Supplementary Materials for Quantile Bandits for Best Arms Identification

We show experiment details in Section A, the detailed proofs for both concentration inequalities (Section B) bandit task (Section C), and discussion in Section D.

A. Experiments Details

In this section, we illustrate experiment details, including simulation details (Section A.1), vaccine allocation strategy description (Section A.2), Q-SR algorithm (Section A.3).

A.1. Illustrative Example

We provide more details about the environments setting in Section 5. We consider two distributions which satisfy our assumptions: absolute Gaussian distribution (Definition 3), and exponential distribution (Definition 4).

Definition 3 (Absolute Gaussian Distribution). *Given a Gaussian random variable X with mean μ and variance σ^2, the random variable $Y = |X|$ has a absolute Gaussian distribution with p.d.f and c.d.f. shown as,*

$$f_{AbsGau}(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x+\mu)^2}{2\sigma^2}}, \tag{25}$$

$$F_{AbsGau}(\mu, \sigma^2) = \frac{1}{2}\left[\text{erf}\left(\frac{x+\mu}{\sigma\sqrt{2}}\right) + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]. \tag{26}$$

where the error function erf$(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$. We denote the absolute Gaussian distribution random variable with mean μ and variance σ^2 as $|\mathcal{N}(\mu, \sigma^2)|$. When $\mu = 0$, the lower bound of hazard rate $L = \frac{1}{\sigma\sqrt{2\pi}}$.

Definition 4 (Exponential Distribution). *With $\theta > 0$, the p.d.f and c.d.f of exponential distribution are defined as*

$$f_{Exp}(x, \theta) = \theta e^{-\theta x}, \tag{27}$$

$$F_{Exp}(x, \theta) = 1 - e^{-\theta x}, \tag{28}$$

We denote the exponential distribution with θ as $Exp(\theta)$. The hazard rate for exponential distribution is a constant and equal to θ, i.e. $h(x) = \theta$.

We design our experimental environments based on three configurations of reward distributions: A) $|\mathcal{N}(0, 2)|$ B) $|\mathcal{N}(3.5, 2)|$ C) Exp(1/4). The histogram of these three arms is shown below.

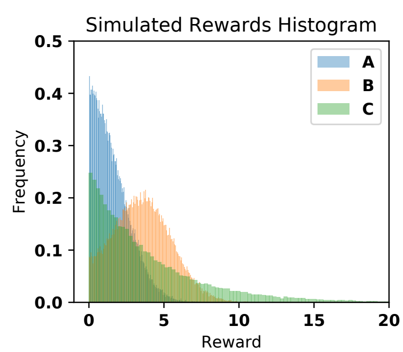

Figure 7. Simulated Arm Rewards Histogram.

A.2. Vaccine Allocation Strategy

We provide more details about the vaccine allocation strategy in this section. We allocate 100 vaccine doses (5% of the population) to 5 age groups (0-4 years, 5-18 years, 19-29 years, 30-64 years and >65 years). We consider all combinations of groups (resulting in $K = 32$ arms), and denote the allocation scheme as a Boolean 5-tuple, with each position corresponds to the respective age group (1 represents allocation; 0 otherwise). We use the median ($\tau = 0.5$) as a robust summary statistic for each strategy. For the task of identifying the best subset of ages ($m = 1$) Q-SAR finds that the optimal arm is $(0, 1, 0, 0, 0)$, i.e. only allocation to 5-18 years old group. For identifying the $m = 3$ best arms, the optimal arms are $(0, 1, 0, 0, 0)$, $(1, 1, 0, 0, 0)$ and $(0, 1, 1, 0, 0)$, indicated as arms 8, 24, and 12 in Figure 5 respectively.

A.3. Q-SR

We extend the Successive Rejects algorithm (Audibert et al., 2010) to a quantile version and adapt it to recommend more than one arm.

