_2/2}} + \frac{5a_0 2^{-i}}{n^{\alpha_1}} \left[C_2 + \frac{2BC_1}{2^{i/2} n^{\alpha_2/2}} + B^2 \right] \right) \\
&\leq \mathbb{E}[h(\theta_{v_0})] + \sum_{i=0}^{\infty} \left(\frac{10DC_1 2^{-i/2}}{n^{\alpha_2/2}} + \frac{5a_0 2^{-i}}{n^{\alpha_1}} \left[C_2 + \frac{2BC_1}{2^{i/2} n^{\alpha_2/2}} + B^2 \right] \right) \\
&= \mathbb{E}[h(\theta_{v_0})] + \frac{10DC_1}{n^{\alpha_2/2}(1 - 1/\sqrt{2})} + \frac{5(C_2 + B^2)a_0}{n^{\alpha_1}(1 - 1/2)} + \frac{10BC_1 a_0}{n^{\alpha_1 + \alpha_2/2}(1 - 2^{-3/2})} \\
&\leq \inf_{\lceil \frac{n}{4} \rceil \leq k \leq n_1} \mathbb{E}[h(\theta_k)] + \frac{35DC_1}{n^{\alpha_2/2}} + \frac{10(C_2 + B^2)a_0}{n^{\alpha_1}} + \frac{16BC_1 a_0}{n^{\alpha_1 + \alpha_2/2}}. \qquad (5.22)
\end{aligned}$$

For $k \leq n_1$, $a_k = \frac{a_0}{n^{\alpha_1}}$ and $m_k = n^{\alpha_2}$. Using the fact that infimum is smaller than the weighted average, we get

$$\begin{aligned}
\inf_{\lceil \frac{n}{4} \rceil \leq k \leq n_1} \mathbb{E}[h(\theta_k) - h(\theta^*)] &\leq \frac{1}{n_1 - \lceil \frac{n}{4} \rceil + 1} \sum_{k = \lceil \frac{n}{4} \rceil}^{n_1} \mathbb{E}[h(\theta_k) - h(\theta^*)] \\
&\leq \frac{2}{n_1} \sum_{k=1}^{n_1} \mathbb{E}[h(\theta_k) - h(\theta^*)] \qquad (5.23) \\
&\leq \frac{2}{n_1} \left[\frac{D^2 n^{\alpha_1}}{2a_0} + \frac{2n_1 DC_1}{n^{\alpha_2/2}} + \frac{n_1 a_0 B^2}{2} \right] \qquad (5.24) \\
&\leq \frac{4D^2}{a_0 n^{1-\alpha_1}} + \frac{4DC_1}{n^{\alpha_2/2}} + \frac{a_0 B^2}{n^{\alpha_1}}, \qquad (5.25)
\end{aligned}$$

where (5.23) follows from $n_1 \leq 2(n_1 - \lceil \frac{n}{4} \rceil + 1)$, (5.24) follows from Lemma 4 and (5.25) follows from the fact that $n_1 \geq \frac{n}{4}$. Plugging (5.25) in (5.22), we obtain the following:

$$\mathbb{E}[h(\theta_n) - h(\theta^*)] \leq \frac{4D^2}{a_0 n^{1-\alpha_1}} + \frac{39DC_1}{n^{\alpha_2/2}} + \frac{(10C_2 + 11B^2)a_0}{n^{\alpha_1}} + \frac{16BC_1 a_0}{n^{\alpha_1 + \alpha_2/2}}. \qquad (5.26)$$

The values for α_1 and α_2 which will result in the tightest bound are 1/2 and 1 respectively. Substituting these values, we get the main claim of Theorem 4. □

5.3 Non-Asymptotic Bound for UBSR Optimization with Markov Sampling

The non-asymptotic bound for UBSR optimization provided in the previous section assumed that the samples used to estimate UBSR, $\xi_1, \xi_2, ..., \xi_m$ are i.i.d samples obtained from the distribution of $-X$. In this section, we generalize the bound for the case when the samples are obtained from a Markov chain with stationary distribution μ and transition kernel (Markov kernel) P. The Markov chain under consideration here has a continuous state space and is supported on a compact set $K \subset \mathbb{R}$.

Definition 3. ***(Empirical stationary distribution)*** *Let $\xi_0, \xi_1, ...$ be a Markov chain with stationary distribution μ. For $m \in \mathbb{N}$, empirical distribution μ_m is given by*

$$\mu_m = \frac{1}{m} \sum_{i=1}^{m} \delta_{\xi_i}, \tag{5.27}$$

where δ_{ξ_i} is the Kronecker delta function.

