Estimation of Spectral Risk Measures

A THESIS

submitted by

AJAY KUMAR PANDEY

for the award of the degree

of

MASTER OF SCIENCE

(by Research)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS

JUNE 2020

THESIS CERTIFICATE

This is to certify that the thesis titled **Estimation of Spectral Risk Measures**, submitted by **Ajay Kumar Pandey**, to the Indian Institute of Technology, Madras, for the award of the degree of **Master of Science (by Research)**, is a bonafide record of the research work done by him under our 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. Prashanth L.A.
Research Advisor
Assistant Professor
Department of Computer Science and Engineering
Indian Institute of Technology Madras, 600036

Place: Chennai

Date:

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere thankfulness to my research advisor, Dr. Prashanth L.A., who has the substance of an expert: he convincingly guided and encouraged me to be professional and do the right thing even when the road got tough. The door to my advisor's office was always open whenever I ran into a trouble spot or had a question about my research or writing. Without his steady assistance, the objective of this thesis would not have been completed.

Second, I would like to thank Dr. Sanjay P. Bhat, Tata Consultancy Services Limited, Hyderabad, whose valuable feedback and ideas improved the quality of this thesis considerably. I would also like to thank my General Test Committee members, Dr. Sutanu Chakraborti, Dr. Puduru Viswanadha Reddy, and Dr. Madhu Mutyam, for their immense guidance and valuable feedback.

Third, I am also grateful to the Computer Science and Engineering department in particular and the Indian Institute of Technology Madras in general for providing an excellent environment for doing research. Furthermore, I want to thank my friends: Nirav, Sidharth, Rajendra, Pawandeep, and many others, who made my stay pleasant as ever on the campus for the last three years.

Finally, I must express my very profound gratitude to my family: my parents, my brother, and my sister, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of research and writing this thesis. This accomplishment would not have been possible without them. Thank you.

ABSTRACT

KEYWORDS: Spectral risk measures, Value-at-Risk, Conditional Value-at-Risk, Estimation technique, Concentration bounds, Bounded distributions, Gaussian distribution, Exponential distribution.

Traditional approach to sequential decision-making under uncertainity is an optimization problem to minimize the expected value of the accumulated loss L: $\min_{\mathbf{x}} \mathbb{E}_y[L(\mathbf{x}, y)]$, where x is a vector of decision variables, and y is a random variable drawn from a loss distribution. However, decision-makers are often risk-averse as they would rather minimize the chance of having a very low reward than focus purely on the average. This is a rational behavior when failure can have large consequences. For instance, if a corporation suffers a disastrous loss, they may go out of business. Or in many cases, low performance entails safety issues. Hence, it is natural to move beyond average-case analysis and optimize a risk-aware objective function.

Various risk measures have been proposed in the literature, e.g., mean-variance tradeoff (Markowitz, 1952), value-at-risk (VaR) and conditional value-at-risk (CVaR) (Rockafellar *et al.*, 2000; Nski, 2010; Shen *et al.*, 2013), spectral risk measures (SRM) (Acerbi, 2002), prospect theory (Tversky and Kahneman, 1979) and its later enhancement, cumulative prospect theory (CPT) (Tversky and Kahneman, 1992).

In this thesis, we consider the problem of estimating SRM from independent and identically distributed (i.i.d.) samples, and propose a novel method that is based on numerical integration. We show that our SRM estimate concentrates exponentially, when the underlying distribution has bounded support. Further, we also consider the case when the underlying distribution is either Gaussian or exponential, and derive a concentration bound for our estimation scheme. Further, we specialize our results to handle CVaR, which is a popular risk measure in finance.

We solve a SRM-sensitive multi-armed bandit (MAB) problem using the best arm identification (BAI) paradigm. BAI is suitable because of simulation optimization, and

also the fact that SRM relates to rare events, making samples hard to obtain in real-world settings. Further, we practically validate our algorithm using SUMO, a state-of-the-art traffic simulator in a vehicular traffic routing application. Also, we consider a portfolio optimization application with CVaR-based criteria, and perform simulation experiments that show the efficacy of the CVaR estimator.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

OR	Operation Research
AI	Artificial Intelligence
VaR	Value-at-Risk
CVaR	Conditional Value-at-Risk
CPT	Cummulative Prospect Theory
SRM	Spectral Risk Measures
EDF	Empirical Distribution Function
SR	Successive Rejects
SUMO	Simulation of Urban Mobility
MVRM	Mean-Variance Risk Measure
PDF	Probability density function
CDF	Cumulative distribution function

CHAPTER 1

INTRODUCTION

In the context of risk-sensitive optimization, Conditional Value-at-Risk (CVaR) is a popular risk measure. CVaR is a conditional expectation of a random variable (r.v.) that usually models the losses in an application (e.g., finance), where the conditioning is based on Value-at-Risk (VaR). The latter denotes the maximum loss that could be incurred, with high probability. CVaR is an appealing risk measure because it is coherent (Artzner *et al.*, 1999), while VaR is not, as it violates the sub-additivity assumption required for coherency.

CVaR is a special instance of the class of spectral risk measures (SRM). The advantage of employing SRM, instead of CVaR in a risk-sensitive optimization setting is that, in addition to being coherent, SRM allows for better risk-aversion. This is because in CVaR all the tail-losses recieve the same weight, whereas SRM has a risk-aversion function, which can ensure that higher losses receive a higher weight, or at least, the same weight as lower losses.

In practical applications, the behaviour of the underlying distribution is unknown. However, one can obtain samples from the distribution, either directly in a real-world application, or through a simulator. We focus on SRM estimation from i.i.d. samples. First, we propose an estimator for SRM using the trapezoidal rule. Next, we derive concentration bounds for this estimator, when the underlying distribution has either bounded support, or is unbounded, but is either Gaussian or exponential.

As a practical application, we consider the vehicular traffic routing application. Here, there are a fixed number of routes, and each route has a delay distribution. The objective is to find a route that has the lowest SRM delay. To perform this experiment, we use SUMO, a vehicular traffic simulator, because it is hard to obtain samples in real-world settings. This vehicular traffic routing application falls under the realm of multi-armed bandits (MAB) (Slivkins *et al.*, 2019), which is the setting for sequential decision-making under uncertainty.

The rest of this chapter is organised as follows: In Section 1.1, we provide background material on risk measures. In Section 1.2, we describe SRM. In Section 1.3, we outline the contributions of this thesis. In Section 1.4, we survey related work, and finally, in Section 1.5, we provide an outline of the remaining chapters.

1.1 Background

This section provides background on risk measures and different types of frameworks to measure the risk. First, we discuss the mean-variance framework along with its functionality and limitations. Subsequently, we describe risk measures VaR and CVaR.

We start with an example to motivate the need for incorporating a risk measure in decision making. Suppose a person has a meeting at his office in 90 minutes. He/She has two routes to choose from. The first route, say A, has a mean delay of 50 minutes and a worst-case delay of 5 hours, whereas the second say B, route has a mean delay of 60 minutes and a worst-case delay of 80 minutes. If the person chooses a route based on minimum expected delay, i.e., route A, then there is a chance that he/she is very late for the meeting, owing to the large worst-case delay of route A. To address the needs of a risk-averse decision maker, we need a risk measure that considers worst-case delay, and maps the loss distribution to a real number.

Next, we will discuss some of the risk measures present in the literature with their properties and limitations.

1.1.1 Mean-Variance risk measure (MVRM)

Mean-variance risk measure (Markowitz, 1952) is a traditional approach of measuring risk, where risk is formulated in terms of mean and variance of the loss distribution. We assume that the underlying loss distribution follows a normal distribution. A r.v. X is normally distributed with mean μ and variance σ^2, if its probability density function (PDF) has the form:

$$\mathrm{f(x)} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\mathrm{x}-\mu)^2}{2\sigma^2}\right]$$

where $\mathrm{x} \in (-\infty, +\infty)$.

A PDF gives the idea about possible outcomes and how likely these outcomes are. The normal distribution PDF is the bell-like curve, which implies that outcomes are likely to occur close to the mean μ, and the spread around the mean depends on the standard deviation σ. In the mean-variance framework, the standard deviation is used as a measure of risk.

To explain how the mean-variance method works, suppose we wish to choose a particular route from a set of routes. We are only concerned about the expected delay on the route and the variance of its delays. Using the interpretation from [(Dowd, 2005), Section 2.1], the various possibilities of routes with their expected delay and variance of delay are shown by the efficient frontier curve in Figure 1.1. Since the user regards a lower expected delay as 'best' and higher variance of delays (i.e higher risk) as 'worst', the user wants to achieve the lowest possible expected delay for any given level of risk; or equivalently, wants to achieve the lowest possible level of risk for any given expected delay. A user who is more risk-averse will choose a point on the efficient curve with low risk or lower variance route, while a less risk-averse user will choose a point with higher risk, which might end up choosing a route with lower delay.

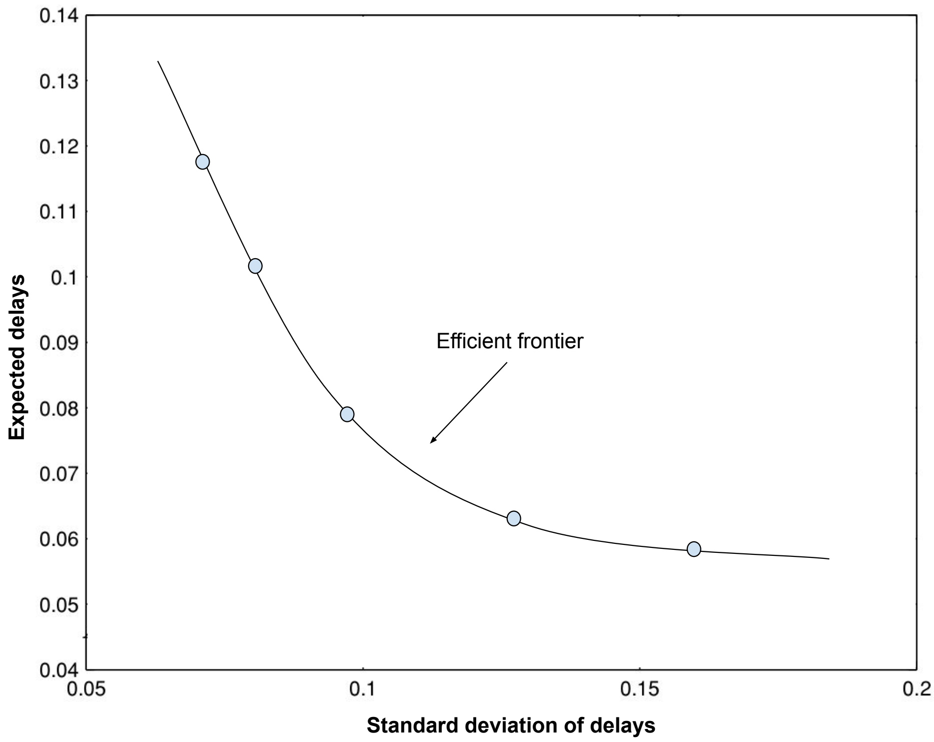

Figure 1.1: The mean-variance efficient frontier curve.

However, the variance is not an adequate risk measure when underlying loss distribution is not Gaussian cf. [(Dowd, 2005), Section 2.1]. For example, in Figure 1.2, we have two distributions with the same mean and variance. The mean-variance method results in the same risk for both the distributions. Nevertheless, the heavy-tailed distribution has a longer tail, having more likely outcomes in the extreme tail region.

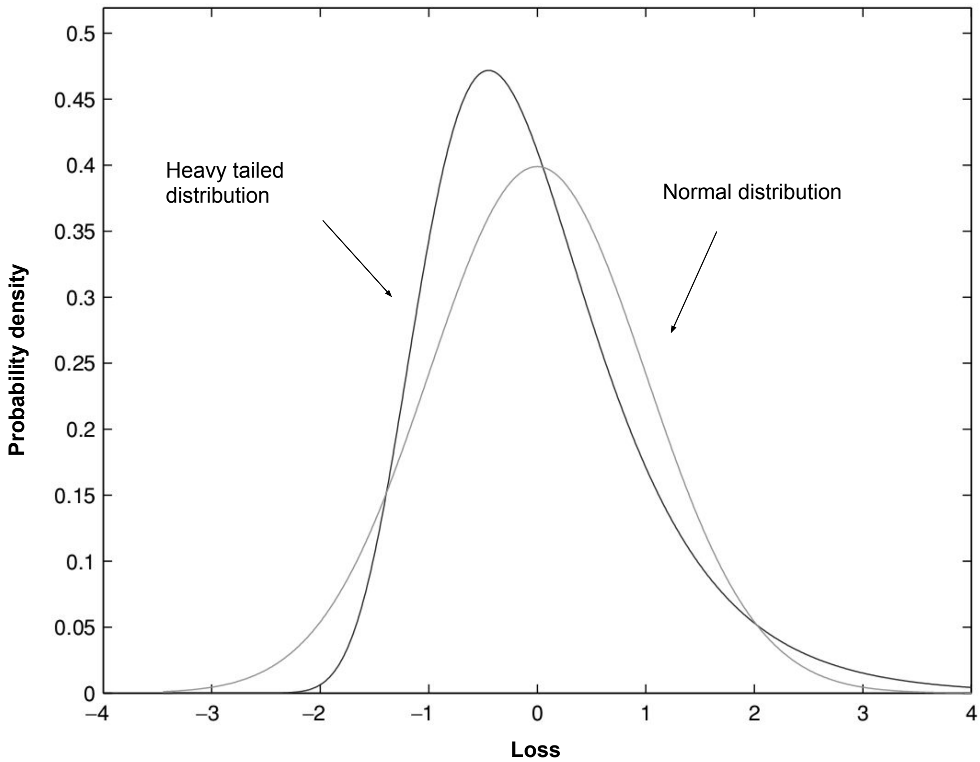

Figure 1.2: Normal distribution and heavy tailed distribution

1.1.2 Value-at-Risk (VaR)

For a given confidence level $\beta \in (0, 1)$ VaR at level β denotes the maximum loss that can occur with $(\beta \times 100)\%$ confidence, and it is defined below:

Definition 1.1.1 (Value-at-Risk)**.** *For a r.v.* X*,* VaR $\mathrm{V}_\beta(X)$ *at level* β*,* $\beta \in (0, 1)$*, is defined as follows:*

$$\mathrm{V}_\beta(X) := \inf\{c : \mathbb{P}(X \leq c) \geq \beta\}.$$

Note that, if X has a continuous and strictly increasing cumulative distribution function

(CDF) F, then $\mathrm{V}_\beta(X)$ is a solution to the following:

$$\mathbb{P}[X \leq \xi] = \beta, \text{ i.e., } \mathrm{V}_\beta(X) = F^{-1}(\beta).$$

Estimation of VaR

Let X_i, $i = 1, \dots, n$ denote i.i.d. samples from the distribution of X. Then, the estimate of $\mathrm{V}_\beta(X)$, denoted by $\widehat{\mathrm{V}}_{n,\beta}$, is formed as follows (Serfling, 2009):

$$\widehat{\mathrm{V}}_{n,\beta} = \widehat{F}_n^{-1}(\beta) = \inf\{x : \widehat{F}_n(x) \geq \beta\}, \tag{1.1}$$

where $\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}[X_i \leq x]$ is the EDF of X. Letting $X_{(1)}, \dots, X_{(n)}$ denote the order statistics, i.e., $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, we have $\widehat{\mathrm{V}}_{n,\beta} = X_{(\lceil n\beta \rceil)}$.

Derivative of VaR

We recall a result from (Dufour, 1995) below.

