Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

AN UPPER CONFIDENCE BOUND APPROACH TO ESTIMATING COHERENT RISK MEASURES

Guangwu Liu

College of Business
City University of Hong Kong
83 Tat Chee Avenue, Kowloon, Hong Kong

Wen Shi

Business School
Central South University
Hunan, China 410083

Kun Zhang

Institute of Statistics and Big Data
Renmin University of China
Beijing, China 100872

ABSTRACT

Coherent risk measures have received increasing attention in recent years among both researchers and practitioners. The problem of estimating a coherent risk measure can be cast as estimating the maximum expected loss taken under a set of probability measures. In this paper, we consider the set of probability measures is finite, and study the estimation of a coherent risk measure via an upper confidence bound (UCB) approach, where samples of the portfolio loss are simulated sequentially from one of the probability measures. We study in depth the so-called *Grand Average* estimator, and establish statistical guarantees, including its strong consistency, asymptotic normality, and asymptotic mean squared error. We also construct asymptotically valid confidence intervals.

1 INTRODUCTION

Risk measures find a wide range of important applications in financial industries, such as calculation of regulatory capital charge, risk pricing, and portfolio optimization. Among various popular risk measures, Value-at-Risk (VaR) has been widely used, which is defined as the quantile of the probability distribution of the loss at a given confidence level. Despite its popularity in the banking industry, VaR has been criticized because it does not take into account the magnitude of extreme losses and may discourage risk diversification. As a remedy for these drawbacks, a class of risk measures, referred to as coherent risk measures, have been proposed by Artzner, Delbaen, Eber, and Heath (1999). A notable example of coherent risk measures is conditional Value-at-Risk (CVaR, also known as expected shortfall), defined as the average size of the losses beyond the VaR value. In the recent Basel Accords (Basel III & IV), the Basel Committee on Banking Supervision has moved away from VaR towards CVaR in its market risk framework (Basel Committee on Banking Supervision 2016). Another application of coherent risk measure is on the construction of good deal bounds when pricing derivative securities, where bid and ask prices of a derivative security can be represented as coherent risk measures; see, e.g., Jaschke and Küchler (2001) and Staum (2004).

978-1-7281-3283-9/19/$31.00 ©2019 IEEE

It has been well known that under mild continuity conditions, a coherent risk measure ρ can be represented in the form of

$$\rho(Y) = \sup_{P \in \mathscr{P}} \mathbb{E}_P[-Y/r],$$

where Y is the value of the portfolio at a future time, $1/r$ is the discount factor, and $\mathscr{P}$ denotes a set of probability distributions. To further simplify the problem, we consider the setting in Lesnevski, Nelson, and Staum (2007), where the set $\mathscr{P}$ has only a finite number K of elements $P_1, \ldots, P_K$. This assumption holds, for instance, when the coherent risk measure is defined by the specification of K generalized scenarios. Under this setting, the problem of estimating the coherent risk measure is reduced to the estimation of the maximum mean of K stochastic systems, which is closely related to identifying the best system (Kim and Nelson 2006), but arguably more difficult (Lesnevski, Nelson, and Staum 2007). The main difficulty lies in that maximum-mean estimation requires sampling budget allocated to the best system as much as possible, not just find the best system out. The problem of estimating the maximum mean also finds a variety of applications in other areas of management science and machine learning, ranging from Markov decision processes (MDPs) and reinforcement learning to Monte Carlo tree search; see, e.g., Chang, Fu, Hu, and Marcus (2005), Kocsis and Szepesvári (2006) and Fu (2017).

This paper focuses on how to estimate the maximum mean in the context of coherent risk measures. In this context, one of the key issues is to develop efficient estimators with sound statistical guarantees, especially their consistency, asymptotic normality and asymptotic mean squared errors, which seem to be missing in the literature, to the best of our knowledge. To this end, we study estimators of the maximum mean under the upper confidence bound (UCB) sampling policy framework. The UCB policy (Auer et al. 2002) has been studied extensively in the literature of the stochastic bandit problem where a sample from a system is called a (random) reward and the objective of the decision maker is to maximize the total rewards. At the heart of the stochastic bandit problem is thus how to accumulate higher rewards while sequentially learning which system is the best, referred to as the tradeoff between exploration and exploitation. It has been well known that the UCB policy serves as an effective way of balancing this tradeoff. In the UCB policy for bandit problems, the exploration rate is set to be a logarithm function, ensuring that the expected amount of sampling budget allocated to the non-maximum systems is at most of a logarithmic order of the total budget, which implies that majority of the sampling budget is allocated to the system with the maximum mean. In our study, we consider a generalized UCB policy that allows the exploration rate to take a range of functional forms, and aim to construct efficient estimators of the maximum mean with desirable statistical guarantees.

The estimator being studied in this paper is the sample average of all the samples, whether drawn from the system with the maximum mean or not. We refer to it as the *Grand Average (GA)* estimator. Rationale of the GA estimator stems from the fact that under the UCB framework, majority of the samples are drawn from the system with the maximum mean, implying that the grand average is dominated by the sample average of this system. The GA estimator is not new in the literature, and has served as a key ingredient for algorithms in MDPs and reinforcement learning; see, e.g., Chang, Fu, Hu, and Marcus (2005) and Kocsis and Szepesvári (2006). However, to the best of our knowledge, very little is known about the statistical properties of the GA estimator, except its asymptotic unbiasedness. In this paper, we fill this gap by establishing statistical guarantees for the GA estimator, including its strong consistency, central limit theorem (CLT), and asymptotic rate of mean squared error (MSE), leading to both an efficient point estimator and asymptotically valid confidence intervals for the coherent risk measure.

The rest of the paper is organized as follows. We propose the GA estimator under the UCB framework in Section 2. Asymptotic properties of the GA estimator are established in Section 3, as well as a way of constructing asymptotically valid confidence intervals. We demonstrate the performance of the point and interval estimates in Section 4, followed by conclusions in Section 5. Lengthy proofs are provided in the appendix.