It can be shown that μ_m converges to μ as $m \to \infty$ under suitable conditions.

Definition 4. ***(1-Wasserstein Distance)*** *Let μ and ν be probability measures on $\mathbb{R}$. Wasserstein distance of order 1 between μ and ν is given by*

$$W_1(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |x - y| \, d\pi(x, y), \tag{5.28}$$

where $\Pi(\mu, \nu)$ is defined as a set of all couplings between μ and ν.

In order to derive the non-asymptotic bound for the optimization of UBSR, it is required to first obtain the convergence rate of the UBSR derivative estimate in case of Markov sampling. For that, we need the rate of convergence of empirical distribution μ_m to the stationary distribution μ in expectation with respect to the 1-Wasserstein distance. We refer to [27] for the same. Here, we state the necessary assumption and the rate of convergence result provided there.

The following assumption is required in order to provide the rate of convergence of empirical distribution.

Assumption 12. *There are constants $D \geq 1$ and $\kappa \in (0,1)$ such that*

$$W_1(P^m(x,.), P^m(y,.)) \leq D\kappa^m|x-y| \tag{5.29}$$

for all $m \in \mathbb{N}$ and $x, y \in \mathbb{R}$.

The above assumption can be understood in this way: two Markov chains starting at x and y can be coupled in such a way that they approach each other as $m \to \infty$. The following lemma is obtained from Theorem 1.1 of [19].

Lemma 5. *Suppose Assumption 12 holds and the Markov chain is supported on a compact set $K \subset \mathbb{R}$, then there is a constant C_3 depending on K and D such that for all m large enough*

$$\mathbb{E}[W_1(\mu, \mu_m)] \leq C_3\sqrt{\frac{log((1-\kappa)m)}{(1-\kappa)m}} \tag{5.30}$$

Next, we derive the consistency property of UBSR derivative estimate with Markov sampling in a manner similar to the derivation of Lemma 1.

Lemma 6. *Suppose the samples used to estimate UBSR and its derivative follow a Markov chain with stationary distribution μ and transition kernel P and supported on a compact set $K \subset \mathbb{R}$. Under Assumptions 4–9 and 12, the UBSR derivative estimator* (5.3) *satisfies* *for all m large enough*

$$\mathbb{E}\left|h_m'(\theta) - \frac{dSR_\lambda(\theta)}{d\theta}\right| \leq C_4\sqrt{\frac{log((1-\kappa)m)}{m}}, \quad and \quad \mathbb{E}\left|h_m'(\theta) - \frac{dSR_\lambda(\theta)}{d\theta}\right|^2 \leq C_2,$$

where $C_4 = \frac{C_3\sqrt{\beta_1}(L_1L_3+2M_2L_2)}{\sqrt{1-\kappa}\eta^2}$ *and* $C_2 = \frac{16\beta_1^2M_2^2}{\eta^4}$*. Here the constants* $\beta_1, L_1, L_2, L_3, M_1, M_2,$ η*,* C_3 *and* κ *are as specified in assumptions 4–9 and 12.*

Proof. For some $t \in [t_l, t_u]$, define

$$u_m(t) = \frac{1}{m}\sum_{i=1}^{m}\ell'(\xi_i(\theta) - t), \text{ and } u(t) = \mathbb{E}[\ell'(\xi(\theta) - t)].$$

One can write $u(t)$ and $u_m(t)$ in the following manner:

$$u_m(t) = \int \ell' d\mu_m, \text{ and } u(t) = \int \ell' d\mu.$$

Since ℓ' is L_2 Lipschitz, using Assumption 7, we get

$$|u_m(t) - u(t)| \leq L_2 W_1(\mu_m, \mu), \tag{5.31}$$

Using Lemma 5, we obtain

$$\mathbb{E}\,|u_m(t) - u(t)| \leq L_2C_3\sqrt{\frac{\log((1-\kappa)m)}{(1-\kappa)m}}, \tag{5.32}$$

Next, we define

$$\tilde{u}_m(t) = \frac{1}{m}\sum_{i=1}^{m}\ell'\left(\xi_i(\theta) - t\right)\xi'(\theta), \text{ and } \tilde{u}(t) = E[\ell'(\xi(\theta) - t)\xi'(\theta)].$$