Lemma 1.1.2. *Let F and f are respectively CDF and PDF of continuous r.v. X. Suppose, the density f is positive in a neighborhood of $\mathrm{V}_\beta(\mathrm{X})$ (denoted V_β for notational convenience), where $0 < \beta < 1$, then we have*

$$\mathrm{V}_\beta^{'} = \frac{1}{f\left(\mathrm{V}_\beta\right)}, \quad \mathrm{V}_\beta^{''} = -\frac{f^{'}\left(\mathrm{V}_\beta\right)}{f\left(\mathrm{V}_\beta\right)^3}$$

Proof. Notice that $F(F^{-1}(\beta)) = \beta$, which implies

$$F'(F^{-1}(\beta))F^{-1'}(\beta) = 1, \text{ and} \tag{1.2}$$

$$F^{''}(F^{-1}(\beta))(F^{-1'}(\beta))^2 + F^{'}(F^{-1}(\beta))F^{-1''}(\beta) = 0 \tag{1.3}$$

From (1.2) and (1.3), we have

$$\begin{aligned}
\mathrm{V}_\beta^{'} = F^{-1'}(\beta) &= \frac{1}{F'(F^{-1}(\beta))} = \frac{1}{f(F^{-1}(\beta))} = \frac{1}{f\left(\mathrm{V}_\beta\right)}, \text{ and} \\
\mathrm{V}_\beta^{''} = F^{-1''}(\beta) &= -\frac{F^{''}(F^{-1}(\beta))(F^{-1'}(\beta))^2}{F^{'}(F^{-1}(\beta))} = -\frac{f^{'}(F^{-1}(\beta))(F^{-1'}(\beta))^2}{f(F^{-1}(\beta))} \\
&= -f^{'}(F^{-1}(\beta))(F^{-1'}(\beta))^3 = -\frac{f^{'}\left(F^{-1}\left(\beta\right)\right)}{f\left(F^{-1}\left(\beta\right)\right)^3} = -\frac{f^{'}\left(\mathrm{V}_\beta\right)}{f\left(\mathrm{V}_\beta\right)^3}
\end{aligned}$$

□

Limitations of VaR

VaR at level β does not provide information about the tail loss that occur with probability (1-β). Moreover, if the tail event occurs, the loss incurred is more than the VaR, and this can lead to undesirable outcomes. For example, some route has a low expected delay but also involves the possibility of higher delays, and a VaR risk measure based decision might lead to adopting this route, regardless of the size of higher delay outcomes.

1.1.3 Coherent risk measure

In (Artzner *et al.*, 1999) the authors postulated a set of four axioms that need to be true in order to qualify a risk measure as coherent. For any two loss r.v.s X and Y, a risk measure $\mathbb{R}(.)$ is said to be coherent if it satisfies the following conditions:

1. *Monotonicity*: $X \leq Y \implies \mathbb{R}(X) \leq \mathbb{R}(Y)$ for all X, Y.
2. *Sub-additivity*: $\mathbb{R}(X+Y) \leq \mathbb{R}(X) + \mathbb{R}(Y)$ for all X, Y.
3. *Positive homogeneity*: $\mathbb{R}(\lambda X) = \lambda \mathbb{R}(X)$ for all $X, \lambda \geq 0$.
4. *Translational invariance*: $\mathbb{R}(X+c) = \mathbb{R}(X) + c$ for all X, c.

VaR is not a coherent risk measure as it violates the sub-additivity condition. CVaR, unlike VaR, is a coherent risk measure, which can be used to model worst-case losses. We introduce this risk measure next.

1.1.4 Conditional Value-at-Risk (CVaR)

Definition 1.1.3. *For a r.v.* X*,* CVaR $\mathrm{C}_\beta(X)$ *at the level* β*,* $\beta \in (0,1)$*, is defined as follows:*

$$\mathrm{C}_\beta(X) := \mathrm{V}_\beta(X) + \frac{1}{1-\beta}\mathbb{E}[X - \mathrm{V}_\beta(X)]^+, \tag{1.4}$$

where $[x]^+ = \max(0,x)$ *for a real number* x*, and* $\mathrm{V}_\beta(X)$ *is the* VaR *at level* β *of a r.v.* X.

Let X be a continuous r.v. representing loss. Then, CVaR $\mathrm{C}_\beta(X)$ can be interpreted as the expected loss, conditional on the event that the loss exceeds $\mathrm{V}_\beta(X)$, i.e., $\mathrm{C}_\beta(X) = \mathbb{E}[X|X \geq \mathrm{V}_\beta(X)]$. As illustrated in Figure 1.3, unlike VaR, CVaR gives an idea about how adverse can be outcomes on an average after VaR. CVaR is a sub-additive risk measure, and it also satisfies other conditions of coherence given in Section 1.1.3.

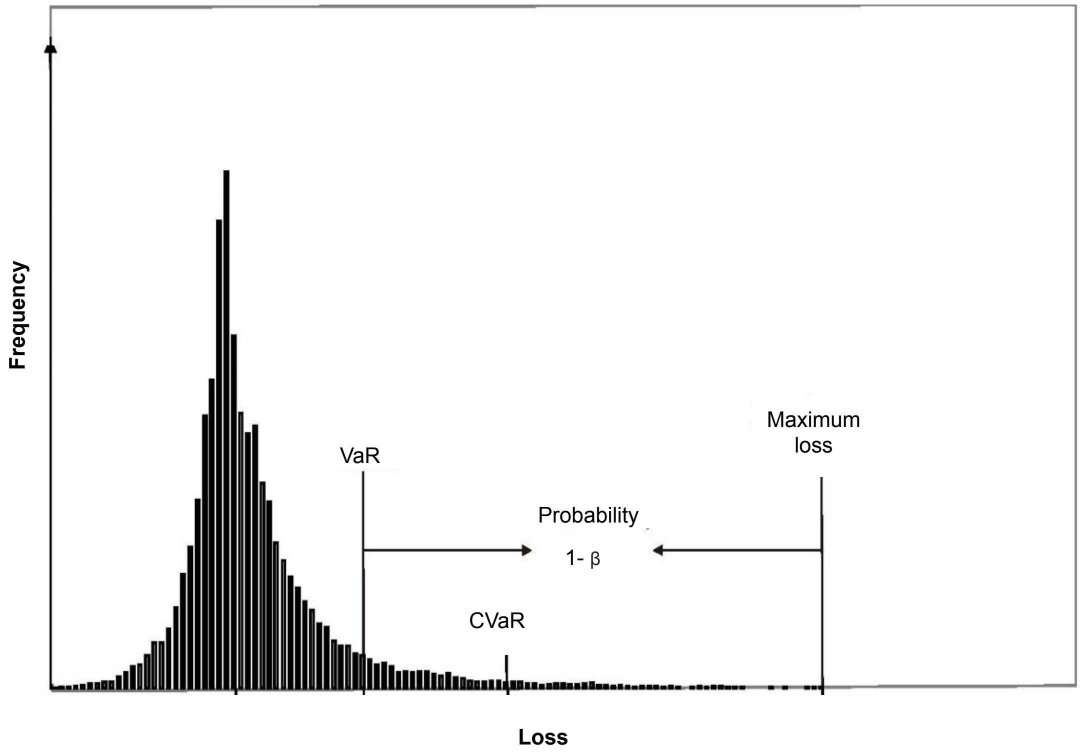

Figure 1.3: VaR and CVaR at level β of a r.v. X representing loss.

Acerbi's formula (Acerbi and Tasche, 2002), an alternative form for $\mathrm{C}_\alpha(X)$, is as follows:

$$\mathrm{C}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{V}_\beta(X)\, d\beta. \tag{1.5}$$

From the expression above, $\mathrm{C}_\alpha(X)$ can be interpreted as the average of $\mathrm{V}_\beta(X)$ for $\beta \in [\alpha, 1)$.

Estimation of CVaR

Let X_i, $i = 1, \ldots, n$ denote i.i.d. samples from the distribution of X. Then, the estimates of $\mathrm{C}_\alpha(X)$, denoted by $\widehat{\mathrm{CVaR}}_{n,\alpha}$, is formed as follows (Serfling, 2009):

$$\widehat{\mathrm{CVaR}}_{n,\alpha} = \widehat{\mathrm{V}}_{n,\alpha} + \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} \left[X_i - \widehat{\mathrm{V}}_{n,\alpha}\right]^{+}, \tag{1.6}$$

where $\widehat{\mathrm{V}}_{n,\alpha}$ is the VaR estimator given in (1.1).

1.2 Spectral risk measure (SRM)

A spectral risk measure $\mathrm{S(X)}$ of a r.v X is defined as follows:

$$\mathrm{S}(X) = \int_0^1 \varphi(\beta) \mathrm{V}_\beta(X)\, d\beta, \tag{1.7}$$

where $\varphi(\cdot)$ is a risk-aversion function, and $\mathrm{V}_\beta(\mathrm{X})$ is the VaR at level β of the r.v. X.

SRM can be seen as a weighted average of the VaR of the underlying distribution. Moreover, CVaR can be recovered by setting:

$$\varphi(\beta) = \begin{cases} 0 & \beta < \alpha \\ 1/(1-\alpha) & \beta \geq \alpha \end{cases} \tag{1.8}$$

The latter choice translates to an equal weight for all tail-loss VaR values.

The risk aversion function $\varphi(\cdot)$ can be chosen to ensure that SRM is a coherent risk measure (Acerbi, 2002). In particular, the following properties ensure coherence.

- *Postivity:* $\varphi(\beta) \geq 0$ for all $\beta \in (0, 1)$;
- *Increasingness:* $\varphi'(\beta) \geq 0$ for all $\beta \in (0, 1)$; and
- *Normalization:* $\int_0^1 \varphi(\beta) d\beta = 1$.

SRM can model a user's risk-aversion better, since the function $\varphi(.)$ can be chosen such that higher losses receive a higher weight, or at least, the same weight as lower losses (Dowd and Blake, 2006). In contrast, CVaR assigns the same weight to all tail losses as evident in (1.8). An example of a risk-aversion function is the exponential utility function, defined by

$$\varphi(\beta) = \frac{ke^{-k(1-\beta)}}{1-e^{-k}}, \tag{1.9}$$

where $\beta \in (0,1)$, and $k \in (0,\infty)$ reflects the user's degree of risk-aversion.

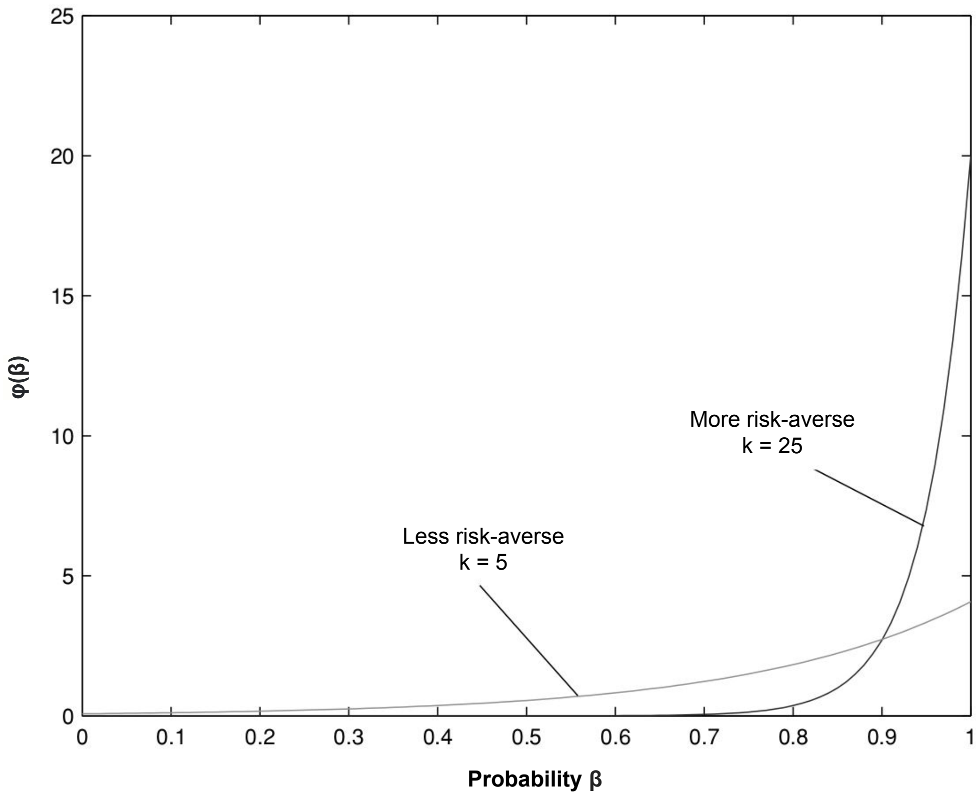

Figure 1.4: An example of risk-aversion function using (1.9).

In Figure 1.4, an exponential utility function is illustrated. As illustrated in Figure 1.4, larger values of k ensure higher risk-aversion.

1.3 Contributions of the thesis

In this work, we consider the problem of estimating SRM of a r.v., given i.i.d samples from the underlying distribution. In this context, our contributions are as follows:

First, we provide a natural estimation scheme for SRM that uses the empirical distribution function (EDF) to estimate VaR, together with a trapezoidal rule-based approximation.

Second, we provide a two-sided concentration bound for our proposed SRM estimate, for the case when the underlying distribution either has a bounded support, or is unbounded, but either Gaussian or exponential. Our tail bounds are of the order $O\left(c_1 \exp\left(-c_2 n \epsilon^2\right)\right)$, where n is the number of samples, ϵ is the accuracy parameter, and c_1, c_2 are universal constants.

Third, we consider the CVaR risk measure, a special case of SRM. For CVaR, we specialize the SRM estimator, and provide concentration bounds for distributions that have bounded support, or an unbounded, but Gaussian or exponential.

Fourth, we perform simulation experiments to show the efficacy of our proposed SRM estimation scheme. In particular, we consider a synthetic setup and show that our scheme provides accurate estimates of SRM. Next, we incorporate our SRM estimation scheme in the inner loop of the successive rejects (SR) algorithm (Audibert *et al.*, 2010), which is a popular algorithm in the best arm identification framework for multi-armed bandits. We test the resulting SR algorithm variant in a vehicular traffic routing application using the Simulation of Urban Mobility (SUMO) traffic simulator (Behrisch *et al.*, 2011). The application is motivated by the fact that, in practice, human road users may not always prefer the route with the lowest mean delay. Instead, a route that minimized worst-case delay, while doing reasonably well on the average, is preferable, and such a preference can be encoded into the risk aversion function $\varphi(\cdot)$ in (1.7). Further, we consider a portfolio optimization application with CVaR-based criteria, and perform simulation experiments that show the efficacy of the CVaR estimator.

1.4 Related work

Concentration bounds for empirical SRM have been derived recently in (Bhat and Prashanth, 2019). In comparison to (Bhat and Prashanth, 2019), our bounds, using a different estimator, exhibits a similar rate of exponential convergence around true SRM, for distributions with bounded support and special case of Gaussian distribution. And, our bound exhibits exponential concentration for special case of exponential distribution, while the corresponding bound in (Bhat and Prashanth, 2019) shows a polynomial decay for accuracy parameter $\epsilon > 1$.

The bounds that we derive for SRM estimation could be specialized to the case of

CVaR. In (Brown, 2007; Wang and Gao, 2010) concentration bounds for the classic CVaR estimator are derived. Our bound matches the rate obtained in (Brown, 2007; Wang and Gao, 2010) for distributions with bounded support. For the case of distributions with unbounded support, concentration bounds for empirical CVaR have been derived recently in (Thomas and Learned-Miller, 2019; Kolla *et al.*, 2019; Prashanth *et al.*, 2019; Bhat and Prashanth, 2019). In (Thomas and Learned-Miller, 2019; Kolla *et al.*, 2019) (resp. (Prashanth *et al.*, 2019; Bhat and Prashanth, 2019)), the authors derive an one-sided concentration bound (resp. two-sided bounds), when the underlying distributions are either sub-Gaussian or sub-exponential (Wainwright, 2019). In comparison to (Thomas and Learned-Miller, 2019; Kolla *et al.*, 2019), we derive two-sided concentration bounds for the special case of Gaussian and exponential distributions. Also, the results in (Thomas and Learned-Miller, 2019) does not allow a bandit application. Moreover, in (Prashanth *et al.*, 2019) concentration bounds for the classic CVaR estimator are derived for heavy-tailed and light-tailed distributions. Our bound matches the rate obtained in (Prashanth *et al.*, 2019) for the special case of Gaussian and exponential distributions. Finally, in comparison to a recent result in (Bhat and Prashanth, 2019), for the special case of exponential distribution, our bound exhibits exponential concentration, while the corresponding bound in (Bhat and Prashanth, 2019) shows a polynomial decay for accuracy parameter $\epsilon > 1$.