2 AN UPPER CONFIDENCE BOUND APPROACH

We formulate the problem of estimating coherent risk measures using simplified notations. Consider K stochastic systems, with performances denoted by random variables $\{X_k, k = 1,\ldots,K\}$, respectively. For each k, we assume that X_k follows an unknown probability distribution, while independent samples can be drawn. Here, X_k represents the discounted loss of a portfolio sampled from probability distribution P_k.

Let $\mu_k = \mathbb{E}(X_k)$. We are interested in estimating the maximum mean, defined as

$$\mu^* \triangleq \max_{k=1,\ldots,K} \mu_k.$$

Throughout the paper, we assume that X_k's have bounded supports. While this assumption simplifies the analysis and helps to convey the main idea in a clearer manner, our results can be extended to the setting when X_k's follow sub-Gaussian distributions.

Let k^* denote the index of the system with the largest mean, i.e., $\mu^* = \mu_{k^*}$. Without loss of generality, k^* is assumed to be unique. Given samples of X_k, denoted by $\{X_{kj}, j = 1,\ldots,n_k\}$, $k = 1,\ldots,K$, a straightforward estimator of μ^*, referred to as the *maximum estimator*, is given by

$$\max_{k=1,\ldots,K} \bar{X}_k,$$

where $\bar{X}_k \triangleq \sum_{j=1}^{n_k} X_{kj}/n_k$ is the sample average of X_k.

It has been well known that the maximum estimator has a positive bias and may overestimate μ^*. It may lead to overestimation of risk, resulting in unduly high capital charges for risky activities (Lesnevski, Nelson, and Staum 2007). Better estimators of μ^* are thus desirable.

Sampling from the K systems is essential for constructing any estimator of μ^*. In developing an efficient estimator, two issues are of major concern. The first issue is on how to efficiently allocate the sampling budget, while the second issue is on methods of constructing an estimator such that statistical guarantees can be established.

To address the first issue, dynamic sampling policies have been studied in the literature on MDPs and machine learning. A popular dynamic sampling policy is the UCB policy that was originally proposed for the multi-armed bandit (MAB) problem; see Auer, Cesa-Bianchi, and Fischer (2002). In traditional MAB problem, in each round, one of the K systems is chosen and a random sample (reward) is drawn from the chosen system, and the decision maker aims to maximize the total rewards collected over the first n rounds. At the heart of the MAB problem is to find a dynamic sampling policy that decides which system to sample from in each round so as to balance the tradeoff between exploration and exploitation. Among various sampling policies for MAB, it has been well known that the UCB policy achieves the optimal rate of regret, defined as the expected loss due to the fact that a policy does not always choose the system with highest expected reward; see Lai and Robbins (1985) and Auer, Cesa-Bianchi, and Fischer (2002). Specifically, let $T_k(t)$ denote the number of samples drawn from system k during the first t rounds, for $k = 1,\ldots,K$. The UCB sampling policy can be described as in Algorithm 1.

The UCB policy offers important insights into the sequential allocation of sampling budget. Essentially, it balances the tradeoff between exploration and exploitation using the UCB defined on the right-hand-side (RHS) of (1). On the one hand, systems with higher on-going sample averages have higher chance to be chosen in the current round, contributing to higher total rewards. On the other hand, systems that are chosen less frequently during the previous rounds, i.e., $T_k(t-1)$ being smaller, may also have sufficient chance to be chosen so that such systems will be sufficiently explored. The function $\log t$ in (1) can be interpreted as the exploration rate that controls the speed of exploring systems that may not have the largest mean.

In the MAB context, the exploration rate is set to be $\log t$ simply because it leads to the optimal rate of regret. However, it should be pointed out that when the objective is not to miminize the regret, as the case in our setting, the rate function may take different forms. Therefore, we shall henceforth allow the exploration rate, denoted by ν_t, to take different forms as a function of t, and refer to the resulting UCB policy as a *generalized UCB (GUCB) policy*, which is described in Algorithm 2.

Algorithm 1: UCB Sampling Policy

1. **Initialization**: During the first K rounds, draw a sample from each system.
2. **Repeat**: For $t \geq K+1$, draw a sample from the system indexed by

$$\kappa_t = \arg\max_{k\in\{1,\ldots,K\}} \left\{ \bar{X}_k[T_k(t-1)] + \sqrt{\frac{2\log t}{T_k(t-1)}} \right\}, \tag{1}$$

where $\bar{X}_k[T_k(t-1)]$ denote the sample mean of system k during the first $(t-1)$ rounds. Update the sample average of system κ_t.

Algorithm 2 (GUCB): Generalized UCB Sampling Policy

1. **Initialization**: During the first K rounds, draw a sample from each system.
2. **Repeat**: For $t \geq K+1$, draw a sample from the system indexed by

$$I_t = \arg\max_{k\in\{1,\ldots,K\}} \left\{ \bar{X}_k[T_k(t-1)] + \sqrt{\frac{2\nu_t}{T_k(t-1)}} \right\},$$

where $\bar{X}_k[T_k(t-1)]$ denote the sample mean of system k during the first $(t-1)$ rounds. Update the sample average of system I_t.

2.1 Grand Average (GA) Estimator

Before moving on to the construction of estimators for μ^*, we highlight some of the key properties of the GUCB policy. To convey the main idea, for a while we focus on a special case where the exploration rate $\nu_t = \log t$. In this case, it has been known that (see Auer, Cesa-Bianchi, and Fischer (2002)), for system k $(k \neq k^*)$,

$$\mathbb{E}[T_k(n)] \leq C\log n,$$

for some constant C that depends on the gap between μ_k and μ^*. Then it can be easily seen that $\mathbb{E}[T_{k^*}(n)]$, the expected number of times system k^* is chosen, is of order n during the first n rounds, because the summation of $T_k(n)$'s equals n.

