Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

A SEQUENTIAL ELIMINATION APPROACH TO VALUE-AT-RISK AND CONDITIONAL VALUE-AT-RISK SELECTION

Adam J. Hepworth
Michael P. Atkinson
Roberto Szechtman

Department of Operations Research
Naval Postgraduate School
1 University Circle
Monterey, CA 93943, USA

ABSTRACT

Conditional Value-at-Risk (CVaR) is a widely used metric of risk in portfolio analysis, interpreted as the expected loss when the loss is larger than a threshold defined by a quantile. This work is motivated by situations where the CVaR is given, and the objective is to find the portfolio with the largest or smallest quantile that meets the CVaR constraint. We define our problem within the classic stochastic multi-armed bandit (MAB) framework, and present two algorithms. One that can be used to find the portfolio with largest or smallest loss threshold that satisfies the CVaR constraint with high probability, and another that determines the portfolio with largest or smallest probability of exceeding a loss threshold implied by a CVaR constraint, also at some desired probability level.

1 INTRODUCTION

The MAB is a classic iterative decision learning problem in which the learner (agent) seeks to make an optimal selection of a single arm from an initial S arms of the bandit available for selection. For every iteration n the agent obtains a reward from arm s, $X_{s,n}$, where the aim is to select the optimal arm with high probability. In the stochastic form of the problem, an underlying distribution that is not initially known forms the basis for the rewards obtained at each iteration. Whilst the unobserved distribution for each arm does not change throughout the problem, the reward observed, being random, does.

As one of the most extensively studied problems in decision theory and reinforcement learning, the MAB has seen a wide range of applications in contemporary settings such as in medical research (Gittins 1989) and financial portfolio optimization (Shen, Wang, Jiang, and Zha 2015). An in-depth background into this class of problems is given in the work of (Even-Dar, Mannor, and Mansour 2002) and (Bubeck and Cesa-Bianchi 2012), well defining the range of MAB problems and providing the foundational context and derivations for subsequent research. The successive elimination algorithms presented in (Even-Dar, Mannor, and Mansour 2002) and (Even-Dar, Mannor, and Mansour 2006) broadly underpins the novel work presented herein. In the last two papers the goal is to find the arm with largest expected reward, and as such the MAB setting is related to the problem ranking and selection in stochastic simluation; see (Kim and Nelson 2006) for an overview.

This work is motivated by situations where the maximum expected loss over a threshold is given, and the agent wishes to discover the arm with largest or smallest threshold that satisfies the constraint, or the arm with largest or smallest probability of exceeding the threshold that meets the constraint. Such problems arise naturally in portfolio risk analysis, where the loss threshold is known as Value-at-Risk, and the expected conditional loss over the worst 100α percent scenarios is known as Conditional Value-at-Risk

978-1-5386-3428-8/17/$31.00

at level α; see, for example, (Rockafellar and Uryasev 2000), (Rockafellar and Uryasev 2002), and (Brown 2007).

Essentially, our work involves bounding the error probability in a root estimation context, and as such is related to MAB problems dealing with quantiles. The quantile-based learning setting given in (Szörényi, Busa-Fekete, Weng, and Hüllermeier 2015) is formalized within the Probably Approximately Correct (PAC) framework and is shown in the Qualitative PAC (`QPAC`) algorithm. `QPAC` uses qualitative ordinal data, that would otherwise not be successfully analyzed within the traditional framework. The `QPAC` algorithm is an adaptive elimination strategy and was further improved by the work of (David and Shimkin 2016) through their `MaxQ` and `Double-MaxQ` algorithms. The quantity of interest for these algorithms is modified to select the arm with the greatest α-quantile value.

We study the problem faced by an agent who has a constraint on the CVaR for a set of candidate portfolios or activities, meaning that

$$E[X_s|X_s > k_{s,\alpha_s}] = C, \tag{1}$$

where X_s is a random loss associated with a portfolio or activity s, k_{s,α_s} is the VaR, and $C \in R$ is the input constraint. The CVaR can be viewed as the expected loss for the fraction α of worst scenarios. The constraint C can be interpreted as the amount of resources (be it money, people allocated to some task, parts damaged, etc) that the agent can afford to lose or allocate in the worst or best 100α percent cases, depending on the situation. The solution of the root problem also is called the *buffered Probability of Exceedance* (bPOE) (Uryasev 2014), defined as the inverse of a CVaR level and is a generalization of the so-called *buffered Probability of Failure* (bPOF), as defined in (Rockafellar and Royset 2010).