1.5 Outline of chapters

The rest of the thesis is organized as follows:

Chapter 2 first presents a novel method for SRM estimation. Second, it presents concentration bounds for SRM estimation with their convergence proofs, for the case when the underlying distribution has bounded support, or unbounded, but is either Gaussian or exponential. And third, it provides concentration bounds for the case of CVaR estimation.

Chapter 3 first presents the applications of SRM and CVaR using simulation experiments. Second, it presents a vehicular traffic routing application using SUMO vehicular traffic simulator.

Finally, Chapter 4 concludes the thesis, and discusses a few interesting directions for future research.

CHAPTER 2

SRM ESTIMATION AND CONCENTRATION BOUNDS

This chapter first presents a novel method for SRM estimation. Second, it presents concentration bounds for SRM estimation with their convergence proofs, for the case when the underlying distribution has bounded support, or unbounded, but is either Gaussian or exponential. And third, it provides the CVaR specific results.

2.1 SRM Estimation scheme: Bounded case

We estimate $\mathrm{S}(X)$, given i.i.d. samples $X_1, \ldots, X_n$ from the distribution of X, by approximating the integral in SRM definition (1.7). Notice that the integrand $V_\beta(X)$ in (1.7) has to be estimated using the samples. Recall that $\widehat{\mathrm{V}}_{n,\beta}$ is the estimate of $\mathrm{V}_\beta(X)$, given by (1.1). We use the weighted VaR estimate to form a discrete sum to approximate the integral, an idea motivated by the trapezoidal rule (Cruz-Uribe and Neugebauer, 2003). The estimate $\widehat{\mathrm{S}}_{n,m}$ of $\mathrm{S}(X)$ is formed as follows:

$$\widehat{\mathrm{S}}_{n,m} = \sum_{k=1}^{m} \frac{\varphi(\beta_{k-1})\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \varphi(\beta_k)\widehat{\mathrm{V}}_{n,\beta_k}}{2} \Delta\beta. \tag{2.1}$$

In the equation above, $\{\beta_k\}_{k=0}^{m}$ is a partition of $[0,1]$ such that $\beta_0 = 0$ and $\beta_k = \beta_{k-1} + \Delta\beta$, where $\Delta\beta = 1/m$ is the length of each sub-interval.

2.2 SRM Estimation scheme: Unbounded case

For the case of unbounded distributions, we use a truncation-based estimator, which is described below.

Let $X_1, \ldots, X_n$ denote i.i.d. samples from the distribution of X. We form a truncated set of samples as follows:

$$\bar{X}_i = X_i \mathbb{I}\left\{X_i \leq B_n\right\},$$

where B_n is a truncation threshold that depends on the underlying distribution. For the case of Gaussian distribution with mean zero and variance σ^2, we set $B_n = \sqrt{2\sigma^2 \log{(n)}}$, and for the case of exponential distribution with mean $1/\lambda$, we set $B_n = \frac{\log(n)}{\lambda}$.

We form an SRM estimate along the lines of (2.1), except that the samples used are truncated samples, i.e.,

$$\widetilde{\mathrm{S}}_{n,m} = \sum_{k=1}^{m} \frac{\varphi(\beta_{k-1})\widetilde{\mathrm{V}}_{n,\beta_{k-1}} + \varphi(\beta_k)\widetilde{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta, \tag{2.2}$$

where, $\widetilde{\mathrm{V}}_{n,\beta} = \widetilde{F}_n^{-1}(\beta)$, with $\widetilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[\bar{X}_i \leq x]$.

2.3 Concentration bounds

This section presents concentration bounds for the case when the underlying distribution has bounded support, or unbounded, but is either Gaussian or exponential.

The motivation behind choosing Gaussian distribution is that many continuous data in nature and psychology display this bell-shaped curve when compiled and graphed. For example, if we randomly sampled 100 individuals, we would expect to see a Gaussian distribution frequency curve for many continuous variables, such as IQ, height, weight, and blood pressure. Similarly, the exponential distribution is also essential as it is often concerned with the amount of time until some specific event occurs. For example, the amount of time until an earthquake occurs and the amount of time, in months, a car battery lasts have an exponential distribution.

2.3.1 Distributions with bounded support

For notational convenience, we shall use V_β and S to denote $\mathrm{V}_\beta(X)$ and $\mathrm{S}(X)$, for any $\beta \in (0, 1)$.

For all the results presented below, we let $\widehat{\mathrm{S}}_{n,m}$ denote the SRM estimate formed from n i.i.d. samples of X and with m sub-intervals, using (2.1). Let F and f denote the distribution and density of X, respectively.

For the sake of analysis, we make one of the following assumptions:

(A1) Let $\varphi(\beta)$ be a risk-aversion function such that $|\varphi(\beta)| \leq C_1$ and $|\varphi'(\beta)| \leq C_2, \ \forall \beta \in [0,1]$.

(A1′) The conditions of (A1) hold. In addition, $|\varphi''(\beta)| \leq C_3, \ \forall \beta \in [0,1]$.

Theorem 2.3.1 (SRM concentration: bounded case). *Let the r.v. X be continuous and $X \leq B$ a.s. Fix $\epsilon > 0$.*

(i) *Assume (A1) holds and $f(x) \geq 1/\delta_1 > 0$, $x \leq B$. If $|\, B\,C_2 + \delta_1\,C_1 \,| \leq K_1$, and $m \geq \frac{K_1}{2\epsilon}$, then*

$$\mathbb{P}\left(\left|\mathrm{S} - \widehat{\mathrm{S}}_{n,m}\right| > \epsilon\right) \leq \frac{2K_1}{\epsilon} \exp\left(\frac{-n\,c\,\epsilon^2}{2C_1^2}\right), \tag{2.3}$$

where $c = \min\{c_0, c_1, \ldots, c_m\}$ and $c_k, k \in \{0, \ldots, m\}$, is a constant that depends on the value of the density f of the r.v. X in a neighborhood of V_{β_k}, with β_k as in (2.1).

(ii) *Assume (A1′) holds and $\frac{\left|f'(x)\right|}{f(x)^3} \leq \delta_2$, $x \leq B$. If $|\, B\,C_3 + 2\,\delta_1\,C_2 + \delta_2\,C_1 \,| \leq K_2$, and $m \geq \sqrt{\frac{K_2}{6\epsilon}}$, then*

$$\mathbb{P}\left(\left|\mathrm{S} - \widehat{\mathrm{S}}_{n,m}\right| > \epsilon\right) \leq \sqrt{\frac{8K_2}{3\epsilon}} \exp\left(\frac{-n\,c\,\epsilon^2}{2C_1^2}\right), \tag{2.4}$$

where c is as in the case above.

Proof. See Section 2.4.1. □

For small values of ϵ, the bound in (2.4) is better than that in (2.3). However, the bound in (2.3) is derived under weaker assumptions on the r.v. X and the risk-aversion function φ, as compared to the bound in (2.4).

In part (i) of the theorem above, we assumed that the density f of X is bounded below by $\frac{1}{\delta_1} > 0$. This implies that the derivative of VaR is bounded above. The latter condition is required for the trapezoidal rule to provide a good approx to the integral in (1.7). Moreover, the assumption that the first derivative of VaR w.r.t. the confidence level β is bounded implies that the underlying r.v. X is bounded. This claim can be made precise as follows: For any $\beta \in (0,1)$, it can be shown that (see Lemma 1.1.2 for a proof)

$$\mathrm{V}_{\beta}^{'} = \frac{1}{f\left(\mathrm{V}_{\beta}\right)}, \text{ and } \mathrm{V}_{\beta}^{''} = -\frac{f^{'}\left(\mathrm{V}_{\beta}\right)}{f\left(\mathrm{V}_{\beta}\right)^{3}}. \tag{2.5}$$

Notice that the first derivative of VaR involves a $1/f$ term, and if the r.v. is unbounded, then for every $\epsilon > 0$, there is an x such that $0 < f(x) < \epsilon$. This implies $1/f$ cannot be bounded above uniformly w.r.t x, and hence the derivative of VaR cannot be bounded either.

The stronger condition $\mid f^{'}(x) \mid / f(x)^3 \leq \delta_2$ used in part (ii) of Theorem 2.3.1, in conjunction with (2.5), implies that the second derivative of VaR is bounded. Now, as before, a bounded second derivative implies that the underlying r.v. X is bounded. To see this, the expression for the second derivative of VaR involves a $\frac{f'}{f^3}$ term, and if the r.v. X is unbounded, then a uniform bound on $\frac{f'}{f^3}$ would mean that, as $x \to \infty$, f decays too slowly to integrate to something finite, leading to a contradiction. More precisely, the differential inequality $\frac{|f'|}{f^3} < K$ can be "solved" to get $f(x) > \frac{C}{\sqrt{a+bx}}$ for large x and suitable constants a, b, and C. However, the expression on the RHS integrates to infinity, and hence, no density f with unbounded support can have $\frac{f'}{f^3}$ bounded.

2.3.2 Gaussian and exponential distributions

Here, we present concentration bounds for our SRM estimator assuming that the samples are either from a Gaussian distribution with mean zero and variance σ^2, or from the exponential distribution with mean $1/\lambda$. Note that the estimation scheme is not provided this information about the underlying distribution. Instead $\widetilde{\mathrm{S}}_{n,m}$ is formed from n i.i.d. samples and with m sub-intervals, using the truncation-based scheme in (2.2).

Theorem 2.3.2 (**SRM concentration: Gaussian case**). *Assume (A1). Suppose that the r.v. X is Gaussian with mean zero and variance σ^2, with $\sigma \leq \sigma_{max}$. If $m \geq \frac{1}{5}\sqrt{\frac{\sigma_{max}}{\epsilon}} \exp\left(\frac{n c \epsilon^2}{4C_1^2}\right)$ and $\epsilon > \frac{2\sigma_{max} C_1}{\sqrt{n}}$, then*

$$\mathbb{P}\left[\left|\mathrm{S} - \widetilde{\mathrm{S}}_{n,m}\right| > \epsilon\right] \leq \frac{2\sigma\left(\sqrt{2\log(n)}\, C_2 + \sqrt{2\pi} n\, C_1\right)}{\left(\epsilon - \frac{2\sigma C_1}{\sqrt{n}}\right)} \exp\left[-\frac{nc\left(\epsilon - \frac{2\sigma C_1}{\sqrt{n}}\right)^2}{2C_1^2}\right],$$

where c is as in Theorem 2.3.1 (i).

Proof. See Section 2.4.2. $\square$

Theorem 2.3.3 (SRM concentration: Exponential case). *Assume (A1). Suppose that the r.v. X is exponentially distribution with parameter λ, and $0 < \lambda_{min} \leq \lambda$. If $m \geq \frac{1}{8}\sqrt{\frac{1}{\lambda_{min}\epsilon}}\exp\left(\frac{nc\epsilon^2}{4C_1^2}\right)$ and $\epsilon > \frac{C_1(n+1)}{\lambda_{min}\, n}$, then*

$$\mathbb{P}\left[\left|\mathrm{S} - \widetilde{\mathrm{S}}_{n,m}\right| > \epsilon\right] \leq \frac{2\left(\frac{log(n)\, C_2}{\lambda} + n\, C_1\right)}{\left(\epsilon - \frac{C_1(n+1)}{\lambda n}\right)} \exp\left[-\frac{nc\left(\epsilon - \frac{C_1(n+1)}{\lambda n}\right)^2}{2C_1^2}\right],$$

where c is as in Theorem 2.3.1 (i).

Proof. See Section 2.4.3. $\square$

Remark 1. *Note that concentration bounds for CVaR estimation can be derived using a completely parallel argument to that of the proof of the theorems above, together with following choice for risk aversion function $\varphi(\beta) = 1/(1-\alpha)\mathbb{I}\{\beta > \alpha\}$, $\alpha \in (0,1)$. The CVaR-specific results are provided in Section 2.5.*

2.4 Convergence proofs

2.4.1 Proof of Theorem 2.3.1

For establishing the bound in Theorem 2.3.1, we require a result concerning the error of a trapezoidal-rule-based approximation, and a concentration bound for the VaR estimate in (1.1). We state these results below, and subsequently provide a proof of Theorem 2.3.1.

Lemma 2.4.1. *Let $0 < a \leq b < 1$, and $\{\beta_k\}_{k=0}^m$ be a partition of $[a,b]$ such that $\beta_0 = a$ and $\beta_k = \beta_{k-1} + \Delta\beta$, $\Delta\beta = \frac{(b-a)}{m}$ is length of each sub-interval.*

***(i)** If $\left|(\varphi(\beta)V_\beta)'\right| \leq K_1$ for $\beta \in [a,b]$, then*

$$\left|\int_a^b \varphi(\beta)\mathrm{V}_\beta\, d\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\mathrm{V}_{\beta_{k-1}} + \varphi(\beta_k)\mathrm{V}_{\beta_k}}{2}\Delta\beta\right| \leq \frac{K_1(b-a)^2}{4m}.$$

(2.6)

(ii) *If* $\left|(\varphi(\beta)V_\beta)''\right| \leq K_2$ *for* $\beta \in [a,b]$*, then*

$$\left|\int_a^b \varphi(\beta)\mathrm{V}_\beta \, d\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\mathrm{V}_{\beta_{k-1}} + \varphi(\beta_k)\mathrm{V}_{\beta_k}}{2}\Delta\beta\right| \leq \frac{K_2(b-a)^3}{12m^2}. \tag{2.7}$$

Proof. See Appendix A. □

Lemma 2.4.2 (VaR concentration). *Let the r.v. X be continuous. Fix $\epsilon > 0$, then we have*

$$\mathbb{P}\left[\left|\mathrm{V}_\beta - \widehat{\mathrm{V}}_{n,\beta}\right| \geq \epsilon\right] \leq 2\exp\left(-2n\bar{c}\epsilon^2\right)$$

where $\bar{c}$ is a constant that depends on the value of the density f of the r.v. X in a neighborhood of V_β*.*