In other words, it is expected that among the first n rounds, system k^* is chosen for a majority of the rounds. Recall that our objective is to estimate μ^*, the mean of system k^*. It is, therefore, reasonable to use the grand average of the samples of all the n rounds as an estimator of μ^*, which we refer to as the *Grand Average (GA)* estimator. Although it takes into account the samples that are not drawn from system k^*, the validity of the GA estimator can be justified by the fact that the estimation error due to such samples may phase out when n is sufficiently large, as the number of such samples is negligible compared to those drawn from system k^*.

Specifically, the GA estimator is defined by

$$\widetilde{M}_n = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{T_k(n)} X_{kj}, \tag{2}$$

where $\{X_{kj}, j = 1,\ldots,T_k(n)\}$ denotes the samples drawn from system k during the first n rounds.

The GA estimator is not new, which has served as a key ingredient for algorithms for MDPs and reinforcement learning under the special case that $\nu_t = \log t$. However, research on the statistical properties of the estimator has been underdeveloped. To the best of our knowledge, only asymptotic unbiasedness of the estimator has been established (Kocsis and Szepesvári 2006), while other statistical properties seem to be missing. One of the contributions of this paper is to fill this gap. In particular, we generalize the GA estimator by allowing ν_t to take a range of functional forms, and establish its strong consistency, CLT, and asymptotic MSE, which shall be discussed in detail in the following section.

3 ASYMPTOTIC PROPERTIES

In this section, we establish strong consistency, asymptotic MSE and asymptotic normality for the GA estimator. To facilitate analysis, we first establish a proposition on the moments of $T_k(n)$. Due to page limit, the proof of the proposition is omitted.

Proposition 1 If $\nu_n \geq \log n$, then for $k \neq k^*$ and any positive integer p,

$$\mathbb{E}T_k(n)^p \leq \begin{cases} \left\lceil \dfrac{8\nu_n}{\Delta_k^2} \right\rceil + 4 & if\ p = 1; \\ \left(\left\lceil \dfrac{8\nu_n}{\Delta_k^2} \right\rceil + 2\left[\log(n+1)+2\right]^{\frac{1}{2}} \right)^2 & if\ p = 2; \\ \left\{ \left\lceil \dfrac{8\nu_n}{\Delta_k^2} \right\rceil + \left\lceil \dfrac{2p}{p-2}(n+1)^{p-2} + O\left(n^{p-3}\right) \right\rceil^{\frac{1}{p}} \right\}^p & if\ p \geq 3, \end{cases}$$

where $\Delta_k = \mu^* - \mu_k$, and the notation $O(\cdot)$ means that $\limsup_{n\to\infty} a_n/b_n \leq C$ for some constant C if $a_n = O(b_n)$.

Proposition 1 provides an upper bound for the pth moment of $T_k(n)$, the number of samples drawn from system k during the first n rounds, for $k \neq k^*$. This upper bound relies on the exploration rate ν_n. In a special case when $\nu_n = \log n$ and $p = 1$, the result is the same as that in Theorem 1 of Auer, Cesa-Bianchi, and Fischer (2002).

Proposition 1 implies that the sampling ratios $T_k(n)/n$ may satisfy certain convergence properties. In particular, strong consistency of the sampling ratios are summarized in the following theorem, whose proof is provided in Section A.1 of the appendix.

Theorem 1 If $\nu_n \in \left[\log n, n^{1-\delta}\right]$ with $0 < \delta < 1$, then, as $n \to \infty$,

$$\frac{T_k(n)}{n} \xrightarrow{a.s.} \begin{cases} 0 & for\ k \neq k^*; \\ 1 & for\ k = k^*. \end{cases}$$

where the notation $\xrightarrow{a.s.}$ denotes convergence almost surely (or with probability 1).

Let $X_{I_j,j}$ denote the sample drawn at the jth round. Note that

$$\widetilde{M}_n = \frac{1}{n}\sum_{k=1}^{K} T_k(n)\bar{X}_k[T_k(n)] = \frac{1}{n}\sum_{j=1}^{n} X_{I_j,j} = \frac{1}{n}\sum_{j=1}^{n}\left(X_{I_j,j} - \mu_{I_j}\right) + \frac{1}{n}\sum_{j=1}^{n}\mu_{I_j}.$$

Define

$$Z_n \triangleq \sum_{j=1}^{n}\left(X_{I_j,j} - \mu_{I_j}\right).$$

Then,

$$\widetilde{M}_n = \frac{Z_n}{n} + \frac{1}{n}\sum_{j=1}^{n}\mu_{I_j}. \tag{3}$$

Let $\mathscr{F}_n$ be the σ-field generated by the first n samples for $n \geq 1$, and $\mathscr{F}_0 = \{\Omega, \emptyset\}$. Note that I_n is $\mathscr{F}_{n-1}$-measurable. It follows that for $n \geq 1$,

$$\mathbb{E}\left[X_{I_n,n} | \mathscr{F}_{n-1}\right] = \mu_{I_n}, \text{ and then } \mathbb{E}X_{I_n,n} = \mathbb{E}\mu_{I_n}.$$

Therefore, $\mathbb{E}Z_n = 0$ and

$$\mathbb{E}\left[Z_n | \mathscr{F}_{n-1}\right] = Z_{n-1} + \mathbb{E}\left[X_{I_n,n} - \mu_{I_n} | \mathscr{F}_{n-1}\right] = Z_{n-1},$$

i.e., Z_n is a martingale with mean 0.

Because X_{kj} is assumed to have bounded support, we have

$$|Z_{n+1} - Z_n| = \left|X_{I_{n+1},n+1} - \mu_{I_{n+1}}\right| \leq C.$$

for some constant C. By Azuma-Hoeffding Inequality (Durrett 2019) and letting $Z_0 = 0$,

$$\mathbb{P}\left(|Z_n/n| \geq \varepsilon\right) = \mathbb{P}\left(|Z_n| \geq n\varepsilon\right) \leq 2\exp\left\{-n\varepsilon^2/(2C^2)\right\},$$

and hence $\sum_{n=1}^{\infty} \mathbb{P}\left(|Z_n/n| > \varepsilon\right) < \infty$. By Borel-Cantelli lemma, $Z_n/n \xrightarrow{a.s.} 0$ as $n \to \infty$. i.e., the first term on the RHS of (3) converges to 0 almost surely. Moreover, the second term on the RHS of (3) satisfies

$$\frac{1}{n}\sum_{j=1}^{n} \mu_{I_j} = \frac{1}{n}\sum_{k=1}^{K} T_k(n)\mu_k \xrightarrow{a.s.} \mu_{k^*},$$

where the convergence follows from Theorem 1 and the continuous mapping theorem (Durrett 2019). Therefore, by (3), we establish strong consistency of $\widetilde{M}_n$ that is summarized as follows.