We consider two situations. First, the agent wishes to find the portfolio with largest or smallest threshold quantile (known as the Value-at-Risk at level α), represented by the root k_{s,α_s} in (1). Second, the agent's objective is to identify the portfolio with largest or smallest probability α of exceeding the VaR, where the root k_{s,α_s} is set to satisfy (1). In the former case the level α_s plays no role in finding the VaR that satisfies the CVaR constraint, while in the latter α_s can be obtained from the root k_{s,α_s}.

Another potential application is online marketing, where each arm is an online marketing campaign for some product. The constant C (an input) is the average quality (e.g., a function of age, income, gender, etc) of the individuals wanted by the seller. $1-\alpha_s(C)$ is the fraction of people generated by the marketing campaign who have an average quality C, with quality at least $k_s(C)$. Hence, the retailer wishes to find the marketing campaign with lowest $\alpha_s(C)$. In this case a sample $X_{s,i}$ is the quality of individual i generated by marketing campaign s.

In this paper we present a new approach to find the arms with largest or smallest Value-at-Risk and CVaR level α under a loss constraint, within a PAC setting. From a technical standpoint, we extend the classical PAC framework, involving mean estimators, to a novel situation involving root estimators. The paper is organized as follows. Section 2 deals with the problem of finding the arm with largest VaR. The problem of finding the arm with the largest CVaR at level α is treated in Section 3. Appendices at the rear of the paper include proofs of the main results.

2 SELECTING THE LARGEST VALUE-AT-RISK

Without loss of generality, we work with the problem of finding the arm with the largest root k and the one with largest CVaR level α. The problem of finding the arm with the smallest root or CVaR level is handled by working with arms driven by the negative random variables $-X_s$.

More formally, consider a finite set of candidate arms $\mathscr{S} = \{1,\ldots,S\}$. For each arm $s \in \mathscr{S}$ there is an uncertain loss, defined by a random variable X_s with an unknown continuous distribution. The agent can draw independent and identically distributed (i.i.d.) random samples $X_{s,1}, X_{s,2}, \ldots$ from a distribution with a density $f_{X_s}(\cdot)$, and $k_s(C)$ is the root of

$$C = E[X_s|X_s > k].$$

The goal is to find the arm $s^* \in \mathscr{S}$ with the largest root, $k_{s^*}(C) = \max_s k_s(C)$.

The main assumptions we make are:

A1. $C - E[X_s] \geq \gamma > 0$, $\forall s \in \mathscr{S}$.
A2. The random variables $X_{s,1}, X_{s,2}, \ldots$, have bounded support over (a,b), $\forall s \in \mathscr{S}$, with $-\infty < a < C < b < \infty$.
A3. The random variables X_s, for all $s \in \mathscr{S}$, have a probability density function $f_{X_s}(\cdot)$ that is uniformly bounded below by $\zeta > 0$.

Assumptions A1 and A2 ensure that the root $k_s(C)$ is well defined, whilst Assumption A3 is used to bound the error probability of the root estimator. In quantile estimation settings (see Section 2.3.2 of (Serfling 2008)) a positive density is required in the neighborhood of the quantile to control the estimation error; in the CVaR setting this assumption is extended to the entire support. Following this, the arm index s is dropped unless it is needed.

The idea is to adapt the sequential elimination approach of (Even-Dar, Mannor, and Mansour 2002), for which one needs to show that for $0 < \delta < 1$ there exists $\varepsilon_n > 0$ such that

$$P(|k_n - k(C)| > \varepsilon_n) \leq \frac{6\delta}{\pi^2 n^2 S},$$

where k_n is the root estimator based off n i.i.d. samples. The analysis is simplified by dealing exclusively with the root of the function

$$g(k) = E[(X - C)I(X > k)],$$

the result of a simplification of

$$C = E[X|X > k] = \frac{E[X; X > k]}{P(X > k)} \iff g(k) = 0.$$

Assumptions A1 and A2 guarantee that $\lim_{k \to -\infty} g(k) = E[X] - C < 0$, and $g(\cdot)$ increases to reach its maximum at C, with $g(C) = E[(X - C)I(X > C)] > 0$. From there $g(k)$ decreases towards 0, as k approaches b. It follows that there is only one root $k(C) < C$ that solves $g(k) = 0$. One implication of Assumption A3 is that $C - k(C) \geq \psi > 0$ for some $\psi > 0$; see Lemma 5.1 in the Appendix for details. Intuitively, the error probability in the root estimation gets larger as C approaches $k(C)$. An illustration of the function $g(\cdot)$ is shown in Figure 1.