Proof: See Proposition 2 in (Kolla *et al.*, 2019). □

Proof of Theorem 2.3.1. First, we prove the claim in part (i). Notice that

$$\begin{aligned}
\mathbb{P}\left[\left|\mathrm{S} - \widehat{\mathrm{S}}_{n,m}\right| > \epsilon\right] &= \mathbb{P}\left[\left|\int_0^1 \varphi(\beta)\mathrm{V}_\beta \, d\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \varphi(\beta_k)\widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \epsilon\right] \\
&= \mathbb{P}\left[\left|\int_0^1 \varphi(\beta)\mathrm{V}_\beta \, d\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\mathrm{V}_{\beta_{k-1}} + \varphi(\beta_k)\mathrm{V}_{\beta_k}}{2}\Delta\beta\right.\right. \\
&\quad + \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\mathrm{V}_{\beta_{k-1}} + \varphi(\beta_k)\mathrm{V}_{\beta_k}}{2}\Delta\beta \\
&\quad \left.\left. - \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \varphi(\beta_k)\widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \epsilon\right] \\
&\leq \mathbb{P}\left[\left|\sum_{k=1}^m \frac{\varphi(\beta_{k-1})\mathrm{V}_{\beta_{k-1}} + \varphi(\beta_k)\mathrm{V}_{\beta_k}}{2}\Delta\beta\right.\right. \\
&\quad \left.\left. - \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \varphi(\beta_k)\widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \frac{\epsilon}{2}\right],
\end{aligned}$$

where the final inequality follows by using Lemma 2.4.1(i) to infer that for $m \geq \frac{K_1}{2\epsilon}$, we have $\left|\int_0^1 \varphi(\beta)\mathrm{V}_\beta \, d\beta - \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\mathrm{V}_{\beta_{k-1}}+\varphi(\beta_k)\mathrm{V}_{\beta_k}}{2}\Delta\beta\right| < \frac{\epsilon}{2}$. Now, we have

$$\begin{aligned}
\mathbb{P}\left[\left|\mathrm{S}-\widehat{\mathrm{S}}_{n,m}\right|>\epsilon\right] &\leq \mathbb{P}\left[\left|\sum_{k=1}^m \frac{\varphi(\beta_{k-1})\mathrm{V}_{\beta_{k-1}}+\varphi(\beta_k)\mathrm{V}_{\beta_k}}{2}\Delta\beta\right.\right. \\
&\quad \left.\left. - \sum_{k=1}^m \frac{\varphi(\beta_{k-1})\widehat{\mathrm{V}}_{n,\beta_{k-1}}+\varphi(\beta_k)\widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \frac{\epsilon}{2}\right] \\
&= \mathbb{P}\left[\left|\sum_{k=1}^m((\varphi(\beta_{k-1})\mathrm{V}_{\beta_{k-1}}+\varphi(\beta_k)\mathrm{V}_{\beta_k})\right.\right. \\
&\quad \left.\left. - (\varphi(\beta_{k-1})\widehat{\mathrm{V}}_{n,\beta_{k-1}}+\varphi(\beta_k)\widehat{\mathrm{V}}_{n,\beta_k}))\right| > \frac{\epsilon}{\Delta\beta}\right] \\
&= \mathbb{P}\left[\left|\varphi(\beta_0)\mathrm{V}_{\beta_0}-\varphi(\beta_0)\widehat{\mathrm{V}}_{n,\beta_0}+2(\varphi(\beta_1)\mathrm{V}_{\beta_1}-\varphi(\beta_1)\widehat{\mathrm{V}}_{n,\beta_1})\right.\right. \\
&\quad + \cdots + 2(\varphi(\beta_{m-1})\mathrm{V}_{\beta_{m-1}}-\varphi(\beta_{m-1})\widehat{\mathrm{V}}_{n,\beta_{m-1}}) \\
&\quad \left.\left. + \varphi(\beta_m)\mathrm{V}_{\beta_m}-\varphi(\beta_m)\widehat{\mathrm{V}}_{n,\beta_m}\right| > \frac{\epsilon}{\Delta\beta}\right] \\
&\leq \mathbb{P}\left[\left|\varphi(\beta_0)\mathrm{V}_{\beta_0}-\varphi(\beta_0)\widehat{\mathrm{V}}_{n,\beta_0}\right| > \frac{\epsilon}{2m\,\Delta\beta}\right] \\
&\quad + 2\mathbb{P}\left[\left|\varphi(\beta_1)\mathrm{V}_{\beta_1}-\varphi(\beta_1)\widehat{\mathrm{V}}_{n,\beta_1}\right| > \frac{\epsilon}{2m\,\Delta\beta}\right] + \cdots + \\
&\quad 2\mathbb{P}\left[\left|\varphi(\beta_{m-1})\mathrm{V}_{\beta_{m-1}}-\varphi(\beta_{m-1})\widehat{\mathrm{V}}_{n,\beta_{m-1}}\right| > \frac{\epsilon}{2m\,\Delta\beta}\right] \\
&\quad + \mathbb{P}\left[\left|\varphi(\beta_m)\mathrm{V}_{\beta_m}-\varphi(\beta_m)\widehat{\mathrm{V}}_{n,\beta_m}\right| > \frac{\epsilon}{2m\,\Delta\beta}\right]
\end{aligned}$$

We now apply Lemma 2.4.2 to bound each of the terms on the RHS above, to obtain

$$\begin{aligned}
\mathbb{P}\left[\left|\mathrm{S}-\widehat{\mathrm{S}}_{n,m}\right|>\epsilon\right] &\leq 2\exp\left(-2nc_0\left(\frac{\epsilon}{2\,m\varphi(\beta_0)\Delta\beta}\right)^2\right) \\
&\quad + 4\exp\left(-2nc_1\left(\frac{\epsilon}{2\,m\varphi(\beta_1)\Delta\beta}\right)^2\right) \\
&\quad + \cdots + 4\exp\left(-2nc_{m-1}\left(\frac{\epsilon}{2\,m\varphi(\beta_{m-1})\Delta\beta}\right)^2\right) \\
&\quad + 2\exp\left(-2nc_m\left(\frac{\epsilon}{2\,m\varphi(\beta_m)\Delta\beta}\right)^2\right),
\end{aligned}$$

where c_i is a constant that depends on the value of the density f in the neighborhood of V_{β_i}, for i = $0\ldots m$. Thus,

$$\mathbb{P}\left[\left|\mathrm{S}-\widehat{\mathrm{S}}_{n,m}\right|>\epsilon\right] \leq 4m\exp\left(-2nc\left(\frac{\epsilon}{2\,mC_1\Delta\beta}\right)^2\right) \tag{2.8}$$

$$= 4m \exp\left(-\frac{n\,c\,\epsilon^2}{2C_1^2}\right) = \frac{2K_1}{\epsilon}\,\exp\left(-\frac{n\,c\,\epsilon^2}{2C_1^2}\right).$$

Note that $c = \min\{c_0, c_1, \ldots, c_m\}$ in (2.8). The claim in part (i) follows.

The proof of the result in part (ii) follows in a similar manner. In particular, using part (ii) in Lemma 2.4.1, with $m \geq \sqrt{\frac{K_2}{6\epsilon}}$, we obtain

$$\begin{aligned}
\mathbb{P}\left[\left|\mathrm{S} - \widehat{\mathrm{S}}_{n,m}\right| > \epsilon\right] &\leq 4m \exp\left(-\frac{n\,c\,\epsilon^2}{2C_1^2}\right) \\
&= \sqrt{\frac{8K_2}{3\epsilon}}\,.\exp\left(-\frac{n\,c\,\epsilon^2}{2C_1^2}\right)
\end{aligned}$$

□

2.4.2 Proof of Theorem 2.3.2

Proof. Recall that the truncation threshold $B_n = \sqrt{2\sigma^2 \log(n)}$. Letting $\eta = F(B_n)$, we have

$$\begin{aligned}
\mathbb{P}\left[\mathrm{S} - \widetilde{\mathrm{S}}_{n,m} > \epsilon\right] &\leq \mathbb{P}\left[\int_0^1 \varphi(\beta)\mathrm{V}_\beta\,d\beta - \sum_{k=1}^{m} \frac{\varphi(\beta_{k-1})\widetilde{\mathrm{V}}_{n,\beta_{k-1}} + \varphi(\beta_k)\widetilde{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta > \epsilon\right] \\
&= \mathbb{P}\left[\int_0^\eta \varphi(\beta)\mathrm{V}_\beta\,d\beta - \sum_{k=1}^{m} \frac{\varphi(\beta_{k-1})\widetilde{\mathrm{V}}_{n,\beta_{k-1}} + \varphi(\beta_k)\widetilde{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta \right.\\
&\qquad \left. + \int_\eta^1 \varphi(\beta)\mathrm{V}_\beta\,d\beta > \epsilon\right] \\
&= \mathbb{P}\left[I_1 + I_2 > \epsilon\right],
\end{aligned} \tag{2.9}$$

where $I_1 = \int_0^\eta \varphi(\beta)\mathrm{V}_\beta\,d\beta - \sum_{k=1}^{m} \frac{\varphi(\beta_{k-1})\widetilde{\mathrm{V}}_{n,\beta_{k-1}} + \varphi(\beta_k)\widetilde{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta$, and $I_2 = \int_\eta^1 \varphi(\beta)\mathrm{V}_\beta\,d\beta$. We bound I_2 as follows:

$$1 - \beta = \mathbb{P}\left(X > \mathrm{V}_\beta\right) \leq \exp\left(-\frac{{\mathrm{V}_\beta}^2}{2\sigma^2}\right), \tag{2.10}$$

since X is Gaussian with mean zero, and variance σ^2. Using $\log x \leq \frac{x}{e}\ \forall x > 0$, we obtain

$$\mathrm{V}_\beta \leq \sqrt{2\sigma^2 \log\left(\frac{1}{1-\beta}\right)} \leq \sqrt{\frac{2\sigma^2}{e(1-\beta)}},$$

leading to

$$
\begin{aligned}
\int_{\eta}^{1} \mathrm{V}_{\beta}\, d\beta &\leq \sqrt{\frac{2\sigma^2}{e}} \int_{\eta}^{1} \frac{d\beta}{\sqrt{1-\beta}} = 2\sqrt{\frac{2\sigma^2}{e}}\sqrt{1-\eta} \\
&\leq 2\sqrt{\frac{2\sigma^2}{e}} \exp\left(-\frac{{\mathrm{V}_{\eta}}^2}{4\sigma^2}\right) && \text{(using (2.10))} \\
&= 2\sqrt{\frac{2\sigma^2}{e}} \exp\left(-\frac{{B_n}^2}{4\sigma^2}\right) && \text{(since } \mathrm{V}_{\eta} = B_n\text{)}
\end{aligned}
$$

Hence,

$$
I_2 = \int_{\eta}^{1} \varphi(\beta)\mathrm{V}_{\beta}\, d\beta \leq C_1 \int_{\eta}^{1} \mathrm{V}_{\beta}\, d\beta \leq \frac{2\sigma C_1}{\sqrt{n}}. \tag{2.11}
$$

Applying the bound in the Theorem 2.3.1 to the truncated r.v. $Z = X\mathbb{I}\left\{X \leq B_n\right\}$, we bound I_1 as follows:

$$
\mathbb{P}\left[I_1 > \epsilon\right] \leq \frac{K_1}{\epsilon} \exp\left(-\frac{nc\epsilon^2}{2C_1^2}\right). \tag{2.12}
$$

Hence,

$$
\begin{aligned}
\mathbb{P}\left[I_1 + I_2 > \epsilon\right] &\leq \frac{K_1}{\left(\epsilon - \frac{2\sigma C_1}{\sqrt{n}}\right)} \exp\left(-\frac{nc\left(\epsilon - \frac{2\sigma C_1}{\sqrt{n}}\right)^2}{2C_1^2}\right) && \text{(using (2.11) and (2.12))} \\
&= \frac{(B_n\, C_2 + \delta_1\, C_1)}{\left(\epsilon - \frac{2\sigma C_1}{\sqrt{n}}\right)} \exp\left(-\frac{nc\left(\epsilon - \frac{2\sigma C_1}{\sqrt{n}}\right)^2}{2C_1^2}\right) \\
&\leq \frac{\sqrt{2}\sigma\left[\sqrt{\log(n)}\, C_2 + \sqrt{\pi}n\, C_1\right]}{\left(\epsilon - \frac{2\sigma C_1}{\sqrt{n}}\right)} \exp\left[-\frac{nc\left[\epsilon - \frac{2\sigma C_1}{\sqrt{n}}\right]^2}{2C_1^2}\right],
\end{aligned}
$$

where the final inequality follows from the fact that $\delta_1 = \sqrt{2\pi\sigma^2}\exp\left(\frac{B_n^2}{2\sigma^2}\right) = \sqrt{2\pi\sigma^2}n$, which holds since the underlying Gaussian distribution is truncated at B_n.

By using a parallel argument, a concentration result for bounding the lower semi-deviations can be derived, and we omit the details. □

2.4.3 Proof of Theorem 2.3.3

Proof. The proof for the exponential case follows in a similar manner as that of the proof of Theorem 2.3.2. In particular, the proof up to (2.12) holds for the exponential case, with a different bound on I_2.

We derive the bound on $I_2 = \int_\eta^1 \varphi(\beta)\mathrm{V}_\beta \, d\beta$. Using arguments similar to that in the Gaussian case, we obtain

$$\begin{aligned}
\int_\eta^1 \mathrm{V}_\beta \, d\beta &\leq \frac{1}{\lambda}\int_\eta^1 \log\left(\frac{1}{1-\beta}\right) d\beta \\
&= \frac{(1-\eta)}{\lambda}\left(1+\log\left(\frac{1}{1-\eta}\right)\right) \\
&\leq \frac{(1-\eta)}{\lambda}\left(1+\frac{1}{(1-\eta)e}\right) \\
&= \frac{\exp\left(-\lambda \mathrm{V}_\eta\right)}{\lambda}\left(1+\exp\left(\lambda \mathrm{V}_\eta - 1\right)\right) \\
&= \frac{\exp\left(-\lambda B_n\right)}{\lambda} + \frac{1}{\lambda e}.
\end{aligned}$$

The final inequality hold since $\mathrm{V}_\eta = B_n$.
Also, we have

$$\int_\eta^1 \varphi(\beta)\mathrm{V}_\beta \, d\beta \leq C_1 \int_\eta^1 \mathrm{V}_\beta \, d\beta.$$

Choosing $B_n = \frac{\log(n)}{\lambda}$, we obtain

$$I_2 = \int_\eta^1 \varphi(\beta)\mathrm{V}_\beta \, d\beta \leq \frac{C_1(n+1)}{\lambda n} \tag{2.13}$$

Now, as in the proof of Theorem 2.3.2, we have

$$\mathbb{P}\left[I_1 > \epsilon\right] \leq \frac{K_1}{\epsilon}\exp\left(-\frac{nc\epsilon^2}{2C_1^2}\right). \tag{2.14}$$

Thus,

$$\mathbb{P}\left[I_1 + I_2 > \epsilon\right] \leq \frac{K_1}{\left(\epsilon - \frac{C_1(n+1)}{\lambda n}\right)}\exp\left(-\frac{nc\left(\epsilon - \frac{C_1(n+1)}{\lambda n}\right)^2}{2C_1^2}\right)$$

(using (2.13) and (2.14))

$$
\begin{aligned}
&= \frac{(B_n\, C_2 + \delta_1\, C_1)}{\left(\epsilon - \frac{C_1(n+1)}{\lambda n}\right)} \exp\left(- \frac{nc\left(\epsilon - \frac{C_1(n+1)}{\lambda n}\right)^2}{2C_1^2} \right) \\
&\leq \frac{\left(\frac{log(n)\, C_2}{\lambda} + n\, C_1\right)}{\left(\epsilon - \frac{C_1(n+1)}{\lambda n}\right)} \exp\left(- \frac{nc\left(\epsilon - \frac{C_1(n+1)}{\lambda n}\right)^2}{2C_1^2} \right),
\end{aligned}
$$

where the final inequality follows from the fact that $\delta_1 = \exp(\lambda B_n) = n$, which holds since the underlying exponential distribution is truncated at B_n.

By using a parallel argument, a concentration result for bounding the lower semi-deviations can be derived, and we omit the details. □

2.5 CVaR results

This section includes CVaR related results. CVaR is a specific case of SRM, and it can be recovered by setting risk-aversion function $\varphi(\beta)$ as in (1.8). In particular, we discuss the estimation technique for CVaR defined in (1.5), together with concentration bounds and their convergence proofs for both the cases when the underlying distribution has bounded support, and unbounded support of either Gaussian or exponential.

2.5.1 CVaR Estimation scheme: Bounded case

Here, we propose to estimate $\mathrm{C}_\alpha(X)$, given n i.i.d. samples $X_1, \ldots, X_n$ from the distribution of X, by approximating the integral in Acerbi's formula. Notice that the integrand V_β in (1.5) has to be estimated using the samples. Let $\widehat{\mathrm{V}}_{n,\beta}$ denote the estimate of $\mathrm{V}_\beta(X)$, as given in (1.1). We use the VaR estimates to form a discrete sum to approximate the integral in Acerbi's formula, an idea motivated by the trapezoidal rule (Cruz-Uribe and Neugebauer, 2003). More precisely, the estimate $\widehat{\mathrm{C}}_{n,m,\alpha}$ of $\mathrm{C}_\alpha(X)$ is formed as follows:

$$
\widehat{\mathrm{C}}_{n,m,\alpha} = \frac{1}{1-\alpha} \sum_{k=1}^{m} \frac{\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \widehat{\mathrm{V}}_{n,\beta_k}}{2} \Delta\beta. \tag{2.15}
$$

In the above, $\{\beta_k\}_{k=0}^m$ is a partition of $[\alpha, 1]$ such that $\beta_0 = \alpha$ and $\beta_k = \beta_{k-1} + \Delta\beta$, where $\Delta\beta = (1-\alpha)/m$ is the length of each sub-interval.