Algorithm 2 Q-SR

Denote the active set $\mathcal{A}_1 = \{1, ..., K\}$, $\widetilde{\log}(K) = \frac{m}{m+1} + \sum_{p=1}^{K-m} \frac{1}{K+1-p}$, $n_0 = 0$, and for $k \in \{1, ..., K-1\}$,

$$n_p = \left\lceil \frac{1}{\widetilde{\log}(K)} \frac{N-K}{K+1-p} \right\rceil$$

For each phase $p = 1, 2, ..., K - m$:
(1) For each $i \in \mathcal{A}_p$, select arm i for $n_p - n_{p-1}$ rounds.
(2) Let $\mathcal{A}_{p+1} = \mathcal{A}_p / \operatorname{argmin}_{i \in \mathcal{A}_p} \hat{Q}^{\tau}_{i,n_p}$
The recommended set is $\mathcal{A}_{K-m+1}$.

We provide justifications for the design choice of our proposed Q-SR algorithm shown in Algorithm 2. Although that both SR and SAR are analysed on reward distributions with support $[0, 1]$, they can both be directly extended to subgaussian reward distributions (Audibert et al., 2010; Bubeck et al., 2013). We propose Quantile-based Successive Rejects (Q-SR), adapted from Successive Rejects (SR) algorithm (Audibert et al., 2010). To be able to recommend multiple arms, the total phase is designed to be $K - m$ instead of $K - 1$, and the number of pulls for each round is modified to make sure all budgets are used. More precisely, one is pulled $n_1 = \left\lceil \frac{1}{\widetilde{\log}(K)} \frac{N-K}{K} \right\rceil$ times, one is pulled $n_2 = \left\lceil \frac{1}{\widetilde{\log}(K)} \frac{N-K}{K-1} \right\rceil$ times, $\ldots$, $m+1$ is pulled $n_{K-m} = \left\lceil \frac{1}{\widetilde{\log}(K)} \frac{N-K}{K+1-(K-m)} \right\rceil$ times, then

$$n_1 + \cdots + (m+1)n_{K-m} \leq K + \frac{N-K}{\widetilde{\log}(K)} \left(\frac{m}{m+1} + \sum_{p=1}^{K-m} \frac{1}{K+1-p} \right) = N. \tag{29}$$

As shown in Section 4, when $m = 1$, the Q-SAR algorithm can be reduced to the Q-SR algorithm. So the theoretical performance of the Q-SR algorithm is guaranteed. We leave the theoretical analysis of Q-SR for $m > 1$ for the future work.

B. Concentration Inequality Proof

This section shows the proofs of the concentration results shown in Section 2. In the following, we will walk through the key statement and show how we achieve our results in details. For the reader's convenience, we restate our theorems in the main paper whenever needed. We first introduce the Modified logarithmic Sobolev inequality, which gives the upper bound of the entropy (Eq. (15)) of $\exp(\lambda W)$.

Theorem 5 (Modified logarithmic Sobolev inequality (Ledoux, 2001))**.** *Consider independent random variables $X_1, \ldots, X_n$, let a real-valued random variable $W = f(X_1, \ldots, X_n)$, where $f : \mathbb{R}^n \to \mathbb{R}$ is measurable. Let $W_i = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, where $f_i : \mathbb{R}^{n-1} \to \mathbb{R}$ is an arbitrary measurable function. Let $\phi(x) = \exp(x) - x - 1$. Then for any $\lambda \in \mathbb{R}$,*

$$\operatorname{Ent}[\exp(\lambda W)] = \lambda \mathbb{E}[W \exp(\lambda W)] - \mathbb{E}[\exp(\lambda W)] \log \mathbb{E}[\exp(\lambda W)] \tag{30}$$

$$\leq \sum_{i=1}^{n} \mathbb{E}[\exp(\lambda W) \phi(-\lambda(W - W_i))] \tag{31}$$

Consider i.i.d random variables $X_1, \ldots, X_n$, and the corresponding order statistics $X_{(1)} \geq \cdots \geq X_{(n)}$. Define the spacing between rank k and $k+1$ order statistics as $S_k = X_{(k)} - X_{(k+1)}$. By taking W as k rank order statistics (or negative k

rank), and W_i as nearest possible order statistics, i.e. $k \pm 1$ rank (or negative $k \pm 1$ rank), Theorem 5 provides the connection between the order statistics and the spacing between order statistics. The connection is shown in Proposition 1.