In a similar manner, we get

$$E\left|\tilde{u}_m(t) - \tilde{u}(t)\right| \leq (L_1L_3 + M_2L_2)C_3\sqrt{\frac{\log((1-\kappa)m)}{(1-\kappa)m}}. \tag{5.33}$$

$$\begin{aligned}
&E\left|h'_m(\theta) - \frac{dSR_\lambda(\theta)}{d\theta}\right| = E\left|\frac{A_m(\theta)}{B_m(\theta)} - \frac{A(\theta)}{B(\theta)}\right| \\
&\leq \frac{|B(\theta)|E[|A_m(\theta) - A(\theta)|] + |A(\theta)|E[|B_m(\theta) - B(\theta)|]}{\eta^2} \\
&\leq \left(\frac{|B(\theta)|\sup_{t\in[t_l,t_u]} E|\tilde{u}_m(t) - \tilde{u}(t)|]}{\eta^2} + \frac{|A(\theta)|\sup_{t\in[t_l,t_u]} E[|u_m(t) - u(t)|]}{\eta^2}\right) \\
&\leq \frac{C_3\sqrt{\beta_1}(L_1L_3 + 2M_2L_2)\sqrt{\log((1-\kappa)m)}}{\sqrt{(1-\kappa)m}\eta^2}.
\end{aligned}$$

The proof of the second claim is same as that for Lemma 1 and hence omitted here. □

The following theorem provides non-asymptotic bound for UBSR optimization in case of Markov sampling.

Theorem 5. *Suppose the samples used to estimate UBSR and its derivative follow a Markov chain with stationary distribution μ, transition kernel P and supported on a compact set $K \subset \mathbb{R}$. Suppose Assumptions 4–12 hold and suppose the update in* (5.1) *is performed for n iterations with step-size a_k and batch size m_k set as follows:*

$$a_k := \frac{a_0 \cdot 2^{-i}}{\sqrt{n}}, \text{ and } m_k := 2^i \cdot n, \tag{5.34}$$

for some constant a_0 when $n_i < k \leq n_{i+1}$, $0 \leq i \leq p$ with n_i, p as defined in (5.4). *Then for any $n \geq 4$,*

$$\mathbb{E}[h(\theta_n) - h(\theta^*)] \leq \frac{\mathcal{K}_3}{\sqrt{n}} + \frac{\mathcal{K}_4}{n}, \tag{5.35}$$

where $\mathcal{K}_3 = 4D^2/a_0 + DC_4(39\sqrt{log((1-\kappa)n)} + 35) + (10C_2 + 11B^2)a_0$*,*
$\mathcal{K}_4 = a_0BC_4(16\sqrt{log((1-\kappa)n)} + 6)$ *and* $B = L_1M_2/\eta$*.*

Proof. The proof follows similarly to the proof of Theorem 4. Throughout the proof until equation (5.21) constant C_1 is replaced by $C_4\sqrt{\log((1-\kappa)m)}$. (5.21) is rewritten in case of Markov sampling as

$$\mathbb{E}[h(\theta_{v_{i+1}}) - h(\theta_{v_i})] \leq \frac{10DC_4\sqrt{\log((1-\kappa)2^i n^{\alpha_2})}2^{-i/2}}{n^{\alpha_2/2}} + \frac{5a_0 2^{-i}}{n^{\alpha_1}}\left[C_2 + \frac{2BC_4\sqrt{\log((1-\kappa)2^i n^{\alpha_2})}}{2^{i/2}n^{\alpha_2/2}} + B^2\right]. \quad (5.36)$$

Substituting the inequality (5.36) in (5.16), we obtain

$$\begin{aligned}
&\mathbb{E}[h(\theta_n)] \\
&\leq \mathbb{E}[h(\theta_{v_0})] + \sum_{i=0}^{p}\left(\frac{10DC_4\sqrt{\log((1-\kappa)2^i n^{\alpha_2})}2^{-i/2}}{n^{\alpha_2/2}} + \frac{5a_0 2^{-i}}{n^{\alpha_1}}\left[C_2 + \frac{2BC_4\sqrt{\log((1-\kappa)2^i n^{\alpha_2})}}{2^{i/2}n^{\alpha_2/2}} + B^2\right]\right) \\
&\leq \mathbb{E}[h(\theta_{v_0})] + \sum_{i=0}^{\infty}\left(\frac{10DC_4\sqrt{\log((1-\kappa)2^i n^{\alpha_2})}2^{-i/2}}{n^{\alpha_2/2}} + \frac{5a_0 2^{-i}}{n^{\alpha_1}}\left[C_2 + \frac{2BC_4\sqrt{\log((1-\kappa)2^i n^{\alpha_2})}}{2^{i/2}n^{\alpha_2/2}} + B^2\right]\right) \\
&\leq \inf_{\lceil \frac{n}{4}\rceil \leq k \leq n_1} \mathbb{E}[h(\theta_k)] + \frac{10DC_4(3.5\sqrt{\log((1-\kappa)n^{\alpha_2})} + 3.5)}{n^{\alpha_2/2}} + \frac{10(C_2 + B^2)a_0}{n^{\alpha_1}} + \\
&\qquad \frac{10Ba_0C_4(1.6\sqrt{\log((1-\kappa)n^{\alpha_2})} + 0.6)}{n^{\alpha_1+\alpha_2/2}}. \quad (5.37)
\end{aligned}$$