2.5.2 CVaR Estimation scheme: Unbounded case

Let $X_1, \ldots, X_n$ denote i.i.d. samples from the distribution of X. We form a truncated set of samples as follows:

$$\bar{X}_i = X_i \mathbb{I}\left\{X_i \leq B_n\right\},$$

where B_n is a truncation threshold that depends on the underlying distribution. For the case of Gaussian distribution with mean zero and variance σ^2, $B_n = \sqrt{2\sigma^2 \log(n)}$, and for the case of exponential distribution with mean $1/\lambda$, $B_n = \frac{\log(n)}{\lambda}$.

We form a CVaR estimate along the lines of (2.15), except that the samples used are truncated samples, i.e.,

$$\widetilde{\mathrm{C}}_{n,m,\alpha} = \frac{1}{1-\alpha} \sum_{k=1}^{m} \frac{\widetilde{\mathrm{V}}_{n,\beta_{k-1}} + \widetilde{\mathrm{V}}_{n,\beta_k}}{2} \Delta\beta. \tag{2.16}$$

In the above, $\widetilde{\mathrm{V}}_{n,\beta} = \widetilde{F}_n^{-1}(\beta)$, with $\widetilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[\bar{X}_i \leq x]$.

2.5.3 Concentration bounds

In this section, we present the concentration bounds for CVaR estimator in Corollaries 2.5.1 to 2.5.3, which are obtained similarly as SRM concentration bounds in Theorems 2.3.1 to 2.3.3, respectively.

For notational convenience, we shall use V_α and C_α to denote $\mathrm{V}_\alpha(X)$ and $\mathrm{C}_\alpha(X)$, for any $\alpha \in (0,1)$.

Distributions with bounded support

For all the results presented below, we take CVaR estimate as $\widehat{\mathrm{C}}_{n,m,\alpha},\ \ \alpha \in [0,1]$ be formed from n i.i.d. samples of X using (2.15). Let F and f denote the distribution and density of X, respectively.

Corollary 2.5.1 (CVaR concentration: bounded case). *Let the r.v. X be continuous and $X \leq B$ a.s. Fix $\epsilon > 0$.*

(i) *If* $f(x) \geq \frac{1}{\delta_1} > 0$, $\forall x \in [F^{-1}(\alpha), B]$, *and* $m \geq \frac{K_1\,(1-\alpha)}{2\epsilon}$, *then*

$$\mathbb{P}[\mid \mathrm{C}_\alpha - \widehat{\mathrm{C}}_{n,m,\alpha} \mid > \epsilon] \leq \frac{2K_1(1-\alpha)}{\epsilon} \exp\left(\frac{-n\,c\,\epsilon^2}{2}\right),$$

where $c = \min\{c_0, c_1, \ldots, c_m\}$ *and* $c_k, k = \{0, \ldots, m\}$ *is a constant that depends on the value of the density* f *of the r.v.* X *in a neighborhood of* V_{β_k}, *with* β_k *as in* (2.15).

(ii) *If* $\frac{\left|f^{'}(x)\right|}{f(x)^3} \leq \frac{1}{\delta_2}$, $\forall x \in [F^{-1}(\alpha), B]$ *and* $m \geq \sqrt{\frac{K_2\,(1-\alpha)^2}{6\epsilon}}$, *then*

$$\mathbb{P}[\mid \mathrm{C}_\alpha - \widehat{\mathrm{C}}_{n,m,\alpha} \mid > \epsilon] \leq \sqrt{\frac{8\,K_2(1-\alpha)^2}{3\epsilon}} \exp\left(\frac{-n\,c\,\epsilon^2}{2}\right),$$

where c *is as in the case above.*

Proof. See Appendix B.1. □

Gaussian and exponential distributions

Here, we present concentration bounds for our CVaR estimator assuming that the samples are either from a Gaussian distribution with mean zero and variance σ^2, or from the exponential distribution with mean $1/\lambda$. Note that the estimation scheme is not provided this information about the underlying distribution. Instead $\widetilde{\mathrm{C}}_{n,m,\alpha}$ is formed from n i.i.d. samples and with m sub-intervals, using (2.16).

Corollary 2.5.2 (CVaR concentration: Gaussian case). *Suppose that the r.v.* X *is Gaussian with mean zero and variance* $\sigma^2 > 0$, *with* $\sigma \leq \sigma_{max}$. *Fix* $\epsilon > 0$. *If* $m \geq \frac{1}{5}\sqrt{\frac{\sigma_{max}(1-\alpha)}{\epsilon}} \exp\left(\frac{nc\epsilon^2}{4}\right)$, *then*

$$\mathbb{P}\left[\left|\mathrm{C} - \widetilde{\mathrm{C}}_{n,m,\alpha}\right| > \epsilon\right] \leq \frac{2(1-\alpha)\sigma\sqrt{2\pi}n}{\left(\epsilon - \frac{2\sigma}{(1-\alpha)\sqrt{n}}\right)} \exp\left(-\frac{nc\left(\epsilon - \frac{2\sigma}{(1-\alpha)\sqrt{n}}\right)^2}{2}\right),$$

for $\epsilon > \dfrac{2\,\sigma_{max}}{(1-\alpha)\sqrt{n}}$.

In the above, c *is as in Theorem 2.5.1 (i).*

Proof. See Appendix B.2. □

Corollary 2.5.3 (CVaR concentration: Exponential case). *Assume r.v.* $X \sim Exp(\lambda)$ *and* $0 < \lambda_{min} \leq \lambda$. *Fix* $\epsilon > 0$. *If* $m \geq \frac{1}{8}\sqrt{\frac{(1-\alpha)}{\lambda_{min}\epsilon}} \exp\left(\frac{nc\epsilon^2}{4}\right)$, *then we have*

$$\mathbb{P}\left[\left|\mathrm{C} - \widetilde{\mathrm{C}}_{n,m,\alpha}\right| > \epsilon\right] \leq \frac{2(1-\alpha)n}{\left(\epsilon - \frac{(n+1)}{(1-\alpha)\lambda n}\right)} \exp\left(-\frac{nc\left(\epsilon - \frac{(n+1)}{(1-\alpha)\lambda n}\right)^2}{2}\right),$$

$$\textit{for } \epsilon > \frac{(n+1)}{(1-\alpha)\lambda_{min}\, n}.$$

In the above, c *is as in Theorem 2.5.1 (i).*

Proof. See Appendix B.3. □

2.6 Summary

In this chapter, we proposed a novel SRM estimation method. We provided a two-sided concentration bound to support our estimation method, for the case when the underlying distribution either has a bounded support, or is unbounded, but either Gaussian or exponential. Our tail bounds are of the order $O\left(c_1 \exp\left(-c_2 n\epsilon^2\right)\right)$, where n is the number of samples, ϵ is the accuracy parameter, and c_1, c_2 are universal constants. Further, we found the same tail bounds for the CVaR estimator also.

In the next chapter, we will see experiments and applications based on our proposed SRM and CVaR estimator.

CHAPTER 3

SIMULATION EXPERIMENTS

In this chapter, we present the simulation experiments for SRM and CVAR. In particular, we first present the synthetic experiments, after that, we present the applications of vehicular traffic routing using SUMO and portfolio optimization problem. The online supplementary material (link) contains the data and the code for the experiments that will allow the reader to reproduce our experimental results.

3.1 SRM experiments

In this section, we demonstrate the efficacy of our proposed method for SRM estimation (2.1), which we shall refer to as SRM-Trapz. In our experiments, we set the risk aversion function as follows: $\varphi(\beta) = \frac{5\,e^{-5(1-\beta)}}{1-e^{-5}}, \beta \in [0,1]$. In the following sub-section, we consider a synthetic experimental setting to compare the accuracy of SRM estimators. Subsequently, we use SRM-Trapz as a subroutine in a vehicular traffic routing application (see Section 3.1.2).

3.1.1 Synthetic setup

Figure 3.1 presents the estimation error as a function of the sample size for SRM-Trapz. The algorithm is run with two different sub-divisions. The samples are generated using a Gaussian distribution with mean 0.5 and variance 25. We observe that SRM-Trapz with 500 subdivisions performs on par with SRM-Trapz with 150 subdivisions for every sample size. Further, as expected, increasing sample size leads to lower estimation error, while also increasing the confidence (demonstrated by the shrinkage in standard error).

Table 3.1 presents the results obtained by SRM-Trapz with 1000 subdivisions, for four different input distributions. We observe that SRM-Trapz is comparable to SRM-True (calculated using definition 1.7) under each input distribution.

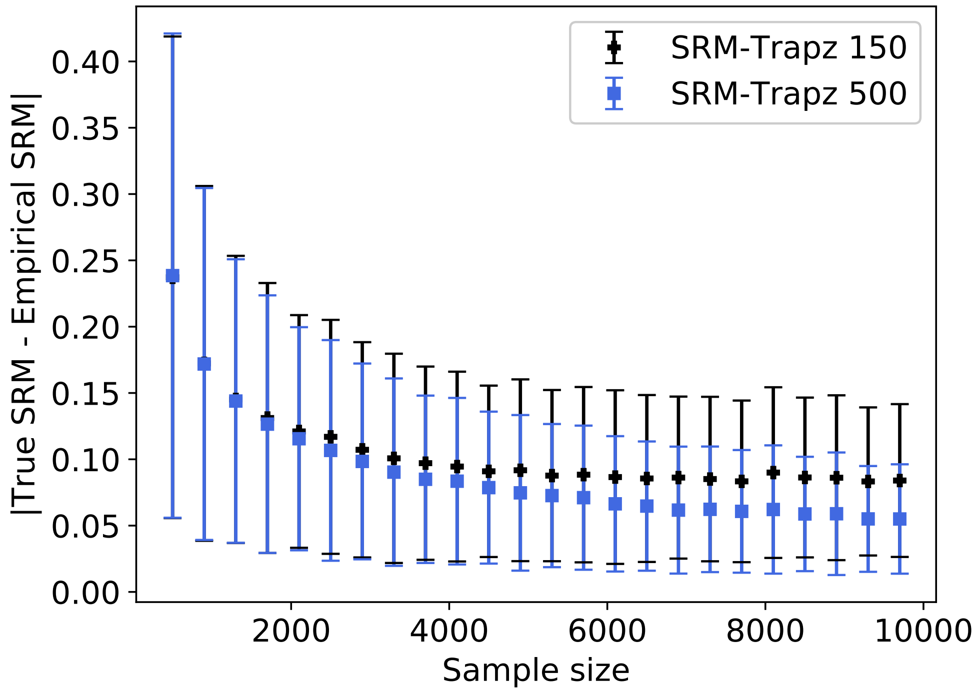

Figure 3.1: Error in SRM estimation (|True SRM - Empirical SRM|) on different sample size. True SRM is calculated using definition 1.7. Empirical SRM is calculated by two methods, **(i)** SRM-Trapz method with $m = 150$ subdivisions (SRM-Trapz 150), and **(ii)** SRM-Trapz method with $m = 500$ subdivisions (SRM-Trapz 500). In both methods, SRM is estimated using (2.1). The underlying distribution considered for this simulation is $X \sim \mathcal{N}(0.5, 5^2)$. The bars in the plot shows standard error averaged over 10^3 iterations.

Table 3.1: The results for SRM estimation, on four distributions, using two methods. Distributions are **(a)** Exponential distribution with mean $1/0.2$ (Exp(0.2)), **(b)** Normal distribution with mean zero and variance 10^2 ($\mathcal{N}(0, 10^2)$), **(c)** Exponential distribution with mean $1/0.01$ (Exp(0.01)), **(d)** Uniform distribution with range -10^3 to 10^3 (U$(-10^3, 10^3)$). Methods are **(i)** Calculation of SRM (SRM-True) using definition 1.7, **(ii)** SRM-Trapz method with $m = 1000$ subdivisions (SRM-Trapz 1000) using (2.1). In method (ii), 10^4 i.i.d. samples are used for estimating SRM on each distribution, and the standard error is averaged over 10^3 iterations.

Distribution	SRM-True	SRM 1000
Exp(0.2)	10.99	11.02$\pm$1.21
N(0, 10^2)	107.36	107.80$\pm$1.32
Exp(0.01)	221.30	221.39$\pm$2.47
U$(-10^3, 10^3)$	612.47	612.65$\pm$4.91

3.1.2 Vehicular traffic routing

In the vehicular routing application, the traditional objective is to find a route with the lowest expected delay. However, such an objective ignores risk factors. An alternative is to consider the weighted-sum delay of each route, and we use SRM to quantify this objective. Thus, given a set of routes, the aim is to find the route (by adaptive sampling) with the lowest SRM of the delay. Simulation of Urban MObility (SUMO) (Behrisch *et al.*, 2011) is an open source, highly portable, microscopic road traffic simulation package designed to handle large road networks. Traffic Control Interface (TraCI) (Wegener *et al.*, 2008) is a library, providing extensive commands to control the behavior of the simulation online, including vehicle state, road configuration, and traffic lights. We implement our routing algorithm using SUMO and TRACI.

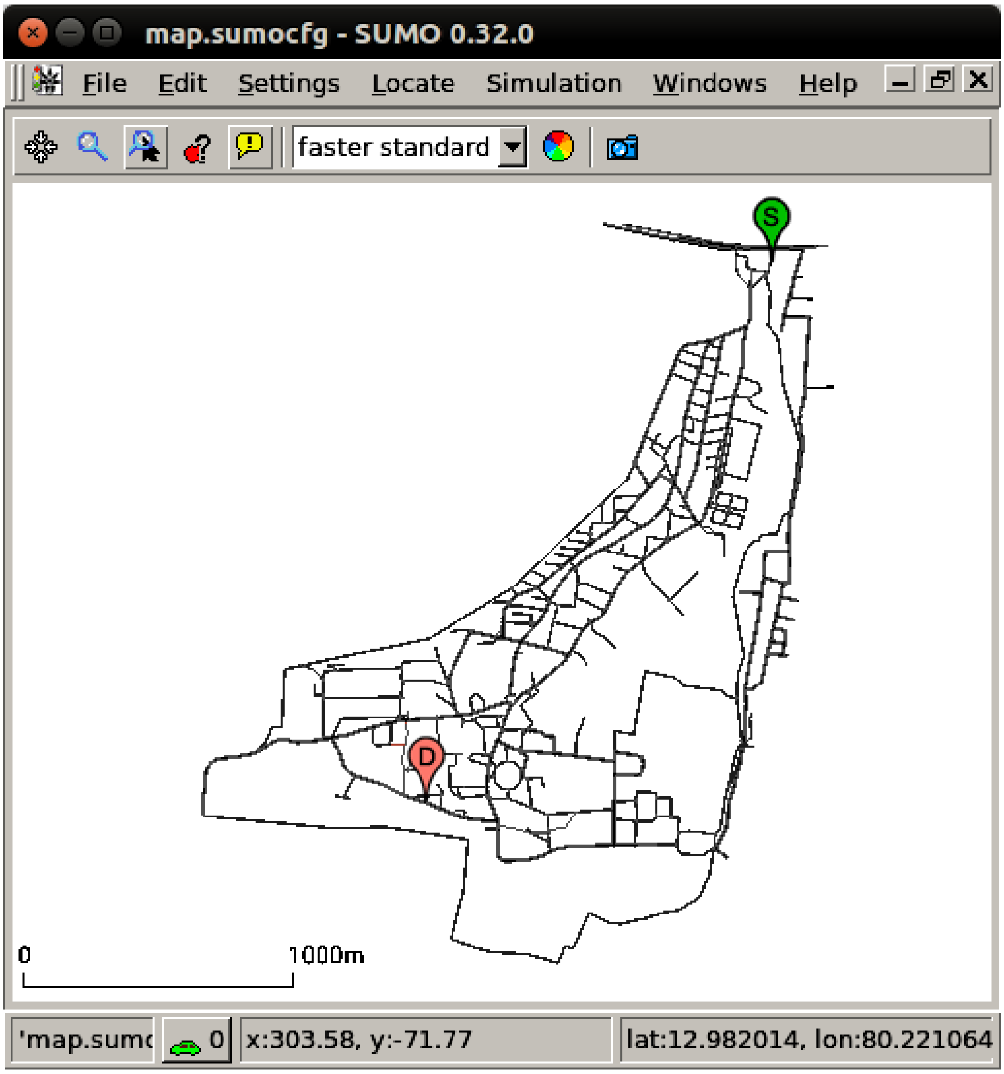

Figure 3.2: Area of an urban city map, used for SUMO network.