Proposition 1 (Entropy upper bounds)**.** *Define* $\phi(x) := \exp(x) - x - 1$ *and* $\zeta(x) := \exp(x)\phi(-x) = 1 + (x-1)\exp(x)$. *For all* $\lambda \geq 0$, *and for* $k \in [1, n) \wedge \mathbb{N}^*$,

$$\operatorname{Ent}\left[\exp(\lambda X_{(k)})\right] \leq k\mathbb{E}\left[\exp(\lambda X_{(k+1)})\zeta\left(\lambda S_k\right)\right]. \tag{18}$$

For $k \in (1, n] \wedge \mathbb{N}^*$,

$$\begin{aligned}&\operatorname{Ent}\left[\exp(-\lambda X_{(k)})\right] \leq \\ &\qquad (n-k+1)\mathbb{E}\left[\exp(-\lambda X_{(k)})\phi\left(-\lambda S_{k-1}\right)\right].\end{aligned} \tag{19}$$

Proof. We prove the upper bound based on Theorem 5. We first prove Eq. (18). We define W, W_i with f^k, f_i^k as following. Let W be the rank k order statistics of $X_1, \ldots, X_n$, i.e. $W = f^k(X_1, \ldots, X_n) = X_{(k)}$; Let W_i be the rank k order statistics of $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ (i.e. with X_i removed from $X_1, \ldots, X_n$), i.e. $f_i = X_{(k+1)}\mathbb{I}(X_i \geq X_{(k)}) + X_{(k)}\mathbb{I}(X_i < X_{(k)})$. So $W_i = X_{(k+1)}$ when the removed element is bigger and equal to $X_{(k)}$, otherwise $W = X_{(k)}$. Then the upper bound of $\operatorname{Ent}\left[\exp(\lambda X_{(k)})\right]$ is

$$\begin{aligned}
&\operatorname{Ent}\left[\exp(\lambda X_{(k)})\right] \\
\leq&\mathbb{E}\big[\sum_{i=1}^{n} \exp(\lambda X_{(k)})\phi\big(-\lambda(X_{(k)} - X_{(k+1)}\mathbb{I}(X_i \geq X_{(k)}) - X_{(k)}\mathbb{I}(X_i < X_{(k)}))\big] && \text{Theorem 5} && (32)\\
=&\mathbb{E}\big[\exp(\lambda X_{(k)})\phi\big(-\lambda(X_{(k)} - X_{(k+1)})\big)\sum_{i=1}^{n}\mathbb{I}(X_i \geq X_{(k)})\big] && && (33)\\
=&k\mathbb{E}\left[\exp(\lambda X_{(k)})\phi\left(-\lambda S_k\right)\right] && && (34)\\
=&k\mathbb{E}\left[\exp(\lambda X_{(k+1)})\exp(\lambda S_k)\phi\left(-\lambda S_k\right)\right] && && (35)\\
=&k\mathbb{E}\left[\exp(\lambda X_{(k+1)})\zeta\left(\lambda S_k\right)\right] && \zeta(x) = \exp(x)\phi(-x) && (36)
\end{aligned}$$

Similarly, for the proof of Eq. (19), We define W, W_i with $\widetilde{f}^k$, $\widetilde{f}_i^{k-1}$. Let W be the negative value of k rank order statistics of $X_1, \ldots, X_n$, i.e. $W = \widetilde{f}^k(X_1, \ldots, X_n) = -X_{(k)}$; Let W_i be the negative value of $k-1$ rank order statistics of $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$. Thus when $X_i \geq X_{(k-1)}$, $W_i = -X_{(k)}$, otherwise $W_i = -X_{(k-1)}$. Then by Theorem 5, we get $\operatorname{Ent}\left[\exp(-\lambda X_{(k)})\right] \leq (n-k+1)\mathbb{E}\left[\exp(-\lambda X_{(k)})\phi\left(-\lambda S_{k-1}\right)\right]$.