Drawing i.i.d. samples $X_1, \ldots, X_n$, the root is estimated by solving

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - C) I(X_i > k) = 0, \tag{2}$$

where the left-hand side of (2) can be interpreted as an empirical $g(\cdot)$ function. More formally, let the estimated root be given by

$$k_n = \inf \left\{ k \geq a : \frac{1}{n} \sum_{i=1}^{n} (X_i - C) I(X_i > k) \geq 0 \right\}. \tag{3}$$

There are three possibilities:

1. $(1/n)\sum_{i=1}^{n} X_i < C$ and $(1/n)\sum_{i=1}^{n} I(X_i > C) > 0$, in which case monotonicity of $(1/n)\sum_{i=1}^{n} (X_i - C) I(X_i > k)$ in k ensures that there is a unique root in (a, C).
2. $(1/n)\sum_{i=1}^{n} X_i \geq C$, leading to $k_n = a$.
3. $(1/n)\sum_{i=1}^{n} I(X_i > C) = 0$, in which case $k_n = \max_{i=1,\ldots,n} X_i \leq C$.

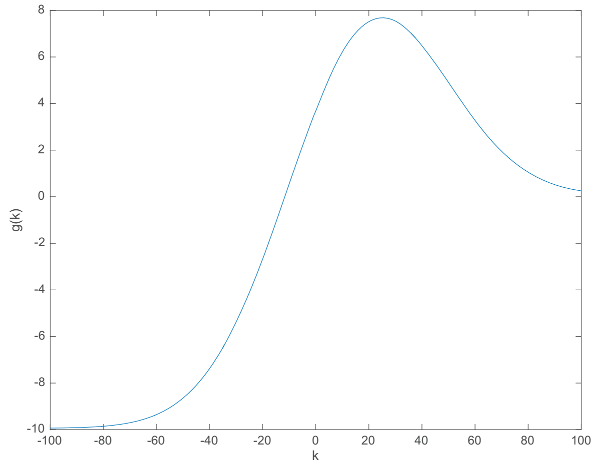

Figure 1: $g(\cdot)$ for a truncated Normal$(\mu = 15, \sigma = 30)$ over the interval $(-100, 100)$ with $C = 25$.

The assumptions imply that the probability of cases 2 or 3 decay to zero exponentially in n. In case 1, given the ordered samples $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, the root finding can be implemented as $k_n = X_{(m^*)}$, where

$$m^* = \min\left\{ m \geq 1 : \sum_{i=m}^{n} \left(X_{(i)} - C \right) \geq 0 \right\}. \tag{4}$$

The average complexity of sorting the samples and finding $X_{(m^*)}$ is of order $O(n \log n)$, as given in (Cormen, Leiserson, Rivset, and Stein 2010).

2.1 Algorithm

Let $\Delta_s = k_{s^*}(C) - k_s(C) > 0$ for all arms $s \neq s^*$. Algorithm 1, shown below, initializes each root $k_{s,n}$ to a, and uses thresholds

$$\varepsilon_n = \left(\log\left(\frac{\pi^2 n^2 S}{3\delta} \right) \frac{1}{2n} \right)^{1/2} \frac{b-a}{\zeta \psi}, \text{ for } n = 1, 2, \ldots, \tag{5}$$

to eliminate suboptimal arms. Theorem 1 shows that the root estimation error $|k_{s,n} - k_s(C)|$ is larger than ε_n with probability δ.

Theorem 1 Under Assumptions A1, A2, and A3,

$$P\left(|k_{s,n} - k_s(C)| \leq \varepsilon_n, \forall n, \forall s = 1, \ldots, S \right) \geq 1 - \delta.$$

Algorithm 1 is a standard implementation of the sequential elimination algorithm of (Even-Dar, Mannor, and Mansour 2002).

Algorithm 1 Sequential VaR Elimination Algorithm

```
Set 𝒜 = {1,...,S}.
Set k_{s,n} = a, ∀s ∈ 𝒜
while |𝒜| > 1 do
  for arm s ∈ 𝒜 do
    Draw one sample from arm s and compute k_{s,n}
    if k_{max,n} = max_{s'∈𝒜}{k_{s',n}} − k_{s,n} > 2ε_n then
      𝒜 = 𝒜 \ {s}
    end if
  end for
  Set n = n+1
end while
```

Since $P(|k_{s,n} - k_s(C)| \leq \varepsilon_n) \geq 1-\delta$, Algorithm 1 selects the best arm with probability at least $1-\delta$ ((Even-Dar, Mannor, and Mansour 2002)). As in (Glynn and Juneja 2015), the expected number of samples $E[N_s]$ generated by a suboptimal arm $s \neq s^*$ is

$$E[N_s] \leq \sum_{n=1}^{\infty} P(k_{\max,n} - k_{s,n} < 2\varepsilon_n) \leq \sum_{n=1}^{\infty} P(k_{s^*,n} - k_{s,n} < 2\varepsilon_n) \leq u_s + \sum_{n=u_s+1}^{\infty} P(k_{s^*,n} - k_{s,n} < 2\varepsilon_n)$$