Algorithm 1 SRM-SR-Trapz

Input: number of rounds n, number of routes K, number of subdivisions m.

Let $A_1 = \{1, \ldots, K\}, \overline{\log}(K) = \frac{1}{2} + \sum_{i=2}^{K} \frac{1}{i}, n_0 = 0$ and for $k \in \{1, \ldots, K-1\}, n_k = \left\lceil \frac{1}{\overline{\log}(K)} \frac{n-K}{K+1-k} \right\rceil$

For each phase $k = 1, 2, \ldots, K-1$:
(1) For each $i \in A_k$, select route i for $n_k - n_{k-1}$ rounds.
(2) Let $A_{k+1} = A_k \backslash \arg\max_{i \in A_k} \widehat{\mathrm{S}}_{n_k,m,i}$ (we only remove one element from A_k if there is a tie, select randomly the route to dismiss among the worst routes).

Output: Let i^* be the unique element of A_K.

$\widehat{\mathrm{S}}_{n_k,m,i}$ is SRM estimate for ith route, using (2.1) with n_k samples, and m subdivisions.

For the experiments, we use the street map of the area around IIT Madras, Chennai, India (see Figure 3.2) obtained from OpenStreetMap (OSM) (Haklay and Weber, 2008), and then used Netconvert tool to load the map in SUMO. The network has 426 junctions and a total edge length of 123 km. We ran SUMO on this network for $30,000$ time-steps, in which 7000 cars, 500 buses, 2000 bikes, 1000 cycles, and 1000 pedestrians were added at different time-steps and in different lanes uniformly. We choose $K = 5$ routes between two fixed points, marked as S and D in Figure 3.2. On these selected routes, we add $n = 1000$ cars and track them. In Table 3.2, $\widehat{X}_{n,i}$ is the estimated average delay of the ith route, and $\widehat{\mathrm{S}}_{n,m,i}$ is the SRM estimate for the ith route, $i = 1, \ldots, K$, using (2.1), and with n samples. We set the number of subdivisions $m = 100$.

Table 3.2: Results for the estimated average delay ($\hat{X}_{n,i}$) and estimated SRM ($\widehat{\mathrm{S}}_{n,m,i}$), for ith route, where $i = 1, \ldots, K$.

	ROUTE$_1$	ROUTE$_2$	ROUTE$_3$	ROUTE$_4$	ROUTE$_5$
$\hat{X}_{n,i}$	283.81	287.15	306.80	266.85	325.86
$\widehat{\mathrm{S}}_{n,m,i}$	431.28	361.81	455.83	378.68	390.95

From Table 3.2 it is apparent that ROUTE$_4$ has the lowest expected delay, and ROUTE$_2$ has the lowest SRM. We consider a best-arm identification (BAI) bandit framework (Audibert *et al.*, 2010), where an algorithm is given a fixed budget. Here, the budget refers to the total number of samples across routes. After the sampling budget, the algorithm is expected to recommend a route, and is judged by the probability that the recommended route is correct (i.e. the best route). We ran successive rejects (SR) (Audibert *et al.*, 2010), which is a popular BAI algorithm, except that SR is modified to find the route with lowest SRM. Note that the regular SR algorithm finds the route with the lowest expected delay. Algorithm 1 presents the pseudocode for the SRM-SR-Trapz algorithm, with SRM-Trapz used to form SRM estimates for each route. The setting of SUMO is as noted above. We set the budget $n = 1000$, number of routes $K = 5$, and $m = 100$ subdivisions for SRM-Trapz. We observed that SR algorithm picks ROUTE_2 with probability 0.91.

Further, we tested algorithm 1 on a bigger network 3.3. The setting of SUMO are: number of rounds $n = 5000$,, number of routes $K = 25$, and $m = 300$ subdivisions for SRM-Trapz. After running the simulation, we found that the yellow route is having the lowest expected delay, but does not lowest SRM delay. Moreover, the SR algorithm picks the green route with a probability of 0.87 as the best route with the lowest SRM delay. It is with the fact that the green route is less risky as compared to the yellow route.

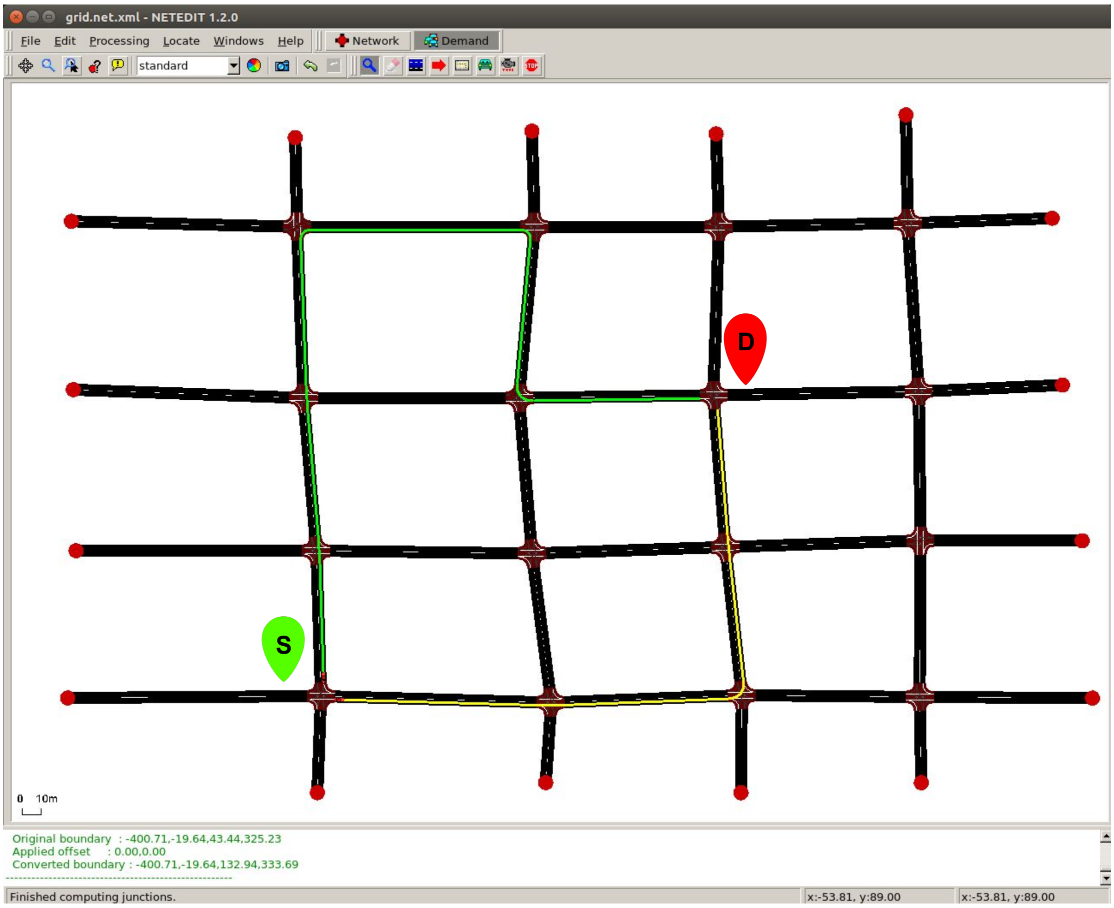

Figure 3.3: Grid network for SUMO.

3.2 CVaR experiments

In this Section, we demonstrate the efficacy of our proposed method for CVaR estimation (2.15), which we shall refer to as CVaR-Trapz. In the following sub-section, we consider a synthetic experimental setting to compare the accuracy of CVaR estimators. Subsequently, we use CVaR-Trapz in a portfolio optimization problem (see Section 3.2.2).

3.2.1 Synthetic setup

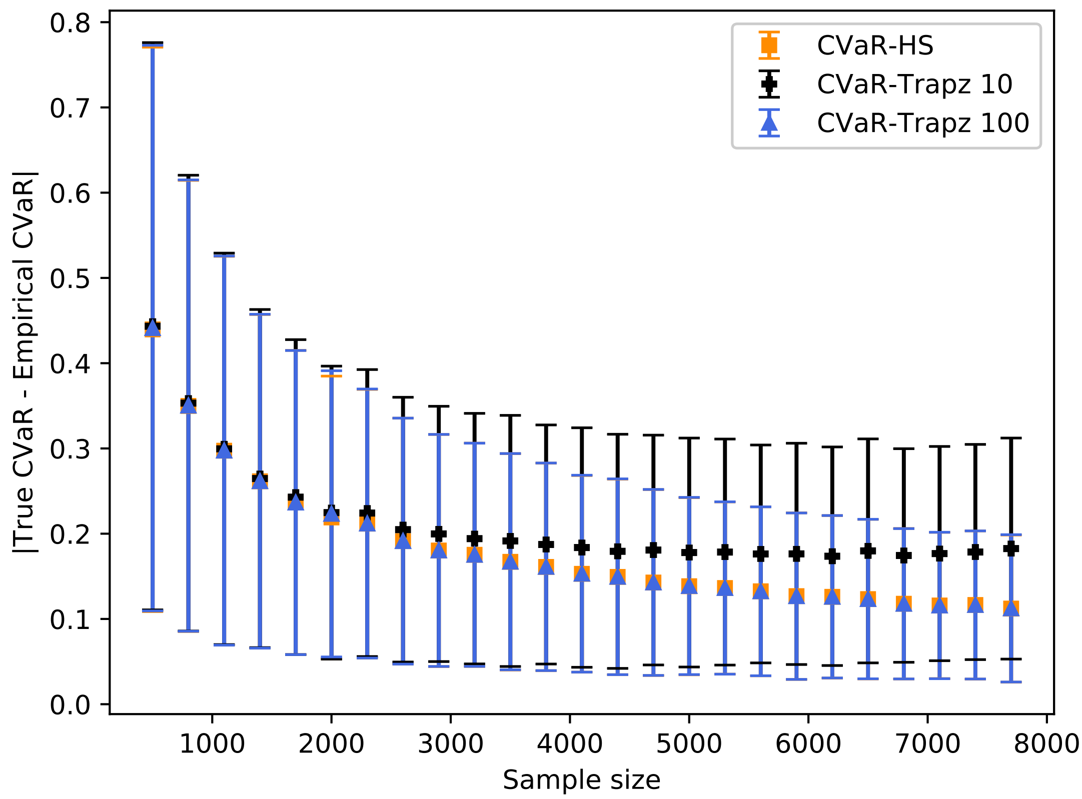

Figure 3.4: Error in CVaR estimation (|True CvaR - Empirical CVaR|) at confidence level $\alpha = 0.95$ on different sample size. True CVaR is calculated using definition 1.4. Empirical CVaR is calculated by three methods, (**i**) Historical simulated method (CVaR-HS) which estimate CVaR using (1.6), (**ii**) CVaR-Trapz method with $m = 10$ subdivisions (CVaR-Trapz 10), and (**iii**) CVaR-Trapz method with $m = 100$ subdivisions (CVaR-Trapz 100). In methods (ii) and (iii), CVaR is estimated using (2.15). The underlying distribution considered for this simulation is $X \sim \mathcal{N}(0.5, 5^2)$. The bars in the plot shows standard error averaged over 5×10^3 iterations.

We compare the performance of CVaR-Trapz with CVaR-HS, which employs the classic estimator given in (1.6). Figure 3.4 presents the estimation error as a function of the sample size for CVaR-HS and CVaR-Trapz. The latter algorithm is run with two different sub-divisions. The samples are generated using a Gaussian distribution with mean 0.5 and variance 25. We observe that CVaR-Trapz with 100 subdivisions performs on par with CVaR-HS for every sample size. Further, as expected, increasing sample size leads to lower estimation error, while also increasing the confidence (demonstrated by the shrink in confidence intervals).

Table 3.3: The results for CVaR estimation at confidence level $\alpha = 0.95$, on four distributions, using three methods. Distributions are **(a)** Exponential distribution with mean $1/0.2$ (Exp(0.2)), **(b)** Normal distribution with mean zero and variance 10^2 ($\mathcal{N}(0, 10^2)$), **(c)** Exponential distribution with mean $1/0.01$ (Exp(0.01)), **(d)** Uniform distribution with range -10^3 to 10^3 (U$(-10^3, 10^3)$). Methods are **(i)** Calculation of CVaR (CVaR-True) using definition 1.4, **(ii)** Historical simulated method (CVaR-HS) using (1.6), **(iii)** CVaR-Trapz method with $m = 500$ subdivisions (CVaR-Trapz 500) using (2.15). In methods (ii) and (iii), 10^4 i.i.d. samples are used for estimating CVaR on each distribution, and the standard error is averaged over 10^3 iterations.

DISTRIBUTION	CVAR-TRUE	CVAR-HS	CVAR-TRAPZ 500
EXP(0.2)	19.97	20.03±1.12	20.01±1.12
N(0, 10^2)	206.27	206.67±2.64	206.45±2.62
EXP(0.01)	399.57	400.39±6.46	399.72±6.38
U(-10^3, 10^3)	950.00	950.97±2.65	950.87±2.65

Table 3.3 presents the results obtained by CVaR-HS and CVaR-Trapz with 500 subdivisions, for four different input distributions. We observe that CVaR-Trapz performs marginally better than CVaR-HS, under each input distribution.

3.2.2 Portfolio optimization

The typical goal in a portfolio optimization problem is to find an investment strategy that maximizes the expected return. In this experiment, we consider a risk-sensitive strategy that is based a CVaR-based objective. In particular, the aim is to find an investment strategy that minimizes the worst-case loss, while guaranteeing a minimum expected return, a problem considered earlier in (Rockafellar *et al.*, 2000).