□

Compared with the proof in Boucheron & Thomas (2012), we do not choose a different initialisation of W_i in terms of the two cases $k \leq n/2$ and $k > n/2$, which does not influence the concentration rates of empirical quantiles, and allows us to extend the proof to all ranks (excluding extremes). We derive upper bounds for both $\operatorname{Ent}\left[\exp(\lambda X_{(k)})\right]$ and $\operatorname{Ent}\left[\exp(-\lambda X_{(k)})\right]$, which allows us to derive two-sided concentration bounds instead of one-sided bound. Now we show the proof of Theorem 3.

Theorem 3 (Extended Exponential Efron-Stein inequality)**.** *With the logarithmic moment generating function defined in Eq. 14, for* $\lambda \geq 0$ *and* $k \in [1, n) \wedge \mathbb{N}^*$,

$$\psi_{Z_k}(\lambda) \leq \lambda\frac{k}{2}\mathbb{E}\left[S_k\left(\exp(\lambda S_k) - 1\right)\right]. \tag{20}$$

For $k \in (1, n] \wedge \mathbb{N}^*$,

$$\psi_{Z'_k}(\lambda) \leq \frac{\lambda^2(n-k+1)}{2}\mathbb{E}[S_{k-1}^2]. \tag{21}$$

Proof. The proof of Eq. (20) is based on Proposition 1 and follows the same reasoning from (Boucheron & Thomas, 2012) Theorem 2.9. Note since Eq. (18) holds for $k \in [1, n)$, Eq. (20) can be proved for $k \in [1, n)$ (Boucheron & Thomas (2012) only proved for $k \in [1, n/2]$).

We now prove Eq. (21). Recall $\phi(x) = \exp(x) - x - 1$. $\phi(x)$ is nonincreasing when $x \leq 0$ and nondecreasing otherwise. By Proposition 1 and Proposition 4 (which will be shown later), for $\lambda \geq 0$,

$$\begin{aligned}\mathrm{Ent}\left[\exp(-\lambda X_{(k)})\right] \leq &(n-k+1)\mathbb{E}\left[\exp(-\lambda X_{(k)})\phi\left(-\lambda S_{k-1}\right)\right] && \text{By Proposition 1} \quad (37)\\ \leq &(n-k+1)\mathbb{E}\left[\exp(-\lambda X_{(k)})\right]\mathbb{E}\left[\phi\left(-\lambda S_{k-1}\right)\right] && \text{By Proposition 4} \quad (38)\end{aligned}$$

Multiplying both sides by $\exp(\lambda \mathbb{E}[X_{(k)}])$,

$$\mathrm{Ent}[\exp(\lambda Z'_k)] \leq (n-k+1)\mathbb{E}[\exp(\lambda Z'_k)]\mathbb{E}[\phi(-\lambda S_{k-1})]. \tag{39}$$

With the fact $\phi(x) \leq \frac{1}{2}x^2$ when $x \leq 0$, and $-\lambda S_{k-1} \leq 0$, we have $\mathbb{E}[\phi(-\lambda S_{k-1})] \leq \frac{\lambda^2}{2}\mathbb{E}[S_{k-1}^2]$. We then obtain

$$\frac{\mathrm{Ent}\left[\exp(\lambda Z'_k)\right]}{\lambda^2\mathbb{E}\left[\exp(\lambda Z'_k)\right]} \leq \frac{n-k+1}{\lambda^2}\mathbb{E}[\phi(-\lambda S_{k-1})] \leq \frac{n-k+1}{2}\mathbb{E}[S_{k-1}^2]. \tag{40}$$