for $u_s = \inf\{n : 4\varepsilon_n \leq \Delta_s\}$. It easily follows that $E[N_s] \leq u_s + 2\delta/S$, meaning that the expected number of samples over the suboptimal arms, $\sum_{s\neq s^*} E[N_s]$, is at most $2\delta + \sum_{s\neq s^*} u_s$. Solving for $n \geq e$ such that $4\varepsilon_n = \Delta_s$, with ε_n as in (5), leads to the dominant term in $\sum_{s\neq s^*} E[N_s]$ being

$$\frac{8(b-a)^2}{\zeta^2\psi^2} \log\left(\frac{\pi^2 S}{3\delta}\right) \sum_{s\neq s^*} \Delta_s^{-2},$$

for small δ; see (Glynn and Juneja 2015).

Hence, finding an arm with largest root changes the expected number of samples by a factor of $(\zeta\psi)^{-2}$ in relation to the case where the goal is to select the arm with largest mean, where the expected number of samples generated by a suboptimal arm is of order $8(b-a)^2 \log(\pi^2 S/(3\delta)) \sum_{s\neq s^*} \Delta_s^{-2}$.

3 SELECTING THE LARGEST CVAR LEVEL

In this section we get back to the problem of finding the arm with the largest CVaR level. More precisely, for $F_s(\cdot)$ the distribution function of X_s, let

$$\alpha_s(C) = F_s(k_s(C)),$$

where $k_s(C)$ is the root of

$$E[X_s | X_s > k] = C.$$

The goal of the agent is to find the arm s^* with largest $\alpha_s(C)$. Let $\alpha_{s,n}$ be the empirical estimator of $\alpha_s(C)$, given by

$$\alpha_{s,n} = \bar{F}(k_n) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq k_n), \tag{6}$$

where $\bar{F}(\cdot)$ is the empirical distribution function, and k_n is defined in (3). It follows from (4) that $\alpha_{s,n} = {}^{m^*}\!/_{n}$ when $({}^{1}\!/_{n})\sum_{i=1}^{n} X_i < C$ and $({}^{1}\!/_{n})\sum_{i=1}^{n} I(X_i > C) > 0$, so that computationally the problem is not any costlier than that of finding the root k_n.

With a view towards the elimination algorithm, define the thresholds

$$\varepsilon_n = \left(\log\left(\frac{\pi^2 n^2 S}{\delta}\right)\frac{2}{n}\right)^{1/2} \times \max\left\{\frac{b-a}{\psi^2\zeta}, \frac{2(b-a)+(b-C)/n}{\psi}, 1\right\}, \tag{7}$$

for $n = 1,2,\ldots$. The maximands in (7) stem from coupling the empirical distribution to the empirical $g(\cdot)$ function (cf., Eq. (2)); see the proof of Theorem 2. Intuitively, the difference between the true and empirical CVaR levels is large if at least one of three events occur: The root estimator is far from the true estimator, the empirical $g(\cdot)$ function at the true root $k(C)$ deviates too much from the true $g(\cdot)$ function, or if the empirical distribution is far from the true distribution at the root $k(C)$.

As in the last section, assume $\Delta_s = \alpha_{s^*}(C) - \alpha_s(C) > 0$ for all arms $s \neq s^*$. Algorithm 2 uses the thresholds in (7) to eliminate suboptimal arms. Theorem 2, whose proof appears in the Appendix, establishes the key condition for the algorithm to work.

Theorem 2 Under Assumptions A1, A2, and A3, for ε_n as in (7) we have

$$P\left(|\alpha_{s,n} - \alpha_s(C)| \leq \varepsilon_n, \forall n, \forall s = 1,\ldots,S\right) \geq 1-\delta.$$

The sequential elimination algorithm for CVaR level selection is detailed next.

Algorithm 2 Sequential CVaR Level Elimination Algorithm

```
Set 𝒜 = {1,...,S}.
Set α_{s,n} = 0, ∀s ∈ 𝒜
while |𝒜| > 1 do
  for arm s ∈ 𝒜 do
    Sample from arm s and compute α_{s,n}
    if max_{s'∈𝒜}{α_{s',n}} − α_{s,n} > 2ε_n then
      𝒜 = 𝒜 \ {s}
    end if
  end for
  Set n = n+1
end while
```

As with Algorithm 1, Theorem 2 implies that the arm with largest CVaR level α is selected with probability at least $1-\delta$.