Theorem 2 If $\nu_n \in \left[\log n, n^{1-\delta}\right]$ with $0 < \delta < 1$, then $\widetilde{M}_n$ is a strongly consistent estimator of μ^*, i.e.,

$$\widetilde{M}_n \xrightarrow{a.s.} \mu^*, \quad \text{as } n \to \infty.$$

Theorem 2 ensures that the GA estimator, $\widetilde{M}_n$, converges to μ^* with probability 1, as the sample size goes to infinity. To further understand its asymptotic properties, it is desirable to conduct error analysis, especially on its bias and variance. In the following theorem, we establish the rate of convergence of its asymptotic MSE. Proof of the theorem is provided in Section A.2 of the appendix.

Theorem 3 If $\nu_n \in \left[\log n, n^{1/2-\delta}\right]$ with $0 < \delta < 1/2$, then,

$$\text{Bias}\left(\widetilde{M}_{I_n^*}\right) = O\left(\frac{\nu_n}{n}\right), \quad \text{Var}\left[\widetilde{M}_n\right] \leq 2K\frac{\sigma_{k^*}^2}{n} + o\left(\frac{1}{n}\right),$$

and thus

$$\text{MSE}\left(\widetilde{M}_n\right) \leq 2K\frac{\sigma_{k^*}^2}{n} + o\left(\frac{1}{n}\right),$$

where $\sigma_k^2 \triangleq \text{Var}[X_k]$, and the notation $o(\cdot)$ means that $\lim_{n\to} a_n/b_n = 0$ if $a_n = o(b_n)$.

Theorem 3 shows that the MSE of the GA estimator is of order n^{-1}. In greater detail, variance is the dominant term in the MSE, compared to square of the bias, which is of order ν_n^2/n^2. This result implies that when the exploration rate takes value in the range $\nu_n \in \left[\log n, n^{1/2-\delta}\right]$ with $0 < \delta < 1/2$, the bias of the GA estimator is negligible, compared to its variance.

To provide a theoretical support for asymptotically valid CIs, we establish a CLT for $\widetilde{M}_n$ in the following theorem, whose proof is provided in Section A.3 of the appendix.

Theorem 4 If $\nu_n \in \left[\log n, n^{\frac{1}{2}-\delta}\right]$ with $0 < \delta < 1/2$, then,

$$\sqrt{n}\left(\widetilde{M}_n - \mu^*\right) \Rightarrow N(0, \sigma^2_{k^*}), \quad as \quad n \to \infty,$$

where $\Rightarrow$ denotes convergence in distribution.

Theorem 4 shows that the GA estimator is asymptotically normally distributed with mean μ^* and variance $\sigma^2_{k^*}/n$. Based on Theorem 4, an asymptotically valid CI can be constructed for the maximum mean μ^*. To do so, a remaining issue is on how to estimate the unknown $\sigma^2_{k^*}$. In light of the proof of Theorem 4, we estimate $\sigma^2_{k^*}$ using

$$\widetilde{\sigma}^2_n = \frac{1}{n}\sum_{k=1}^{K} T_k(n)\hat{\sigma}^2_k, \tag{4}$$

where for $k = 1, ..., K$,

$$\hat{\sigma}^2_k = \frac{1}{T_k(n)}\sum_{j=1}^{T_k(n)} X^2_{k,j} - \left[\frac{1}{T_k(n)}\sum_{j=1}^{T_k(n)} X_{k,j}\right]^2. \tag{5}$$

It can be shown that $\widetilde{\sigma}^2_n$ converges to σ^2_k in probability, as $n \to \infty$. This result is summarized in the following proposition, whose proof is omitted due to page limit.

Proposition 2 If $\nu_n \in \left[\log n, n^{1-\delta}\right]$ with $0 < \delta < 1$, then, as $n \to \infty$,

$$\widetilde{\sigma}^2_n \to \sigma^2_{k^*}, \quad \text{in probability.}$$

From Theorem 4 and Proposition 2, it follows that $\sqrt{n}\widetilde{\sigma}^{-1}_n\left(\widetilde{M}_n - \mu^*\right) \Rightarrow N(0,1)$. Then, an asymptotically valid $100(1-\beta)\%$ CI of μ^* is given by

$$(\widetilde{M}_n - z_{1-\beta/2}\widetilde{\sigma}_n/\sqrt{n}, \widetilde{M}_n + z_{1-\beta/2}\widetilde{\sigma}_n/\sqrt{n}), \tag{6}$$

where $z_{1-\beta/2}$ is the $1-\beta/2$ quantile of the standard normal distribution.

4 A NUMERICAL EXAMPLE

We examine the performances of the GA estimator and the constructed CI through an example. Consider the option portfolio example in Section 2.2 of Lesnevski, Nelson, and Staum (2007), where the risk of the portfolio is assessed via a coherent risk measure based on $K = 4^4 = 256$ generalized scenarios. These scenarios are set by varying the risk factors (i.e., underlying asset prices) under several different conditions such as a large increase, a large decrease, and moderate changes. We follow exactly the same parameter settings as in Lesnevski, Nelson, and Staum (2007), to which interested readers are referred for details.

While the GA estimator is asymptotically unbiased, it is low-biased with finite samples and the bias is more severe especially when the variances of the systems are large. To alleviate this adverse effect of bias, we set up a warm-up period and discard samples in this period during the implementation. We observed that MSEs of the GA estimator are robust with respect to the length of the warm-up period, while it mainly affects the validity of the CIs. The MSEs are also robust with respect to the specification of exploration rate function ν_t, and the results reported in this section are based on $\nu_t = \log^2 t$.