We now solve this integral inequality. Integrating left side, with the fact that $\lim_{\lambda\to 0}\frac{1}{\lambda}\log\mathbb{E}\exp(\lambda Z'_k) = 0$, for $\lambda \geq 0$, we have

$$\int_0^\lambda \frac{\mathrm{Ent}\left[\exp(tZ'_k)\right]}{t^2\mathbb{E}\left[\exp(tZ'_k)\right]}dt = \int_0^\lambda \frac{\mathbb{E}[tZ'_k] - \log\mathbb{E}[\exp(tZ'_k)]}{t^2}dt = \frac{\log\mathbb{E}[\exp(tZ'_k)]}{t}\Big|_0^\lambda = \frac{1}{\lambda}\log\mathbb{E}[\exp(\lambda Z'_k)]. \tag{41}$$

Integrating right side, for $\lambda \geq 0$,

$$\int_0^\lambda \frac{n-k+1}{2}\mathbb{E}[S_{k-1}^2]dt = \frac{\lambda(n-k+1)}{2}\mathbb{E}[S_{k-1}^2]. \tag{42}$$

Combining Eq. (40), (41) and (42), we get

$$\psi_{Z'_k}(\lambda) = \log\mathbb{E}[\exp(\lambda Z'_k)] \leq \frac{\lambda^2(n-k+1)}{2}\mathbb{E}[S_{k-1}^2]. \tag{43}$$

which concludes the proof.

□

To further bound the order statistic spacings in expectation, we introduce the R-transform (Definition 5) and Rényi's representation (Theorem 6). In the sequel, f is a monotone function from (a,b) to (c,d), its generalised inverse $f^{\leftarrow} : (c,d) \to (a,b)$ is defined by $f^{\leftarrow}(y) = \inf\{x : a < x < b, f(x) \geq y\}$. Observe that the R-transform defined in Definition 5 is the quantile transformation with respect to the c.d.f of standard exponential distribution, i.e. $F^{\leftarrow}\left(F_{exp}\left(t\right)\right)$.

Definition 5 (R-transform)**.** *The R-transform of a distribution F is defined as the non-decreasing function on* $[0,\infty)$ *by* $R(t) = \inf\{x : F(x) \geq 1 - \exp(-t)\} = F^{\leftarrow}(1-\exp(-t))$.

Theorem 6 (*Rényi's representation, Theorem 2.5 in (Boucheron & Thomas, 2012)*)**.** *Let* $X_{(1)} \geq \ldots \geq X_{(n)}$ *be the order statistics of samples from distribution F,* $Y_{(1)} \geq Y_{(2)} \geq \ldots \geq Y_{(n)}$ *be the order statistics of independent samples of the standard exponential distribution, then*

$$\left(Y_{(n)},\ldots,Y_{(k)},\ldots,Y_{(1)}\right) \stackrel{d}{=} \left(\frac{E_n}{n},\ldots,\sum_{i=k}^n \frac{E_i}{i},\ldots,\sum_{i=1}^n \frac{E_i}{i}\right), \tag{44}$$

where $E_1,\ldots,E_n$ *are independent and identically distributed (i.i.d.) standard exponential random variables, and*

$$\left(X_{(n)},\ldots,X_{(1)}\right) \stackrel{d}{=} \left(R\left(Y_{(n)}\right),\ldots,R\left(Y_{(1)}\right)\right), \tag{45}$$

where $R\left(\cdot\right)$ *is the R-transform defined in Definition 5, equality in distribution is denoted by* $\stackrel{d}{=}$.

The Rényi's representation shows the order statistics of an Exponential random variable are linear combinations of independent Exponentials, which can be extended to the representation for order statistics of a general continuous F by quantile transformation using R-transform. The following proposition states the connection between the *IHR* and R-transform.