Likewise, the total expected number of samples $\sum_{s\neq s^*} E[N_s]$ generated by the suboptimal arms is

$$\sum_{s\neq s^*} E[N_s] \leq 2\delta + \sum_{s\neq s^*} u_s$$

for $u_s = \inf\{n > e : 4\varepsilon_n \leq \Delta_s\}$. For small $\delta > 0$, a standard argument shows that the dominant term in $\sum_{s\neq s^*} E[N_s]$ is

$$32\left(\max\left\{\frac{b-a}{\psi^2\zeta}, \frac{2(b-a)}{\psi}, 1\right\}\right)^2 \log\left(\frac{\pi^2 S}{\delta}\right) \sum_{s\neq s^*} \Delta_s^{-2},$$

where the impact of the $(b-C)/n$ term showing in (7) is of order smaller than $\log(1/\delta)$.

4 CONCLUDING REMARKS

Conditional Value-at-Risk is a widely used metric of portfolio risk. Motivated by situations where the maximum expected loss over a threshold is given, in this paper we have demonstrated a novel approach to find the arms with largest or smallest Value-at-Risk and CVaR level within a PAC setting. This approach can be generalized to random variables with bounds on $1+\nu$ moments (i.e., $E|X|^{1+\nu} \leq B$, for some finite known constant B, and $\nu > 0$), rather than bounded support; see (Hepworth, Atkinson, and Szechtman 2017).

5 APPENDICES

A. Proof of Theorem 1:

Suppose that $k_n < k(C) - \varepsilon$, for $\varepsilon > 0$; the case $k_n > k(C) + \varepsilon$ is treated below. Then, from (3),

$$\begin{aligned} 0 &\leq \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k_n) \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(k_n < X_i \leq k(C)) + \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C)) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(k(C) - \varepsilon < X_i \leq k(C)) + \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C)), \end{aligned}$$

since $X_i < C$ on the event $\{X_i < k(C) - \varepsilon\}$. It follows that, since $C > k(C)$,

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C)) \\ &\geq -\frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(k(C) - \varepsilon < X_i \leq k(C)) \\ &\geq (C - k(C))\frac{1}{n}\sum_{i=1}^{n}I(k(C) - \varepsilon < X_i \leq k(C)). \end{aligned}$$

Then it must hold that

$$\begin{aligned} &P(k_n < k(C) - \varepsilon) \\ &\leq P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C)) \geq (C - k(C))\frac{1}{n}\sum_{i=1}^{n}I(k(C) - \varepsilon < X_i \leq k(C))\right) \\ &\leq \exp\left(-2n\left(\frac{(C - k(C))P(k(C) - \varepsilon < X_i \leq k(C))}{b-a}\right)^2\right) \end{aligned}$$

by Hoeffding's Lemma. Hence, by Assumption A3 and Lemma 5.1 below,

$$P(k_n < k(C) - \varepsilon) \leq \exp(-2n\psi^2\varepsilon^2\zeta^2/(b-a)^2). \tag{8}$$

Going in the other direction, if $k_n > k(C) + \varepsilon$ then (3) results in

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C) + \varepsilon) < 0 \leq \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k_n),$$

where this covers the third possibility for the root k_n discussed in Section 2. Also, since $E[(X - C)I(X > k(C))] = 0$,

$$\begin{aligned} E[(X - C)I(X > k(C) + \varepsilon)] &= E[(C - X)I(k(C) < X \leq k(C) + \varepsilon)] \\ &> (C - k(C))P(k(C) < X \leq k(C) + \varepsilon) \qquad (9) \\ &\geq \psi\varepsilon\zeta, \end{aligned}$$

by Assumption A3 and Lemma 5.1. If follows that

$$
\begin{aligned}
&P\left(k_n > k(C)+\varepsilon\right) \\
&\le P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C)+\varepsilon) < 0\right) \\
&= P\left(E[(X-C)I(X > k(C)+\varepsilon)] - \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C)+\varepsilon) \ge E[(X-C)I(X > k(C)+\varepsilon)]\right) \qquad (10)\\
&\le P\Big(E[(X-C)I(X > k(C)+\varepsilon)] - \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C)+\varepsilon) \ge \psi\varepsilon\zeta\Big) \\
&\le \exp\left(-2n\left(\frac{\psi\varepsilon\zeta}{b-a}\right)^2\right),
\end{aligned}
$$

by (9) and Hoeffding's Lemma. In summary,

$$P(|k_n - k(C)| > \varepsilon) \le 2\exp\left(-2n\left(\frac{\psi\varepsilon\zeta}{b-a}\right)^2\right).$$

From here, the results are fed into the sequential elimination approach of (Even-Dar, Mannor, and Mansour 2006) to get the elimination algorithm, as follows. For $0 < \delta < 1$ selected by the agent, set

$$P(|k_n - k(C)| > \varepsilon_n) \le 2\exp\left(-2n\left(\frac{\psi\varepsilon\zeta}{b-a}\right)^2\right) = \frac{6\delta}{\pi^2 n^2 S}.$$