To examine the performance of the GA estimator, we estimate its relative bias, standard deviation (std), and square root of MSE (RMSE), as percentages compared to true value of the coherent risk measure that is 4645 in this example. The estimated error metrics reported are based on 1000 independent replications.

Table 1: Relative bias (%), relative standard deviation (%) and RMSE (%), and coverage probabilities (%).

sample size n ($\times 10^5$)	1	2	4	6	8	10
bias	-0.60	-0.24	-0.05	-0.02	-0.007	0.001
std	1.61	1.13	0.60	0.36	0.31	0.19
RMSE	1.72	1.15	0.60	0.36	0.31	0.19
cov. prob.	79.6	83.9	88.9	90.1	89.7	90.8

Table 2: Comparison of CIs.

Fixed width of CI (Lesnevski et al.)	100	90	80	70	60	50	40	30
cov. prob. (Lesnevski et al.)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.99
cov. prob. (GA)	0.90	0.90	0.88	0.90	0.89	0.91	0.90	0.89
Ratio of CI widths	4.5	4.0	3.6	3.1	2.7	2.2	1.8	1.3

We also report the coverage probability (cov. prob.) of the constructed 90% CIs. Performances of the estimators with respect to different sample sizes are summarized in Table 1. From the table it can be seen that when the sample size is 10^5, the RMSE is below 2% of the true value, suggesting a very accurate estimate. It can also be seen that the bias of the GA estimator is negligible compared to its variance that contributes to the major part of its MSE. This coincides with the theoretical result on asymptotic MSE as in Theorem 3. Moreover, when the sample size becomes larger, the coverage probability converges to the nominal one (90%), implying that the constructed CIs are asymptotically valid.

In the second set of experiments, we compare the constructed CIs to the fixed-width CI procedure as proposed in Lesnevski, Nelson, and Staum (2007). Since a fixed width of the CI has to be set at the beginning of the procedure of Lesnevski, Nelson, and Staum (2007) and thus the total sample size required to stop the process may vary across different replications. For the sake of fairness in comparison, we first run the procedure of Lesnevski, Nelson, and Staum (2007), and the same sample size is then used to construct our CIs. We then compared the widths of the 90% CIs to that of the fixed width. The comparison results with respect to varying settings of the fixed width are presented in Table 2. From the table it can be seen that the procedure of Lesnevski, Nelson, and Staum (2007) tends to be conservative in that it leads to coverage probabilities that are close to 100%, while our method produces narrower CIs. For instance, when the fixed width is set to be 100 (about 2% of the true value), the width of CIs of Lesnevski, Nelson, and Staum (2007) is more than four times wider than ours.

5 CONCLUSIONS

In this paper, we have studied a Grand Average estimator of a coherent risk measure under the UCB framework, and established statistical guarantees for the estimator, including its strong consistency, asymptotic normality, and asymptotic rate of MSE. We have shown that the rate of convergence of the estimator is $\sqrt{n}$, where n is the sample size. We have also constructed an asymptotically valid confidence interval. Both the point estimator and the confidence interval perform very well numerically.

ACKNOWLEDGMENTS

The work of the second author was partially supported by the Major Project for National Natural Science Foundation of China under Grant No. 71790615 and the National Natural Science Foundation of China under Grant Nos. 71402048, and 71671060.

A APPENDIX

A.1 Proof of Theorem 1

When $k \neq k^*$, by Markov's inequality and Proposition 1, for any $\varepsilon > 0$ and $p \geq 3$,

$$\begin{aligned}\mathbb{P}\left(\frac{T_k(n)}{n} > \varepsilon\right) \leq \frac{1}{\varepsilon^p}\mathbb{E}\left[\frac{T_k(n)}{n}\right]^p &\leq \frac{1}{\varepsilon^p}\left\{\left\lceil\frac{8\nu_n}{\Delta_k^2}\right\rceil\frac{1}{n} + \left[\frac{2p}{p-2}\frac{1}{n^2} + O\left(n^{-3}\right)\right]^{\frac{1}{p}}\right\}^p \\ &\leq \frac{2^{p-1}}{\varepsilon^p}\left[\left(\left\lceil\frac{8\nu_n}{\Delta_k^2}\right\rceil\frac{1}{n}\right)^p + \frac{2p}{p-2}\frac{1}{n^2} + O\left(n^{-3}\right)\right].\end{aligned}$$

Because $\log n \leq \nu_n \leq n^{1-\delta}$, it follows that for $p \geq 3$,

$$\mathbb{P}\left(\frac{T_k(n)}{n} > \varepsilon\right) \leq \frac{2^{p-1}}{\varepsilon^p}\left[\left(\frac{8}{\Delta_k^2 n^\delta} + \frac{1}{n}\right)^p + \frac{2p}{p-2}\frac{1}{n^2} + O\left(n^{-3}\right)\right] \leq \frac{2^{p-1}}{\varepsilon^p}\left[2^{p-1}\left(\frac{8}{\Delta_k^2 n^{p\delta}} + \frac{1}{n^p}\right) + \frac{2p}{p-2}\frac{1}{n^2} + O\left(n^{-3}\right)\right]. \quad (7)$$

Let $p > \max\{1/\delta, 3\}$. Then, for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\frac{T_k(n)}{n} > \varepsilon\right) < \infty.$$

By Borel-Cantelli lemma, $T_k(n)/n \xrightarrow{a.s.} 0$ as $n \to \infty$, for $k \neq k^*$, and thus

$$\frac{T_{k^*}(n)}{n} = 1 - \sum_{k \neq k^*}\frac{T_k(n)}{n} \xrightarrow{a.s.} 1.$$

A.2 Proof of Theorem 3

We analyze the bias and variance of $\widetilde{M}_n$ separately.