In this problem, the investor can invest in d different assets, and his investment decision vector is denoted by $\mathbf{x}$, and the decision region is $\mathbf{X} \subset \mathbb{R}^d$. The returns of all assets are random, and losses are expressed by a random vector $\mathbf{y} \in \mathbb{R}^d$. The loss that an investor can experience for a decision vector $\mathbf{x}$ is a r.v. denoted by $f(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$. We consider the following portfolio optimization problem (Rockafellar *et al.*, 2000):

$$\begin{aligned} \min_{\mathbf{x} \in \mathbf{X}} \quad & \mathrm{C}_\alpha \left(\mathbf{x}^T \mathbf{y}\right) \\ \text{s.t.} \quad & -\mathbf{x}^T \mathbb{E}\left(\mathbf{y}\right) \geq R, \end{aligned} \tag{3.1}$$

Solving the problem formulated above requires knowledge of the distribution of loss $\mathbf{y}$, and this information is often unavailable in practice. Our solution approach is to obtain i.i.d. samples of the underlying losses, and approximately solve (3.1) by substituting an estimate of CVaR that is based on CVaR-Trapz in the objective of (3.1). The details are as follows. Given i.i.d. loss vector samples $\mathbf{y}_1, \ldots, \mathbf{y}_n$ from distribution of $\mathbf{y}$, assume loss for a portfolio $\mathbf{x}$ is $l_i = \mathbf{x}^T \mathbf{y}_i$, $i \in \{1, \ldots, n\}$. Let $l_{(1)}, \ldots, l_{(n)}$ denote the order statistics for losses, and $\hat{\mathrm{V}}_{n,\beta_k} = l_{(\lceil n\beta_k \rceil)}$ denotes estimate of V_{β_k}. Also, to ensure a minimum expected return R for the investor, we add a constraint -$\mathbf{x}^T \widehat{\mathbf{u}} \geq R$, where $\widehat{\mathbf{u}}$ denote estimate of $\mathbb{E}[\mathbf{y}]$. Then, we can write COP (3.2) as:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbf{X}} \quad & \frac{1}{(1-\alpha)} \sum_{k=1}^{m} \left(\frac{\hat{\mathrm{V}}_{n,\beta_{k-1}} + \hat{\mathrm{V}}_{n,\beta_k}}{2} \right) \Delta\beta \\ \text{s.t.} \quad & -\mathbf{x}^T \widehat{\mathbf{u}} \geq R, \end{aligned} \tag{3.2}$$

where α is the specified confidence level and m is the number of subdivisions used in CVaR-Trapz.

We considered weekly negative returns of three assets IBM, MSFT and WMT of S&P500 stocks during Nov 2004 - Apr 2016, taken from the real stock market data set of (Bruni *et al.*, 2016). Table 3.4 and Table 3.5 presents the mean loss and covariance matrix, obtained as an average over 595 negative return (loss) values of three assets. Table 3.6 presents the weights of three assets, while minimizing CVaR using (3.2), with CVaR-Trapz for CVaR estimation. For obtaining these results, $n = 20,000$ i.i.d. loss vector samples were used. The underlying distribution for sampling is set to multivariate normal with mean as given in Table 3.4 and covariance matrix as in Table 3.5. For solving the problem (3.2), candidate vectors $(x_1, x_2, x_3)^T \in \mathbb{R}^3$ satisfying $x_i >= 0$, $i \in \{1, 2, 3\}$ and $\sum_{i=1}^{3} x_i = 1$, were uniformly generated. We set the confidence parameter $\alpha = 0.95$, and number of subdivisions $m = 1000$.

Further, to see how the portfolio risk (CVaR) changes for different expected returns, we solved (3.2) approximately with parameters are mentioned above, for different values of expected minimum return R, and calculated the resulting portfolio risk. Figure 3.5 presents the plot of risk vs. return. The bars in the plot shows standard error averaged over 100 replications. We observe that CVaR-Trapz is comparable to CVaR-HS, which employs the classic estimator given in (1.6).

Table 3.4: Mean asset losses of three assets, S&P500 stocks

ASSET	MEAN LOSS
IBM	-0.0068772
MSFT	-0.0018255
WMT	-0.0013223

Table 3.5: Covariance matrix of three assets, S&P500 stocks

	IBM	MSFT	WMT
IBM	0.0024851	0.0007456	0.0006083
MSFT	0.0007456	0.0012761	0.0004629
WMT	0.0006083	0.0004629	0.0008997

Table 3.6: Portfolio configuration: assets' weights (%) in the optimal portfolio with minimum CVaR at confidence level 0.95 for different required returns.

REQUIRED RETURN	0.002	0.003	0.005
IBM	12.33%	30.06%	65.72%
MSFT	31.52%	27.32%	17.33%
WMT	56.15%	42.61%	16.94%

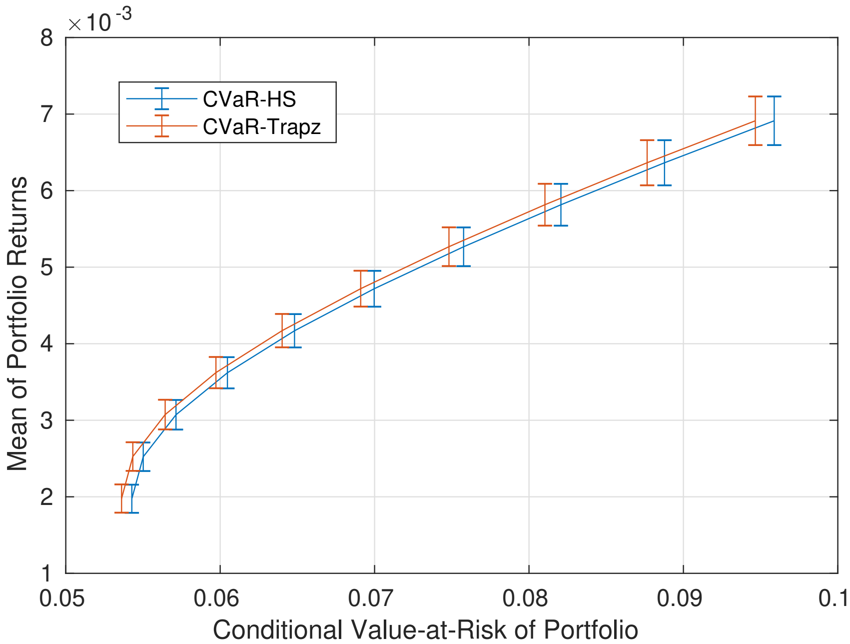

Figure 3.5: Efficient frontier (optimization with minimum CVaR constraint).

CHAPTER 4

CONCLUSIONS

We proposed a novel SRM estimation scheme that is based on numerical integration, and derived concentration bounds for our SRM estimator for the case of distributions with bounded support, as well as Gaussian and exponential distributions. We conducted simulation experiments to validate the theoretical findings. In particular, we studied a bandit application in vehicular traffic routing, and a financial application in portfolio optimization.

As future work, it would be interesting to generalize the bounds for Gaussian/ exponential distributions to the class of sub-Gaussian/sub-exponential distributions. An orthogonal direction for future work is to derive a lower bound for SRM estimation and close the gap (if any) w.r.t. the upper bound that we have derived.

Further, one can explore the application of this study in areas like clinical trials (which drug to prescribe), crowdsourcing (which task to allocate which worker, and at what price), logistic management (effective movement and storage of goods and services from origin to destination), and recommendation systems (news article, advertisement, and product).

APPENDIX A

PROOF OF LEMMA 2.4.1

Proof. The proof follows in a similar fashion as a result in Cruz-Uribe and Neugebauer (2003), and we provide the details below for the sake of completeness. Let $h = \Delta\beta = \frac{b-a}{m}$ and $\beta_k = a + kh$. We look at a single interval and operate integrate by parts twice:

$$\int_{\beta_k}^{\beta_{k+1}} \varphi(\beta)\mathrm{V}_\beta \, d\beta = \int_0^h \varphi(t+\beta_k)\mathrm{V}_{(t+\beta_k)} \, dt$$

$$= \left[(t+A)\varphi(t+\beta_k)\mathrm{V}_{(t+\beta_k)}\right]_0^h - \int_0^h (t+A)\left(\varphi(t+\beta_k)\mathrm{V}_{(t+\beta_k)}\right)' \, dt \tag{A.1}$$

$$\begin{aligned}
&= \left[(t+A)\varphi(t+\beta_k)\mathrm{V}_{(t+\beta_k)}\right]_0^h \\
&- \left[\left(\frac{(t+A)^2}{2} + B\right)\left(\varphi(t+\beta_k)\mathrm{V}_{(t+\beta_k)}\right)'\right]_0^h \\
&\quad + \int_0^h \left(\frac{(t+A)^2}{2} + B\right)\left(\varphi(t+\beta_k)\mathrm{V}_{(t+\beta_k)}\right)'' dt.
\end{aligned} \tag{A.2}$$

Setting $A = -h/2$ and $B = -h^2/8$ in the RHS above, we obtain

$$\begin{aligned}
\int_{\beta_k}^{\beta_{k+1}} \varphi(\beta)\mathrm{V}_\beta \, d\beta &= \frac{h\left(\varphi(\beta_k)\mathrm{V}_{\beta_k} + \varphi(\beta_{k+1})\mathrm{V}_{\beta_{k+1}}\right)}{2} \\
&\quad + \int_0^h \left(\frac{(t-h/2)^2}{2} - \frac{h^2}{8}\right)\left(\varphi(t+\beta_k)\mathrm{V}_{(t+\beta_k)}\right)'' \, dt
\end{aligned}$$

Let $E_T(k)$ denote the difference between the integral above and the corresponding trapezoid. Then, the error in the trapezoidal rule approximation can be simplified as follows:

$$\begin{aligned}
E_T &= E_T(0) + E_T(1) + \cdots + E_T(m-1) \\
&= \int_0^h \left(\frac{(t-h/2)^2}{2} - h^2/8\right)\left(\varphi(t+\beta_0)\mathrm{V}_{(t+\beta_0)}\right)'' \, dt + \cdots + \\
&\quad \int_0^h \left(\frac{(t-h/2)^2}{2} - h^2/8\right)\left(\varphi(t+\beta_{m-1})\mathrm{V}_{(t+\beta_{m-1})}\right)'' \, dt
\end{aligned}$$

$$= \int_0^h \left(\frac{(t-h/2)^2}{2} - h^2/8\right)\Big(\big(\varphi(t+\beta_0)\mathrm{V}_{(t+\beta_0)}\big)'' + \cdots + \big(\varphi(t+\beta_{m-1})\mathrm{V}_{(t+\beta_{m-1})}\big)''\Big)\,dt$$

As in the text, we suppose that $\left|(\varphi(\beta)\mathrm{V}_\beta)''\right| \leq K_2$ for $0 \leq \beta \leq 1$. Then,

$$\begin{aligned}
|E_T| &= \left|\int_0^h \left(\frac{(t-h/2)^2}{2} - h^2/8\right)\Big(\big(\varphi(t+\beta_0)\mathrm{V}_{(t+\beta_0)}\big)'' + \cdots + \big(\varphi(t+\beta_{m-1})\mathrm{V}_{(t+\beta_{m-1})}\big)''\Big)\,dt\right| \\
&\leq \int_0^h \left|\left(\frac{(t-h/2)^2}{2} - h^2/8\right)\Big(\big(\varphi(t+\beta_0)\mathrm{V}_{(t+\beta_0)}\big)'' + \cdots + \big(\varphi(t+\beta_{m-1})\mathrm{V}_{(t+\beta_{m-1})}\big)''\Big)\right| dt \\
&= \int_0^h \left|\frac{(t-h/2)^2}{2} - h^2/8\right| \left|\big(\varphi(t+\beta_0)\mathrm{V}_{(t+\beta_0)}\big)'' + \cdots + \big(\varphi(t+\beta_{m-1})\mathrm{V}_{(t+\beta_{m-1})}\big)''\right| dt \\
&\leq \int_0^h \left|\frac{(t-h/2)^2}{2} - h^2/8\right| \Big(\left|\big(\varphi(t+\beta_0)\mathrm{V}_{(t+\beta_0)}\big)''\right| + \cdots + \left|\big(\varphi(t+\beta_{m-1})\mathrm{V}_{(t+\beta_{m-1})}\big)''\right|\Big)\,dt \\
&\leq mK_2 \int_0^h \left|\frac{(t-h/2)^2}{2} - \frac{h^2}{8}\right| dt.
\end{aligned}$$

The function $\frac{(t-h/2)^2}{2} - \frac{h^2}{8}$ is a parabola opening upward that is zero at $t = 0$ and $t = h/2$. Thus, it is negative for $0 < t < h/2$ and positive elsewhere. Using this fact, we have

$$\begin{aligned}
\int_0^h \left|\frac{(t-h/2)^2}{2} - \frac{h^2}{8}\right| dt &\leq \int_0^{h/2} \left|\left(\frac{(t-h/2)^2}{2} - \frac{h^2}{8}\right)\right| dt \\
&\quad + \int_{h/2}^h \left|\left(\frac{(t-h/2)^2}{2} - \frac{h^2}{8}\right)\right| dt \\
&= \left[\left|\frac{(t-h/2)^3}{6} - \frac{h^2 t}{8}\right|\right]_0^{h/2} + \left[\left|\frac{(t-h/2)^3}{6} - \frac{h^2 t}{8}\right|\right]_{h/2}^h \\
&= \frac{h^3}{24} + \frac{h^3}{24} = \frac{h^3}{12}.
\end{aligned}$$

Putting this all together and using $h = \frac{b-a}{m}$ gives us the following error bound:

$$|E_T| \leq \frac{mK_2h^3}{12} = \frac{K_2(b-a)^3}{12m^2}. \tag{A.3}$$

In the case when the second derivative of VaR is not bounded and instead, we have $\mid (\varphi(\beta)\mathrm{V}_\beta)' \mid \leq K_1$ for $\beta \in [a,b]$, one can employ an argument similar to that used above in arriving at (A.3). In particular, starting with equation (A.1), and constant of integration $A = h/2$, we obtain

$$|E_T| \leq mK_1 \int_0^h |t - h/2| \, dt.$$

The integral of function $|t - h/2|$ is $h^2/4$. Putting it all together, and using $h = \frac{b-a}{m}$ leads to the following error bound:

$$|E_T| \leq \frac{mK_1h^2}{4} = \frac{K_1(b-a)^2}{4m}.$$

$\square$

APPENDIX B

PROOFS FOR CVAR ESTIMATION

B.1 Proof of Corollary 2.5.1

Proof. First, we prove the claim in part (i). Notice that

$$\begin{aligned}
\mathbb{P}[\mid \mathrm{C}_\alpha - \widehat{\mathrm{C}}_{n,m,\alpha} \mid > \epsilon] &= \mathbb{P}\left[\left|\frac{1}{1-\alpha}\int_\alpha^1 \mathrm{V}_\beta\, d\beta - \frac{1}{1-\alpha}\sum_{k=1}^m \frac{\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \epsilon\right] \\
&= \mathbb{P}\left[\left|\int_\alpha^1 \mathrm{V}_\beta\, d\beta - \sum_{k=1}^m \frac{\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \epsilon(1-\alpha)\right] \\
&= \mathbb{P}\left[\left|\int_\alpha^1 \mathrm{V}_\beta\, d\beta - \sum_{k=1}^m \frac{\mathrm{V}_{\beta_{k-1}} + \mathrm{V}_{\beta_k}}{2}\Delta\beta\right.\right. \\
&\qquad + \sum_{k=1}^m \frac{\mathrm{V}_{\beta_{k-1}} + \mathrm{V}_{\beta_k}}{2}\Delta\beta \\
&\qquad \left.\left. - \sum_{k=1}^m \frac{\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \epsilon(1-\alpha)\right] \\
&\leq \mathbb{P}\left[\left|\sum_{k=1}^m \frac{\mathrm{V}_{\beta_{k-1}} + \mathrm{V}_{\beta_k}}{2}\Delta\beta\right.\right. \\
&\qquad \left.\left. - \sum_{k=1}^m \frac{\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \frac{\epsilon(1-\alpha)}{2}\right], \qquad \text{(B.1)}
\end{aligned}$$

where the final inequality follows by using Lemma 2.4.1(i) to infer that for $m \geq \frac{K_1(1-\alpha)}{2\epsilon}$, we have $\left|\int_\alpha^1 \mathrm{V}_\beta\, d\beta - \sum_{k=1}^m \frac{\mathrm{V}_{\beta_{k-1}} + \mathrm{V}_{\beta_k}}{2}\Delta\beta\right| < \frac{\epsilon(1-\alpha)}{2}$. Now, we have