Solving for ε_n,

$$\varepsilon_n = \left(\log\left(\frac{\pi^2 n^2 S}{3\delta}\right)\frac{1}{2n}\right)^{1/2}\frac{b-a}{\zeta\psi}.$$

Thus, for any $n = 1, 2, \ldots$, and ε_n as above,

$$P\left(|k_{s,n} - k_s(C)| > \varepsilon_n\right) \le \frac{6\delta}{\pi^2 n^2 S},$$

so that we obtain

$$\sum_{n=1}^{\infty} P\left(|k_{s,n} - k_s(C)| > \varepsilon_n\right) \le \frac{\delta}{S}\sum_{n=1}^{\infty}\frac{6}{\pi^2 n^2} = \frac{\delta}{S},$$

and due to Basel's problem,

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}.$$

Hence,

$$P\left(\cup_{n,s}|k_{s,n} - k_s(C)| > \varepsilon_n\right) \le \sum_{s,n} P\left(|k_{s,n} - k_s(C)| > \varepsilon_n\right) \le \sum_{s=1}^{S}\frac{\delta}{S} \le \delta.$$

It follows that,

$$P\left(|k_{s,n} - k_s(C)| \le \varepsilon_n, \forall n, \forall s = 1, \ldots, S\right) \ge 1 - \delta,$$

and the proof is complete.

Lemma 5.1 Setting ψ to

$$\psi = \frac{\frac{b-C}{2}\frac{b-C}{2}\zeta}{1-\frac{b-C}{2}\zeta} > 0$$

satisfies $C - k(C) \geq \psi$.

Proof of Lemma 5.1: We argue that

$$\mathbf{E}[X \mid X > C-\psi] \geq C$$

which implies that $C - k(C) \geq \psi$. Indeed,

$$\begin{aligned}
\mathbf{E}[X \mid X > C-\psi] &= \frac{\mathbf{E}[XI(X > C-\psi)]}{\mathbf{P}[X > C-\psi]} \\
&= \frac{\mathbf{E}[XI(C-\psi < X \leq C+\frac{b-C}{2})]}{\mathbf{P}[X > C-\psi]} + \frac{\mathbf{E}[XI(X > C+\frac{b-C}{2})]}{\mathbf{P}[X > C-\psi]} \\
&\geq \frac{\mathbf{E}[(C-\psi)I(C-\psi < X \leq C+\frac{b-C}{2})]}{\mathbf{P}[X > C-\psi]} + \frac{\mathbf{E}[(C+\frac{b-C}{2})I(X > C+\frac{b-C}{2})]}{\mathbf{P}[X > C-\psi]} \\
&= (C-\psi)\frac{\mathbf{P}[C-\psi < X \leq C+\frac{b-C}{2}]}{\mathbf{P}[X > C-\psi]} + (C+\frac{b-C}{2})\frac{\mathbf{P}[X > C+\frac{b-C}{2}]}{\mathbf{P}[X > C-\psi]} \\
&= (C-\psi)\left(1-\frac{\mathbf{P}[X > C+\frac{b-C}{2}]}{\mathbf{P}[X > C-\psi]}\right) + (C+\frac{b-C}{2})\frac{\mathbf{P}[X > C+\frac{b-C}{2}]}{\mathbf{P}[X > C-\psi]} \\
&\geq (C-\psi)\left(1-\mathbf{P}[X > C+\frac{b-C}{2}]\right) + (C+\frac{b-C}{2})\mathbf{P}[X > C+\frac{b-C}{2}] \\
&\geq (C-\psi)\left(1-\mathbf{P}[X > C+\frac{b-C}{2}]\right) + (C+\frac{b-C}{2})(b-C-\frac{b-C}{2})\zeta \\
&\geq (C-\psi)\left(1-\frac{b-C}{2}\zeta\right) + (C+\frac{b-C}{2})(\frac{b-C}{2})\zeta \\
&\geq C-\psi\left(1-\frac{b-C}{2}\zeta\right) + \frac{b-C}{2}\frac{b-C}{2}\zeta.
\end{aligned}$$

We just need to ensure ψ is small enough so that the right hand side is at least C. By inspection,

$$\psi \leq \frac{\frac{b-C}{2}\frac{b-C}{2}\zeta}{1-\frac{b-C}{2}\zeta},$$

which completes the proof.