The Bias. We assert that $T_k(n)$ is a stopping time. In fact, given the filtration generated by $\{X_{ij}, i \neq k, j = 1, 2, \dots\}$, denoted by $\sigma(X_{ij}, i \neq k, j = 1, 2, \dots)$, one has

$$\{T_k(n) \geq m\} = \{T_k(n) \leq m-1\}^c \subset \sigma(X_{k1}, \dots, X_{k,m-1}).$$

Hence $T_k(n)$ is a stopping time with respect to $\{X_{kj}, j \geq 1\}$. Then by Wald's equation (Ross 1996), $\mathbb{E}\left[\sum_{j=1}^{T_k(n)} X_{kj}\right] = \mu_k \mathbb{E}[T_k(n)]$, leading to

$$\mathbb{E}\widetilde{M}_n = \frac{1}{n}\sum_{k=1}^{K}\mu_k \mathbb{E}[T_k(n)].$$

Thus,

$$\begin{aligned}\text{Bias}\left(\widetilde{M}_n\right) &= \mathbb{E}\widetilde{M}_n - \mu^* = \left(-1 + \frac{\mathbb{E}T_{k^*}(n)}{n}\right)\mu_{k^*} + \sum_{k \neq k^*}\frac{\mathbb{E}T_k(n)}{n}\mu_k \\ &= -\sum_{k \neq k^*}\frac{\mathbb{E}T_k(n)}{n}\mu_{k^*} + \sum_{k \neq k^*}\frac{\mathbb{E}T_k(n)}{n}\mu_k = \sum_{k \neq k^*}\frac{\mathbb{E}T_k(n)}{n}(\mu_k - \mu_{k^*}).\end{aligned}$$

By Proposition 1,

$$\text{Bias}\left(\widetilde{M}_n\right) \leq \sum_{k \neq k^*}\left(\frac{8\nu_n}{n\Delta_k^2} + \frac{5}{n}\right)(\mu_{k^*} - \mu_k) = O\left(\frac{\nu_n}{n}\right). \quad (8)$$

The Variance. Note that

$$\begin{aligned}
&\mathrm{Var}\left[\widetilde{M}_n\right] = \mathbb{E}\left(\widetilde{M}_n - \mathbb{E}\left[\widetilde{M}_n\right]\right)^2 \\
= \;& \mathbb{E}\left\{\left(\widetilde{M}_n - \frac{1}{n}\sum_{k=1}^{K}\mu_k T_k(n)\right) + \left(\frac{1}{n}\sum_{k=1}^{K}\mu_k T_k(n) - \frac{1}{n}\sum_{k=1}^{K}\mu_k \mathbb{E}T_k(n)\right)\right\}^2 \\
\leq \;& 2\mathbb{E}\left(\widetilde{M}_n - \frac{1}{n}\sum_{k=1}^{K}\mu_k T_k(n)\right)^2 + 2\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{K}\mu_k T_k(n) - \frac{1}{n}\sum_{k=1}^{K}\mu_k \mathbb{E}T_k(n)\right)^2. \qquad (9)
\end{aligned}$$

We analyze the two terms on the RHS of (9) separately. By the Wald's Lemma (see Theorem 13.2.14 in Athreya and Lahiri (2006)),

$$\mathbb{E}\left(\sum_{j=1}^{T_k(n)} X_{kj} - \mu_k T_k(n)\right)^2 = \sigma_k^2 \mathbb{E}T_k(n). \qquad (10)$$

Then, by (3) and Cauchy-Schwarz inequality, it can be seen that

$$\begin{aligned}
\mathbb{E}\left(\widetilde{M}_n - \frac{1}{n}\sum_{k=1}^{K}\mu_k T_k(n)\right)^2 &\leq \frac{K}{n^2}\sum_{k=1}^{K}\mathbb{E}\left(\sum_{j=1}^{T_k(n)} X_{kj} - \mu_k T_k(n)\right)^2 = \frac{K}{n^2}\sum_{k=1}^{K}\sigma_k^2 \mathbb{E}T_k(n) \\
&= \frac{K}{n^2}\sigma_{k^*}^2 \mathbb{E}T_{k^*}(n) + \frac{K}{n^2}\sum_{k\neq k^*}\mathbb{E}T_k(n) \leq \frac{K}{n}\sigma_{k^*}^2 + \frac{K}{n^2}\sum_{k\neq k^*}\mathbb{E}T_k(n), \qquad (11)
\end{aligned}$$

where the last inequality follows from the fact that $T_{k^*}(n) \leq n$.

We then analyze the second term on the RHS of (9). Note that, by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{K}\mu_k T_k(n) - \frac{1}{n}\sum_{k=1}^{K}\mu_k \mathbb{E}T_k(n)\right)^2 = \frac{1}{n^2}\mathbb{E}\left(\sum_{k=1}^{K}\mu_k T_k(n) - \sum_{k=1}^{K}\mu_k \mathbb{E}T_k(n)\right)^2 \\
\leq \;& \frac{K}{n^2}\sum_{k=1}^{K}\mu_k^2 \mathbb{E}\left[T_k(n) - \mathbb{E}T_k(n)\right]^2 = \frac{K}{n^2}\sum_{k=1}^{K}\mu_k^2 \mathrm{Var}\left[T_k(n)\right] \\
\leq \;& \frac{K}{n^2}\sum_{k\neq k^*}\mu_k^2 \mathrm{Var}\left[T_k(n)\right] + \frac{K}{n^2}\mu_{k^*}^2 \mathrm{Var}\left[T_{k^*}(n)\right]. \qquad (12)
\end{aligned}$$

Furthermore,

$$\mathrm{Var}\left[T_{k^*}(n)\right] = \mathrm{Var}\left[n - \sum_{k\neq k^*} T_k(n)\right] = \mathrm{Var}\left[\sum_{k\neq k^*} T_k(n)\right] \leq \mathbb{E}\left[\sum_{k\neq k^*} T_k(n)\right]^2 \leq K\sum_{k\neq k^*}\mathbb{E}T_k^2(n).$$