Proposition 3 (Proposition 2.7 (Boucheron & Thomas, 2012))**.** *Let F be an absolutely continuous distribution function with hazard rate h (assuming density exists), the derivative of R-transform is $R' = 1/h\,(R)$. Then if the hazard rate h is non-decreasing (Assumption 1), then for all $t > 0$ and $x > 0$, $R\,(t+x) - R(t) \leq x/h\,(R(t))$.*

We now show Proposition 4 based on the Rényi's representation (Theorem 6) and Harris' inequality (Theorem 7). Proposition 4 allows us to upper bound the expectation of multiplication of two functions in terms of the multiplication of expectation of those two functions respectively. We will use this property to prove Theorem 3.

Theorem 7 (Harris' inequality (Boucheron et al., 2013))**.** *Let $X_1, \dots, X_n$ be independent real-valued random variables and define the random vector $X = (X_1, \dots, X_n)$ taking values in $\mathbb{R}^n$. If $f : \mathbb{R}^n \to \mathbb{R}$ is nonincreasing and $g : \mathbb{R}^n \to \mathbb{R}$ is nondecreasing then*

$$\mathbb{E}[f(X)g(X)] \leq \mathbb{E}[f(X)]\mathbb{E}[g(X)]$$

Proposition 4 (Negative Association)**.** *Let the order statistics spacing of rank $k-1$ as $S_{k-1} = X_{(k-1)} - X_{(k)}$. Then $X_{(k)}$ and S_{k-1} are negatively associated: for any pair of non-increasing function f_1 and f_2,*

$$\mathbb{E}\left[f_1(X_{(k)})f_2\left(S_{k-1}\right)\right] \leq \mathbb{E}\left[f_1(X_{(k)})\right]\mathbb{E}\left[f_2\left(S_{k-1}\right)\right]. \tag{46}$$

Proof. From Definition 5 and Theorem 6, let $Y_{(1)}, \dots, Y_{(n)}$ be the order statistics of an exponential sample. Let $E_{k-1} = Y_{(k-1)} - Y_{(k)}$ be the $(k-1)^{\text{th}}$ spacing of the exponential sample. By Theorem 6, E_{k-1} and $Y_{(k)}$ are independent.

$$\begin{aligned}
\mathbb{E}\left[f_1\left(X_{(k)}\right)f_2\left(S_{k-1}\right)\right] =&\mathbb{E}\left[f_1(R(Y_{(k)}))f_2\left(R(Y_{(k-1)}) - R(Y_{(k)})\right)\right] && (47)\\
=&\mathbb{E}\left[\mathbb{E}\left[f_1(R(Y_{(k)}))f_2\left(R(E_{k-1}+Y_{(k)}) - R(Y_{(k)})\right)|Y_{(k)}\right]\right] && (48)\\
=&\mathbb{E}\left[f_1(R(Y_{(k)}))\mathbb{E}\left[f_2\left(R(E_{k-1}+Y_{(k)}) - R(Y_{(k)})\right)|Y_{(k)}\right]\right]. && (49)
\end{aligned}$$

The function $f_1 \circ R$ is non-increasing. Almost surely, the conditional distribution of $(k-1)E_{k-1}$ w.r.t $Y_{(k)}$ is the exponential distribution.

$$\mathbb{E}\left[f_2\left(R(E_{k-1}+Y_{(k)}) - R(Y_{(k)})\right)|Y_{(k)}\right] = \int_0^\infty e^{-x} f_2(R(\frac{x}{k-1}+Y_{(k)}) - R(Y_{(k)}))dx. \tag{50}$$

As F is IHR, $R(\frac{x}{k-1}+y) - R(y) = \int_0^{x/(k-1)} R'(y+z)dz$ is non-increasing w.r.t. y (from Proposition 3 we know R is concave when F is IHR). Then $\mathbb{E}\left[f_2\left(R(E_{k-1}+Y_{(k)}) - R(Y_{(k)})\right)|Y_{(k)}\right]$ is non-decreasing function of $Y_{(k)}$. By Harris' inequality,

$$\begin{aligned}
\mathbb{E}\left[f_1\left(X_{(k)}\right)f_2\left(S_{k-1}\right)\right] \leq&\mathbb{E}\left[f_1(R(Y_{(k)}))\right]\mathbb{E}\left[\mathbb{E}\left[f_2\left(R(E_{k-1}+Y_{(k)}) - R(Y_{(k)})\right)|Y_{(k)}\right]\right] && (51)\\
=&\mathbb{E}\left[f_1(R(Y_{(k)}))\right]\mathbb{E}\left[f_2\left(R(E_{k-1}+Y_{(k)}) - R(Y_{(k)})\right)\right] && (52)\\
=&\mathbb{E}\left[f_1(X_{(k)})\right]\mathbb{E}\left[f_2(S_{k-1})\right]. && (53)
\end{aligned}$$