B. Proof of Theorem 2:

Eq. (6) leads to,

$$P\Big(\alpha_n - \alpha > \varepsilon\Big) \leq P\Big(\bar{F}(k_n) - \bar{F}(k(C)) > q\varepsilon\Big) + P\Big(\bar{F}(k(C)) - F(k(C)) > (1-q)\varepsilon\Big), \tag{11}$$

for $0 < q < 1$ and $\varepsilon > 0$. For the first term

$$P(\bar{F}(k_n) - \bar{F}(k(C)) > q\varepsilon) = P\left(\frac{1}{n}\sum_{i=1}^{n} I(k(C) < X_i \leq k_n) > q\varepsilon\right).$$

If $(1/n)\sum_{i=1}^{n} I(k(C) < X_i \le k_n) > q\varepsilon$ then

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k_n) - \frac{1}{n}\sum_{i=1}^{n}(X_i - C)I(X_i > k(C)) = \frac{1}{n}\sum_{i=1}^{n}(C - X_i)I(k(C) < X_i \le k_n) \ge (C - k_n)q\varepsilon.$$

Since $0 \le (1/n)\sum_{i=1}^{n}(X_i - C)I(X_i > k_n) \le (b - C)/n$, for $0 < \xi < \psi$,

$$\begin{aligned}
&P\left(\frac{1}{n}\sum_{i=1}^{n} I(k(C) < X_i \le k_n) > q\varepsilon\right)\\
&\le P\left(\frac{1}{n}\sum_{i=1}^{n}(C - X_i)I(X_i > k(C)) > (C - k_n)q\varepsilon - (b - C)/n\right)\\
&= P\left(\frac{1}{n}\sum_{i=1}^{n}(C - X_i)I(X_i > k(C)) > (C - k_n)q\varepsilon - (b - C)/n; k_n - k(C) > \xi\right)\\
&\qquad + P\left(\frac{1}{n}\sum_{i=1}^{n}(C - X_i)I(X_i > k(C)\right) > (C - k_n)q\varepsilon - (b - C)/n; k_n - k(C) \le \xi)\\
&\le P(k_n - k(C) > \xi) + P\left(\frac{1}{n}\sum_{i=1}^{n}(C - X_i)I(X_i > k(C)) > (\psi - \xi)q\varepsilon - (b - C)/n\right),
\end{aligned}$$

by Assumption A3 and Lemma 5.1. Hence,

$$P(\bar{F}(k_n) - \bar{F}(k(C)) > q\varepsilon) \le \exp\left(-2n\left(\frac{\psi\xi\zeta}{b - a}\right)^2\right) + \exp\left(-2n\left(\frac{(\psi - \xi)q\varepsilon - (b - C)/n}{b - a}\right)^2\right),$$

by (10) and Hoeffding's Lemma, for $n > (b - C)/((\psi - \xi)q\varepsilon)$. Regarding the second term in (11),

$$P(\bar{F}(k(C)) - F(k(C)) > (1 - q)\varepsilon) \le \exp(-2n(1 - q)^2\varepsilon^2).$$

In summary,

$$\begin{aligned}
P(\alpha_n - \alpha > \varepsilon) &\le \exp\left(-2n\left(\frac{\psi\xi\zeta}{b - a}\right)^2\right) + \exp\left(-2n\left(\frac{(\psi - \xi)q\varepsilon - (b - C)/n}{b - a}\right)^2\right) + \exp(-2n(1 - q)^2\varepsilon^2)\\
&\le 3\exp\left(-2n\left(\min\left\{\frac{\psi\xi\zeta}{b - a}, \frac{(\psi - \xi)q\varepsilon - (b - C)/n}{b - a}, (1 - q)\varepsilon\right\}\right)^2\right).
\end{aligned}$$

To shorten the proof set $\xi = \psi/2$ and $q = 1/2$ (but they could be optimized), so that

$$P(\alpha_n - \alpha > \varepsilon) \le 3\exp\left(-2n\left(\min\left\{\frac{\psi^2\varepsilon\zeta}{2(b - a)}, \frac{\psi\varepsilon/2 - (b - C)/n}{2(b - a)}, \varepsilon/2\right\}\right)^2\right),$$

for $n > 2(b - C)/(\psi\varepsilon)$. The analysis of $P(\alpha_n < \alpha - \varepsilon)$ is similar and results in an identical exponential bound; the proof is omitted for the sake of brevity. The conclusion is that,

$$P(|\alpha_n - \alpha| > \varepsilon) \le 6\exp\left(-2n\left(\min\left\{\frac{\psi^2\varepsilon\zeta}{2(b - a)}, \frac{\psi\varepsilon/2 - (b - C)/n}{2(b - a)}, \varepsilon/2\right\}\right)^2\right),$$

for $n > 2(b-C)/(\psi\varepsilon)$. As in the proof of Theorem 1, for $0 < \delta < 1$ chosen by the agent, ε_n is set so that

$$6\exp\left(-2n\left(\min\left\{\frac{\psi^2\varepsilon_n\zeta}{2(b-a)}, \frac{\psi\varepsilon_n/2-(b-C)/n}{2(b-a)}, \varepsilon_n/2\right\}\right)^2\right) \leq \frac{6\delta}{\pi^2 n^2 S},$$

leading to,

$$\varepsilon_n = \left(\log\left(\frac{\pi^2 n^2 S}{\delta}\right)\frac{2}{n}\right)^{1/2} \max\left\{\frac{b-a}{\psi^2\zeta}, \frac{2(b-a)+(b-C)/n)}{\psi}, 1\right\}.$$