Therefore,

$$\text{RHS of (12)} \leq \frac{K}{n^2}\sum_{k\neq k^*}\mu_k^2 \mathbb{E}T_k^2(n) + \frac{K^2}{n^2}\mu_{k^*}^2\sum_{k\neq k^*}\mathbb{E}T_k^2(n) \leq \frac{K}{n^2}\sum_{k\neq k^*}\left(\mu_k^2 + K\mu_{k^*}^2\right)\mathbb{E}T_k^2(n). \qquad (13)$$

Combining (9), (11)-(13) and Proposition 1 yields

$$\mathrm{Var}\left[\widetilde{M}_n\right] \leq \frac{2K}{n}\sigma_{k^*}^2 + \frac{2K}{n^2}\sum_{k\neq k^*}\left(\frac{8\nu_n}{\Delta_k^2} + 5\right) + 2K^2\bar{\mu}^2\sum_{k\neq k^*}\left(\frac{8\nu_n}{n\Delta_k^2} + \frac{1}{n} + \frac{2\left[\log(n+1)+2\right]^{\frac{1}{2}}}{n}\right)^2, \qquad (14)$$

where $\bar{\mu}^2 = \max\{\mu_1^2, ..., \mu_K^2\}$.

Incorporating the bias in (8), the variance in (14), and $v_n \in \left[\log n, n^{1/2-\delta}\right]$ with $0 < \delta < 1/2$, we have

$$\text{MSE}\left(\widetilde{M}_n\right) = \text{Var}\left[\widetilde{M}_n\right] + \text{Bias}\left(\widetilde{M}_n\right)^2 \leq \frac{2K}{n}\sigma_{k^*}^2 + o\left(\frac{1}{n}\right).$$

A.3 Proof of Theorem 4

To prove the result, we first define a martingale difference array and introduce a lemma on asymptotic normality for martingale different arrays (Theorem 16.1.1 of Athreya and Lahiri (2006)).

Definition 1 Let $\{Y_n\}_{n\geq 1}$ be a collection of random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\{\mathscr{F}_n\}_{n\geq 1}$ be a filtration. Then, $\{Y_n, \mathscr{F}_n\}_{n\geq 1}$ is called a martingale difference array if Y_n is $\mathscr{F}_n$-measurable and $\mathbb{E}[Y_n|\mathscr{F}_{n-1}] = 0$ for each $n \geq 1$.

Lemma 1 For each $n \geq 1$, let $\{Y_{ni}, \mathscr{F}_{ni}\}_{i\geq 1}$ be a martingale difference array on $(\Omega, \mathscr{F}, \mathbb{P})$, with $\mathbb{E}|Y_{ni}|^2 < \infty$ for all $n \geq 1$ and let τ_n be a finite stopping time w.r.t. $\mathscr{F}_{ni}$. Suppose that for some constant $\sigma^2 > 0$, as $n \to \infty$, $\sum_{i=1}^{\tau_n} \mathbb{E}\left[|Y_{ni}|^2 \middle| \mathscr{F}_{ni}\right] \to \sigma^2$ in probability, and that for any $\varepsilon > 0$, $\sum_{i=1}^{\tau_n} \mathbb{E}\left[|Y_{ni}|^2 1_{\{|Y_{ni}|>\varepsilon\}} \middle| \mathscr{F}_{ni}\right] \to 0$ in probability. Then,

$$\sum_{i=1}^{\tau_n} Y_{ni} \Rightarrow N(0, \sigma^2), \text{ as } n \to \infty.$$

Proof of Theorem 4. By (3),

$$\sqrt{n}\left(\widetilde{M}_n - \mu^*\right) = \frac{Z_n}{\sqrt{n}} + \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} \mu_{I_j} - \mu^*\right). \tag{15}$$

We analyze the two terms on the RHS of (15) separately. Note that Z_n is a martingale. So $Z_n - Z_{n-1} = X_{I_n,n} - \mu_{I_n}$ is a martingale difference array. To show the asymptotic normality of $Z_n/\sqrt{n}$, it suffices to prove that the two conditions of Lemma 1 are satisfied for $Y_{nj} = \left(X_{I_j,j} - \mu_{I_j}\right)/\sqrt{n}$, $\tau_n = n$ and $\mathscr{F}_{nj} = \mathscr{F}_j$. First, note that

$$\begin{aligned}
&\sum_{j=1}^{n} \mathbb{E}\left[\left|\left(X_{I_j,j} - \mu_{I_j}\right)/\sqrt{n}\right|^2 \middle| \mathscr{F}_{j-1}\right] = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\left[X_{I_j,j}^2 - 2\mu_{I_j}X_{I_j,j} + \mu_{I_j}^2 \middle| \mathscr{F}_{j-1}\right] \\
= \ & \frac{1}{n}\sum_{j=1}^{n}\left\{\mathbb{E}\left[X_{I_j,j}^2 \middle| \mathscr{F}_{j-1}\right] - 2\mu_{I_j}\mathbb{E}\left[X_{I_j,j} \middle| \mathscr{F}_{j-1}\right] + \mu_{I_j}^2\right\} = \frac{1}{n}\sum_{j=1}^{n}\left\{\mathbb{E}\left[X_{I_j,j}^2 \middle| \mathscr{F}_{j-1}\right] - \mu_{I_j}^2\right\} \\
= \ & \frac{1}{n}\sum_{j=1}^{n}\sigma_{I_j}^2 = \frac{1}{n}\sum_{k=1}^{K} T_k(n)\sigma_k^2 \xrightarrow{a.s.} \sigma_{k^*}^2, \text{ as } n \to \infty,
\end{aligned}$$

where the convergence follows from Theorem 1.