The inequality in (5.25) becomes

$$\inf_{\lceil \frac{n}{4}\rceil \leq k \leq n_1} \mathbb{E}[h(\theta_k) - h(\theta^*)] \leq \frac{4D^2}{a_0 n^{1-\alpha_1}} + \frac{4DC_4\sqrt{\log((1-\kappa)n^{\alpha_2})}}{n^{\alpha_2/2}} + \frac{a_0B^2}{n^{\alpha_1}}. \quad (5.38)$$

Combining the inequalities in (5.37) and (5.38), we obtain

$$\begin{aligned}
&\mathbb{E}[h(\theta_n) - h(\theta^*)] \\
&\leq \frac{4D^2}{a_0 n^{1-\alpha_1}} + \frac{DC_4(39\sqrt{\log((1-\kappa)n^{\alpha_2})} + 35)}{n^{\alpha_2/2}} + \frac{(10C_2 + 11B^2)a_0}{n^{\alpha_1}} + \\
&\qquad\qquad \frac{BC_4(16\sqrt{\log((1-\kappa)n^{\alpha_2})} + 6)a_0}{n^{\alpha_1+\alpha_2/2}}. \qquad (5.39)
\end{aligned}$$

The values for α_1 and α_2 which will result in the tightest bound are 1/2 and 1 respectively. Substituting these values, we get the main claim of Theorem 5. □

CHAPTER 6

Conclusions and Future Work

We considered the problem of optimization of UBSR under two different settings. In the first setting i.e., under bandit feedback, we derived UCB and LCB required for the StoROO algorithm followed by the upper bound on simple regret. Next, we proposed a stochastic gradient descent scheme for UBSR optimization. Here, we made use of the UBSR derivative estimate whose bias depends on the batch size. We proposed a decreasing step size sequence and a batch size sequence for obtaining non asymptotic bound of the order $O(1/\sqrt{n})$. We also derived non-asymptotic bounds when the samples are not i.i.d., instead they follow a Markov chain. The resulting bound is of the order of $O(\sqrt{\log n/n})$. The technique used here for the optimization of UBSR is also applicable in optimizing general convex functions whose bias in the gradient estimate is a function of the batch size used to estimate the gradient.

In the future, one could explore the optimization of UBSR in a risk sensitive reinforcement learning setting. UBSR optimization with linear bandit feedback could be an interesting problem to solve as well. Optimization of UBSR with a constraint on the expected return is another possible research topic with many potential applications. Optimization of other risk measures such as CPT and exponential cost function in various settings provides many future research possibilities.

REFERENCES

[1] **Artzner, P.**, **F. Delbaen**, **J.-M. Eber**, and **D. Heath** (1999). Coherent measures of risk. *Mathematical finance*, **9**(3), 203–228.

[2] **Audibert, J.-Y.**, **R. Munos**, and **C. Szepesvári** (2009). Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, **410**(19), 1876–1902.

[3] **Balasubramanian, K.** and **S. Ghadimi** (2018). Zeroth-order (non)-convex stochastic optimization via conditional gradient and gradient updates. **31**. URL `https://proceedings.neurips.cc/paper/2018/file/36d7534290610d9b7e9abed244dd2f28-Paper.pdf`.

[4] **Bhavsar, N.** and **L. A. Prashanth** (2021). Non-asymptotic bounds for stochastic optimization with biased noisy gradient oracles.

[5] **Bodie, Z.** (1991). Shortfall risk and pension fund asset management. *Financial Analysts Journal*, **47**(3), 57–61. ISSN 0015198X. URL `http://www.jstor.org/stable/4479434`.