By standard arguments, as in the proof of Theorem 1, it follows that,

$$P\left(|\alpha_{s,n} - \alpha_s(C)| \leq \varepsilon_n, \forall n, \forall s = 1, \ldots, S\right) \geq 1-\delta,$$

which completes the proof.

REFERENCES

Brown, D. 2007. “Large deviations bounds for estimating conditional value-at-risk”. *Operations Research Letters* 35:722–730.

Bubeck, S., and N. Cesa-Bianchi. 2012. “Regret Analysis of Stochastic and Non-Stochastic Multi-armed Bandit Problems”. *Foundations and Trends in Machine Learning* 5 (1).

Cormen, T., C. Leiserson, R. Rivset, and C. Stein. 2010. *Introduction to Algorithms*. 3rd ed. MIT Press.

David, Y., and N. Shimkin. 2016. “Pure Exploration for Max-Quantile Bandits”. *Machine Learning and Knowledge Discover in Databases, European Conference, EMCL PKDD Proceedings*:556–571.

Even-Dar, E., S. Mannor, and Y. Mansour. 2002. “PAC Bounds for Multi-Armed Bandit and Markov Decision Processes”. *In Fifteenth Annual Conference on Computational Learning Theory (COLT)*:255–270.

Even-Dar, E., S. Mannor, and Y. Mansour. 2006. “Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems”. *Journal of Machine Learning Research* 7:1079 – 1105.

Gittins, J. 1989. “Multi-armed bandit allocation indices”. *Wiley-Interscience Series in Systems and Optimization*.

Glynn, P., and S. Juneja. 2015. “Ordinal optimization - empirical large deviations rate estimators, and stochastic multi-armed bandits”. *arXiv:1507.04564*.

Hepworth, A., M. Atkinson, and R. Szechtman. 2017. “A Sequential Elimination Approach To Value-At-Risk And Conditional Value-At-Risk Selection”. *In preparation*.

Kim, S.-H., and B. L. Nelson. 2006. “Selecting the Best System”. In *Simulation*, edited by S. G. Henderson and B. L. Nelson, Volume 13 of *Handbooks in Operations Research and Management Science*. Amsterdam: Elsevier.

Rockafellar, R., and J. Royset. 2010. “On Buffered Failure Probability in Design and Optimization of Structures”. *Reliability Engineering and System Safety* 95:499–510.

Rockafellar, R., and S. Uryasev. 2000. “Optimization of conditional value-at-risk”. *Journal of Risk* 2:21–42.

Rockafellar, R., and S. Uryasev. 2002. “Conditional value-at-risk for general loss distributions”. *Journal of Banking and Finance* 26:1443–1471.

Serfling, R. 2008. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Shen, W., J. Wang, Y. Jiang, and H. Zha. 2015. “Portfolio Choices with Orthogonal Bandit Learning”. *Proceedings of International Joint Conferences on Artificial Intelligence*.

Szörényi, B., R. Busa-Fekete, P. Weng, and E. Hüllermeier. 2015. “Qualitative Multi-Armed Bandits: A Quantile-Based Approach”. *Proceedings of the 32nd International Conference on Machine Learning* 37.

Uryasev, S. 2014. “Buffered Probability of Exceedance and Buffered Service Level: Definitions and Properties”. *Research report*.

AUTHOR BIOGRAPHIES

ADAM J. HEPWORTH is an Operations Research and Management Science Analyst at the Australian Defence Force Land Warfare Lab, Canberra, AU. He is a graduate of the Master of Science in Operations Research Program at the Naval Postgraduate School. His email address is adam.hepworth@defence.gov.au.

MICHAEL P. ATKINSON is an Associate Professor of Operations Research at the Naval Postgraduate School. His research focuses on applying stochastic models to study homeland security and military applications. Dr. Atkinson received his Ph.D. in Computational and Mathematical Engineering from Stanford University in 2009. He teaches courses in statistics, computational methods, stochastic models, search theory, decision theory, and game theory. His e-mail address is mpatkins@nps.edu.

ROBERTO SZECHTMAN received his Ph.D. from Stanford University, and currently is Associate Professor in the Operations Research Department at the Naval Postgraduate School. His research interests include applied probability and military operations research. His email address is rszechtm@nps.edu.