Second,

$$\begin{aligned}
&\sum_{j=1}^{n} \mathbb{E}\left[\left|\left(X_{I_j,j} - \mu_{I_j}\right)/\sqrt{n}\right|^2 1_{\left\{\left|\left(X_{I_j,j}-\mu_{I_j}\right)/\sqrt{n}\right|>\varepsilon\right\}} \middle| \mathscr{F}_{j-1}\right] \\
= \ & \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\left[\left|X_{I_j,j} - \mu_{I_j}\right|^2 1_{\left\{\left|X_{I_j,j}-\mu_{I_j}\right|>\varepsilon\sqrt{n}\right\}} \middle| \mathscr{F}_{j-1}\right] \leq \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\left[C \cdot 1_{\left\{\left|X_{I_j,j}-\mu_{I_j}\right|>\varepsilon\sqrt{n}\right\}} \middle| \mathscr{F}_{j-1}\right] \\
= \ & C \cdot \frac{1}{n}\sum_{j=1}^{n} \mathbb{P}\left(\left|X_{I_j,j} - \mu_{I_j}\right| > \varepsilon\sqrt{n} \middle| \mathscr{F}_{j-1}\right) \leq C \cdot \frac{1}{n}\sum_{j=1}^{n} \frac{\mathbb{E}\left[\left|X_{I_j,j} - \mu_{I_j}\right|^2 \middle| \mathscr{F}_{j-1}\right]}{\varepsilon^2 n} \\
= \ & \frac{C}{\varepsilon^2 n} \cdot \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\left[\left|X_{I_j,j} - \mu_{I_j}\right|^2 \middle| \mathscr{F}_{j-1}\right] = \frac{C}{\varepsilon^2 n} \cdot \frac{1}{n}\sum_{j=1}^{n} \sigma_{I_j}^2 = \frac{C}{\varepsilon^2 n} \cdot \frac{1}{n}\sum_{k=1}^{K} T_k(n)\sigma_k^2 \xrightarrow{a.s.} 0, \text{ as } n \to \infty,
\end{aligned}$$

where C is an upper bound of $\left|X_{I_j,j} - \mu_{I_j}\right|^2$ and exists because X_k's are assumed to have bounded supports. Therefore, by Lemma 1, we have

$$Z_n/\sqrt{n} \Rightarrow N(0, \sigma^2_{k^*}), \quad as \quad n \to \infty.$$

It remains to prove that the second term on the RHS of (15) converges to 0 in probability. Note that

$$\begin{aligned}\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\mu_{I_j} - \mu_{k^*}\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{K}T_k(n)\mu_k - \mu_{k^*}\right) = \sqrt{n}\left[\sum_{k\neq k^*}\frac{T_k(n)}{n}\mu_k + \left(\frac{T_{k^*}(n)}{n} - 1\right)\mu_{k^*}\right] \\ &= \sqrt{n}\left[\sum_{k\neq k^*}\frac{T_k(n)}{n}\mu_k - \sum_{k\neq k^*}\frac{T_k(n)}{n}\mu_{k^*}\right] = \sum_{k\neq k^*}\frac{T_k(n)}{\sqrt{n}}(\mu_k - \mu_{k^*}) \xrightarrow{L^1} 0,\end{aligned}$$

where the convergence follows from Proposition 1 and the condition $\nu_n \in \left[\log n, n^{\frac{1}{2}-\delta}\right]$ with $0 < \delta < 1/2$. Applying Slutsky's Theorem (Durrett 2019) to the RHS of (15) leads to the conclusion.

REFERENCES

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. "Coherent Measures of Risk". *Mathematical Finance* 9:203–228.

Athreya, K. B., and S. N. Lahiri. 2006. *Measure Theory and Probability Theory*. Springer Science & Business Media.

Auer, P., N. Cesa-Bianchi, and P. Fischer. 2002. "Finite-Time Analysis of the Multiarmed Bandit Problem". *Machine Learning* 47:235–256.

Basel Committee on Banking Supervision 2016. *Standards, Minimum Capital Requirements for Market Risk*. http://www.bis.org.

Chang, H. S., M. C. Fu, J. Hu, and S. I. Marcus. 2005. "An Adaptive Sampling Algorithm for Solving Markov Decision Processes". *Operations Research* 53:126–139.

Durrett, R. 2019. *Probability: Theory and Examples*. New York: Cambridge University Press.

Fu, M. C. 2017. "An Adaptive Sampling Algorithm for Solving Markov Decision Processes". *INFORMS TutORials in Operations Research*:68–88.

Jaschke, S., and U. Küchler. 2001. "Coherent Risk Measures and Good-Deal Bounds". *Finance and Stochastics* 5:181–200.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". In *Handbooks in OR & MS: Simulation*, edited by S. G. Henderson and B. L. Nelson. New York: Elsevier Science.

Kocsis, L., and C. Szepesvári. 2006. "Bandit Based Monte Carlo Planning". *Machine Learning: ECML*:282–293.

Lai, T. L., and H. Robbins. 1985. "Asymptotically Efficient Adaptive Allocation Rules". *Advances in Applied Mathematics* 6:4–22.

Lesnevski, V., B. L. Nelson, and J. Staum. 2007. "Simulation of Coherent Risk Measures Based on Generalized Scenarios". *Management Science* 53:1756–1769.

Ross, S. M. 1996. *Stochastic Processes*. 2nd ed. New York: John Wiley.

Staum, J. 2004. "Fundamental Theorems of Asset Pricing for Good Deal Bounds". *Mathematical Finance* 14:141–161.

AUTHOR BIOGRAPHIES

GUANGWU LIU is a Professor in the Department of Management Sciences, College of Business at City University of Hong Kong. He holds a PhD in Industrial Engineering and Logistics Management from The Hong Kong University of Science and Technology. His research interests include stochastic simulation, financial engineering, risk management, and machine learning. He currently serves as an Associate Editor for Naval Research Logistics. His email address is msgw.liu@cityu.edu.hk.

WEN SHI is a Professor in Business School at Central South University. He holds a PhD in management science and engineering from Huazhong University of Science and Technology. His research focuses on simulation modeling, simulation experiment design and analysis, and supply chain management. His email address is shi3wen@163.com.

KUN ZHANG is an Assistant Professor in the Institute of Statistics and Big Data at Renmin University of China. He holds a PhD in management science from City University of Hong Kong. His research interests include financial engineering and risk management. His email address is kkunzhang2013@gmail.com.