[6] **Borkar, V.**, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

[7] **Bottou, L.**, **F. E. Curtis**, and **J. Nocedal** (2018). Optimization methods for large-scale machine learning. *Siam Review*, **60**(2), 223–311.

[8] **Bubeck, S.** (2014). Convex optimization: Algorithms and complexity. URL `https://arxiv.org/abs/1405.4980`.

[9] **Duchi, J. C.**, **M. I. Jordan**, **M. J. Wainwright**, and **A. Wibisono** (2012). Finite sample convergence rates of zero-order stochastic optimization methods.

[10] **Dunkel, J.** and **S. Weber** (2010). Stochastic root finding and efficient estimation of convex risk measures. *Operations Research*, **58**(5), 1505–1521.

[11] **Föllmer, H.** and **A. Schied** (2002). Convex measures of risk and trading constraints. *Finance and stochastics*, **6**(4), 429–447.

[12] **Galichet, N.**, **M. Sebag**, and **O. Teytaud** (2014). Exploration vs exploitation vs safety: Risk-averse multi-armed bandits. *Journal of Machine Learning Research*, **29**.

[13] **Giesecke, K.**, **T. Schmidt**, and **S. Weber** (2008). Measuring the risk of large losses. *Journal of Investment Management*, **6**(4), 1–15.

[14] **Gopalan, A.**, **L. A. Prashanth**, **M. Fu**, and **S. Marcus** (2016). Weighted bandits or: How bandits learn distorted values that are not expected.

[15] **Hepworth, A. J.** (2017). A multi-armed bandit approach to superquantile selection. Technical report, Naval Postgraduate School Monterey United States.

[16] **Hong, L. J.**, **Z. Hu**, and **G. Liu** (2014). Monte carlo methods for value-at-risk and conditional value-at-risk: A review. *ACM Trans. Model. Comput. Simul.*, **24**(4). ISSN 1049-3301. URL `https://doi.org/10.1145/2661631`.

[17] **Hu, Z.** and **Z. Dali** (2016). Convex risk measures: efficient computations via monte carlo. *Available at SSRN 2758713.*

[18] **Jain, P.**, **D. Nagaraj**, and **P. Netrapalli** (2021). Making the last iterate of sgd information theoretically optimal. *SIAM Journal on Optimization*, **31**, 1108–1130.

[19] **Kloeckner, B.** (2018). Empirical measures: regularity is a counter-curse to dimensionality. URL `https://arxiv.org/abs/1802.04038`.

[20] **Kolla, R. K.**, **K. Jagannathan**, *et al.* (2019). Risk-aware multi-armed bandits using conditional value-at-risk. *arXiv preprint arXiv:1901.00997.*

[21] **Menon, A. S.**, **L. A. Prashanth**, and **K. Jagannathan** (2021). Online estimation and optimization of utility-based shortfall risk.

[22] **Moulines, E.** and **F. Bach** (2011). Non-asymptotic analysis of stochastic approximation algorithms for machine learning. **24**. URL `https://proceedings.neurips.cc/paper/2011/file/40008b9a5380fcacce3976bf7c08af5b-Paper.pdf`.

[23] **Munos, R.** (2014). From bandits to Monte-Carlo Tree Search: The optimistic principle applied to optimization and planning.

[24] **Pasupathy, R.**, **P. Glynn**, **S. Ghosh**, and **F. S. Hashemi** (2018). On sampling rates in simulation-based recursions. *SIAM Journal on Optimization*, **28**(1), 45–73.

[25] **Prashanth, L. A.** and **S. P. Bhat** (2020). Concentration of risk measures: A Wasserstein distance approach.

[26] **Prashanth L. A., K. J.** and **R. K. Kolla** (2019). Concentration bounds for cvar estimation: The cases of light-tailed and heavy-tailed distributions.

[27] **Riekert, A.** (2021). Wasserstein convergence rate for empirical measures of markov chains. URL `https://arxiv.org/abs/2101.06936`.

[28] **Rockafellar, R. T.**, **S. Uryasev**, *et al.* (2000). Optimization of conditional value-at-risk. *Journal of risk*, **2**, 21–42.

[29] **Sani, A.**, **A. Lazaric**, and **R. Munos** (2013). Risk-aversion in multi-armed bandits. *arXiv preprint arXiv:1301.1936*.

[30] **Torossian, L.**, **A. Garivier**, and **V. Picheny** (2020). X-armed bandits: Optimizing quantiles, cvar and other risks.

[31] **Tversky, A.** and **D. Kahneman** (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, **5**(4), 297–323.