$$\begin{aligned}
\mathbb{P}[\mid \mathrm{C}_\alpha - \widehat{\mathrm{C}}_{n,m,\alpha} \mid > \epsilon] &\leq \mathbb{P}\left[\left|\sum_{k=1}^m \frac{\mathrm{V}_{\beta_{k-1}} + \mathrm{V}_{\beta_k}}{2}\Delta\beta\right.\right. \\
&\qquad \left.\left. - \sum_{k=1}^m \frac{\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \widehat{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right| > \frac{\epsilon(1-\alpha)}{2}\right] \\
&= \mathbb{P}\left[\left|\sum_{k=1}^m ((\mathrm{V}_{\beta_{k-1}} + \mathrm{V}_{\beta_k}) - (\widehat{\mathrm{V}}_{n,\beta_{k-1}} + \widehat{\mathrm{V}}_{n,\beta_k}))\right| > \frac{\epsilon(1-\alpha)}{\Delta\beta}\right] \\
&= \mathbb{P}\left[\left|\mathrm{V}_{\beta_0} - \widehat{\mathrm{V}}_{n,\beta_0} + 2(\mathrm{V}_{\beta_1} - \widehat{\mathrm{V}}_{n,\beta_1})\right.\right.
\end{aligned}$$

$$
\begin{aligned}
&+\cdots+2(\mathrm{V}_{\beta_{m-1}}-\widehat{\mathrm{V}}_{n,\beta_{m-1}})+\ \mathrm{V}_{\beta_m}-\widehat{\mathrm{V}}_{n,\beta_m}\Big|>\frac{\epsilon(1-\alpha)}{\Delta\beta}\Bigg]\\
\le\ & \mathbb{P}\left[\left|\mathrm{V}_{\beta_0}-\widehat{\mathrm{V}}_{n,\beta_0}\right|>\frac{\epsilon(1-\alpha)}{2m\,\Delta\beta}\right]\\
&+\ 2\mathbb{P}\left[\left|\mathrm{V}_{\beta_1}-\widehat{\mathrm{V}}_{n,\beta_1}\right|>\frac{\epsilon(1-\alpha)}{2m\,\Delta\beta}\right]\\
&+\cdots+2\mathbb{P}\left[\left|\mathrm{V}_{\beta_{m-1}}-\widehat{\mathrm{V}}_{n,\beta_{m-1}}\right|>\frac{\epsilon(1-\alpha)}{2m\,\Delta\beta}\right]\\
&+\ \mathbb{P}\left[\left|\mathrm{V}_{\beta_m}-\widehat{\mathrm{V}}_{n,\beta_m}\right|>\frac{\epsilon(1-\alpha)}{2m\,\Delta\beta}\right]\\
\le\ & 2\exp\left(-2nc_0\left(\frac{\epsilon(1-\alpha)}{2\,m\Delta\beta}\right)^2\right)\\
&+\ 4\exp\left(-2nc_1\left(\frac{\epsilon(1-\alpha)}{2\,m\Delta\beta}\right)^2\right)\\
&+\cdots+4\exp\left(-2nc_{m-1}\left(\frac{\epsilon(1-\alpha)}{2\,m\Delta\beta}\right)^2\right)\\
&+2\exp\left(-2nc_m\left(\frac{\epsilon(1-\alpha)}{2\,m\Delta\beta}\right)^2\right)\\
\le\ & 4m\exp\left(-2nc\left(\frac{\epsilon(1-\alpha)}{2\,m\Delta\beta}\right)^2\right) \qquad \text{(B.2)}\\
=\ & 4m\exp\left(-\frac{n\,c\,\epsilon^2}{2}\right)\\
=\ & 2K_1(1-\alpha)/\epsilon\,.\exp\left(-n\,c\,\epsilon^2/2\right).
\end{aligned}
$$

where we have applied Lemma 2.4.2 to arrive at the inequality in (B.2), with $c=\min\{c_0,c_1,\ldots,c_m\}$. The claim follows for the case when the first derivative of VaR is bounded.

The proof of the result in part (ii) follows in a similar manner. In particular, using part (ii) in Lemma 2.4.1, with $m\ge\sqrt{\frac{K_2(1-\alpha)^2}{6\epsilon}}$, we obtain

$$
\begin{aligned}
\mathbb{P}\left[\left|\mathrm{C}-\widehat{\mathrm{C}}_{n,m,\alpha}\right|>\epsilon\right] &\le 4m\exp\left(-\frac{n\,c\,\epsilon^2}{2}\right)\\
&=\sqrt{\frac{8K_2(1-\alpha)^2}{3\epsilon}}\,.\exp\left(-\frac{n\,c\,\epsilon^2}{2}\right)
\end{aligned}
$$

□

B.2 Proof of Corollary 2.5.2

Proof. Recall that the truncation threshold $B_n = \sqrt{2\sigma^2 \log(n)}$. Letting $\eta = F(B_n)$, we have

$$
\begin{aligned}
\mathbb{P}\left[\mathrm{C} - \widetilde{\mathrm{C}}_{n,m,\alpha} > \epsilon\right] &\leq \mathbb{P}\left[\frac{1}{1-\alpha}\int_0^1 \mathrm{V}_\beta \, d\beta - \frac{1}{1-\alpha}\sum_{k=1}^{m} \frac{\widetilde{\mathrm{V}}_{n,\beta_{k-1}} + \widetilde{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta > \epsilon\right] \\
&= \mathbb{P}\left[\frac{1}{1-\alpha}\int_0^\eta \mathrm{V}_\beta \, d\beta - \frac{1}{1-\alpha}\sum_{k=1}^{m} \frac{\widetilde{\mathrm{V}}_{n,\beta_{k-1}} + \widetilde{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta\right. \\
&\quad \left. + \frac{1}{1-\alpha}\int_\eta^1 \mathrm{V}_\beta \, d\beta > \epsilon\right] \\
&= \mathbb{P}\left[I_1 + I_2 > \epsilon\right], \qquad \text{(B.3)}
\end{aligned}
$$

where $I_1 = \frac{1}{1-\alpha}\int_0^\eta \mathrm{V}_\beta \, d\beta - \frac{1}{1-\alpha}\sum_{k=1}^{m} \frac{\widetilde{\mathrm{V}}_{n,\beta_{k-1}} + \widetilde{\mathrm{V}}_{n,\beta_k}}{2}\Delta\beta$, and $I_2 = \frac{1}{1-\alpha}\int_\eta^1 \mathrm{V}_\beta \, d\beta$. We bound I_2 as follows:

$$
1 - \beta = \mathbb{P}\left(X > \mathrm{V}_\beta\right) \leq \exp\left(-\frac{{\mathrm{V}_\beta}^2}{2\sigma^2}\right), \qquad \text{(B.4)}
$$

since X is Gaussian with mean zero, and variance σ^2. Using $\log x \leq \frac{x}{e}\ \forall x > 0$, we obtain

$$
\mathrm{V}_\beta \leq \sqrt{2\sigma^2 \log\left(\frac{1}{1-\beta}\right)} \leq \sqrt{\frac{2\sigma^2}{e(1-\beta)}},
$$

leading to

$$
\begin{aligned}
\int_\eta^1 \mathrm{V}_\beta \, d\beta &\leq \sqrt{\frac{2\sigma^2}{e}}\int_\eta^1 \frac{d\beta}{\sqrt{1-\beta}} = 2\sqrt{\frac{2\sigma^2}{e}}\sqrt{1-\eta} \\
&\leq 2\sqrt{\frac{2\sigma^2}{e}}\exp\left(-\frac{{\mathrm{V}_\eta}^2}{4\sigma^2}\right) \qquad \text{(using (B.4))} \\
&= 2\sqrt{\frac{2\sigma^2}{e}}\exp\left(-\frac{{B_n}^2}{4\sigma^2}\right) \qquad \text{(since } \mathrm{V}_\eta = B_n\text{)}
\end{aligned}
$$

Hence,

$$
I_2 = \frac{1}{1-\alpha}\int_\eta^1 \mathrm{V}_\beta \, d\beta \leq \frac{2\sigma}{(1-\alpha)\sqrt{n}}. \qquad \text{(B.5)}
$$

Applying the bound in the Theorem 2.5.1 to the truncated r.v. $Z = X\mathbb{I}\{X \leq B_n\}$, we bound I_1 as follows:

$$\mathbb{P}[I_1 > \epsilon] \leq \frac{K_1(1-\alpha)}{\epsilon} \exp\left(\frac{-n\,c\,\epsilon^2}{2}\right). \tag{B.6}$$

Hence,

$$\begin{aligned}
\mathbb{P}[I_1 + I_2 > \epsilon] &\leq \frac{K_1(1-\alpha)}{\left(\epsilon - \frac{2\sigma}{(1-\alpha)\sqrt{n}}\right)} \exp\left(-\frac{nc\left(\epsilon - \frac{2\sigma}{(1-\alpha)\sqrt{n}}\right)^2}{2}\right) \\
&\qquad \text{(using (B.5) and (B.6))} \\
&\leq \frac{\sqrt{2\pi}\sigma n(1-\alpha)}{\left(\epsilon - \frac{2\sigma}{(1-\alpha)\sqrt{n}}\right)} \exp\left[-\frac{nc\left[\epsilon - \frac{2\sigma}{(1-\alpha)\sqrt{n}}\right]^2}{2}\right],
\end{aligned}$$

where the final inequality follows from the fact that $\delta_1 = \sqrt{2\pi}\sigma \exp\left(\frac{B_n^2}{2\sigma^2}\right) = \sqrt{2\pi}\sigma n$, which holds since the underlying Gaussian distribution is truncated at B_n.

By using a parallel argument, a concentration result for bounding the lower semi-deviations can be derived, and we omit the details. □

B.3 Proof of Corollary 2.5.3

Proof. The proof for the exponential case follows in a similar manner as that of the proof of Theorem 2.5.2. In particular, the proof up to (B.6) holds for the exponential case, with a different bound on I_2.

We derive the bound on $I_2 = \frac{1}{1-\alpha}\int_\eta^1 \mathrm{V}_\beta\, d\beta$. Using arguments similar to that in the Gaussian case, we obtain

$$\begin{aligned}
\int_\eta^1 \mathrm{V}_\beta\, d\beta &\leq \frac{1}{\lambda}\int_\eta^1 \log\left(\frac{1}{1-\beta}\right) d\beta \\
&= \frac{(1-\eta)}{\lambda}\left(1 + \log\left(\frac{1}{1-\eta}\right)\right) \\
&\leq \frac{(1-\eta)}{\lambda}\left(1 + \frac{1}{(1-\eta)e}\right) \\
&= \frac{\exp(-\lambda \mathrm{V}_\eta)}{\lambda}\left(1 + \exp(\lambda \mathrm{V}_\eta - 1)\right)
\end{aligned}$$

$$= \frac{\exp\left(-\lambda B_n\right)}{\lambda} + \frac{1}{\lambda e}. \qquad \text{(since } \mathrm{V}_\eta = B_n\text{)}$$

Choosing $B_n = \frac{\log(n)}{\lambda}$, we have

$$I_2 = \frac{1}{1-\alpha}\int_\eta^1 \mathrm{V}_\beta \, d\beta \leq \frac{(n+1)}{\lambda n(1-\alpha)} \tag{B.7}$$

Now, as in the proof of Theorem 2.5.2, we have

$$\mathbb{P}\left[I_1 > \epsilon\right] \leq \frac{K_1(1-\alpha)}{\epsilon} \exp\left(-\frac{nc\epsilon^2}{2}\right). \tag{B.8}$$

Thus,

$$\begin{aligned}
\mathbb{P}\left[I_1 + I_2 > \epsilon\right] &\leq \frac{K_1(1-\alpha)}{\left(\epsilon - \frac{(n+1)}{\lambda n(1-\alpha)}\right)} \exp\left[-\frac{nc\left(\epsilon - \frac{(n+1)}{\lambda n(1-\alpha)}\right)^2}{2}\right] \quad \text{(using (B.7) and (B.8))}\\
&\leq \frac{n(1-\alpha)}{\left(\epsilon - \frac{(n+1)}{\lambda n(1-\alpha)}\right)} \exp\left[-\frac{nc\left(\epsilon - \frac{(n+1)}{\lambda n(1-\alpha)}\right)^2}{2}\right].
\end{aligned}$$

where the final inequality follows from the fact that $\delta_1 = \exp(\lambda B_n) = n$, which holds since the underlying exponential distribution is truncated at B_n.

By using a parallel argument, a concentration result for bounding the lower semi-deviations can be derived, and we omit the details. $\square$

REFERENCES

1. **Acerbi, C.** (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, **26**(7), 1505–1518.

2. **Acerbi, C.** and **D. Tasche** (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, **26**(7), 1487–1503.

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

4. **Audibert, J. Y.**, **S. Bubeck**, and **R. Munos**, Best arm identification in multi-armed bandits. *In Conference on Learning Theory*. 2010.

5. **Behrisch, M.**, **L. Bieker**, **J. Erdmann**, and **D. Krajzewicz**, SUMO–Simulation of Urban MObility. *In The Third International Conference on Advances in System Simulation (SIMUL 2011)*, volume 42. 2011.

6. **Bhat, S. P.** and **L. A. Prashanth**, Concentration of risk measures: A Wasserstein distance approach. *In Advances in Neural Information Processing Systems 32*. 2019.

7. **Brown, D. B.** (2007). Large deviations bounds for estimating conditional value-at-risk. *Operations Research Letters*, **35**(6), 722–730.

8. **Bruni, R.**, **F. Cesarone**, **A. Scozzari**, and **F. Tardella** (2016). Real-world datasets for portfolio selection and solutions of some stochastic dominance portfolio models. *Data in brief*, **8**, 858–862.

9. **Cruz-Uribe, D.** and **C. J. Neugebauer** (2003). An elementary proof of error estimates for the trapezoidal rule. *Mathematics magazine*, **76**(4), 303–306.

10. **Dowd, K.**, *Measuring market risk*. John Wiley & Sons, 2005.

11. **Dowd, K.** and **D. Blake** (2006). After VaR: the theory, estimation, and insurance applications of quantile-based risk measures. *Journal of Risk and Insurance*, **73**(2), 193–229.

12. **Dufour, J. M.** (1995). Distribution and quantile functions. Technical report, McGill University, Montreal, Canada.

13. **Haklay, M.** and **P. Weber** (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, **7**(4), 12–18.

14. **Kolla, R. K.**, **L. A. Prashanth**, **S. P. Bhat**, and **K. P. Jagannathan** (2019). Concentration bounds for empirical conditional value-at-risk: The unbounded case. *Operations Research Letters*, **47**(1), 16–20.

15. **Markowitz, H.** (1952). Portfolio selection. *The journal of finance*, **7**(1), 77–91.

16. **Nski, A. R.** (2010). Risk-averse dynamic programming for markov decision process. *Math. Program*, **125**, 235–261.

17. **Prashanth, L. A.**, **K. Jagannathan**, and **R. K. Kolla** (2019). Concentration bounds for cvar estimation: The cases of light-tailed and heavy-tailed distributions. *arXiv preprint arXiv:1901.00997*.

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

19. **Serfling, R. J.**, *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.

20. **Shen, Y.**, **W. Stannat**, and **K. Obermayer** (2013). Risk-sensitive markov control processes. *SIAM Journal on Control and Optimization*, **51**(5), 3652–3672.

21. **Slivkins, A.** *et al.* (2019). Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, **12**(1-2), 1–286.

22. **Thomas, P.** and **E. Learned-Miller**, Concentration inequalities for conditional value at risk. *In International Conference on Machine Learning*. 2019.

23. **Tversky, A.** and **D. Kahneman** (1979). An analysis of decision under risk. *Econometrica*, **47**, 263–292.

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

25. **Wainwright, M. J.**, *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

26. **Wang, Y.** and **F. Gao** (2010). Deviation inequalities for an estimator of the conditional value-at-risk. *Operations Research Letters*, **38**(3), 236–239.

27. **Wegener, A.** *et al.*, TraCI: an interface for coupling road traffic and network simulators. *In Proceedings of the 11th communications and networking simulation symposium*. ACM, 2008.

LIST OF PAPERS BASED ON THESIS

1. A. K. Pandey, L. A. Prashanth and S. P. Bhat, *Estimation of Spectral Risk Measures*, accepted in AAAI Conference on Artificial Intelligence 2021.