□

We prove Proposition 2 in the following by transform the spacing based on Rényi's representation and the property described in Proposition 3.

Proposition 2. *For any $k \in [1, n) \wedge \mathbb{N}^*$, the expectation of spacing S_k defined in Eq. (17) can be bounded under Assumption 1, $\mathbb{E}[S_k] \leq \frac{1}{kL}$.*

Proof. We show the upper bound the expectations of the k^{th} spacing of order statistics, assuming the lower bound hazard rate is L. The following proof uses Proposition 3, which requires Assumption 1 hold.

$$
\begin{aligned}
\mathbb{E}[S_k] &= \mathbb{E}[X_{(k)} - X_{(k+1)}] \\
&= \mathbb{E}[R\left(Y_{(k+1)} + \frac{E_k}{k}\right) - R\left(Y_{(k+1)}\right)] && \text{By Theorem 6} && (54)\\
&= \int_Y \int_E \left(R\left(y + \frac{z}{k}\right) - R\left(y\right)\right) f_Y\left(y\right) f_E\left(z\right) \mathrm{d}z\mathrm{d}y && && (55)\\
&\leq \int_Y \int_E \frac{z}{k \times h\left(R\left(y\right)\right)} f_Y\left(y\right) f_E\left(z\right) \mathrm{d}z\mathrm{d}y && \text{By Proposition 3} && (56)\\
&\leq \int_E \frac{z}{kL} f_E\left(z\right) \mathrm{d}z = \frac{1}{kL} && L \text{ is the lower bound of the hazard rate.} && (57)
\end{aligned}
$$

□

Using the same technique of shown in Proposition 2, we prove Lemma 1 and Lemma 2 by further bounding inequalities shown in Theorem 3.

Lemma 1 (Right Tail Concentration Bounds for Order Statistics)**.** *Define $v^r := \frac{2}{kL^2}$, $c^r := \frac{2}{kL}$. Under Assumption 1 , for all $\lambda \in [0, 1/c^r)$, and all $k \in [1, n) \wedge \mathbb{N}^*$, we have*

$$
\log \mathbb{E}[\exp\left(\lambda\left(X_{(k)} - \mathbb{E}[X_{(k)}]\right)\right)] \leq \frac{\lambda^2 v^r}{2(1 - c^r\lambda)}. \tag{6}
$$

For all $\gamma \geq 0$, we obtain the concentration inequality

$$
\mathbb{P}\left(X_{(k)} - \mathbb{E}[X_{(k)}] \geq \sqrt{2v^r\gamma} + c^r\gamma\right) \leq \exp(-\gamma). \tag{7}
$$

Proof. We first prove Eq. (6). From Theorem 6, we can represent the spacing as $S_k = X_{(k)} - X_{(k+1)} \overset{d}{=} R\left(Y_{(k+1)+E_k/k}\right) - R\left(Y_{(k+1)}\right)$, where E_k is standard exponentially distributed and independent of $Y_{(k+1)}$. The following proof uses Proposition 3, which requires Assumption 1 hold.

$$
\begin{aligned}
\psi_{Z_k}(\lambda) \leq& \lambda\frac{k}{2}\mathbb{E}\left[S_k\left(\exp(\lambda S_k) - 1\right)\right] && \text{By Theorem 3} && (58)\\
\leq& \lambda\frac{k}{2}\int_E\int_Y \frac{z}{h\left(R\left(y\right)\right)k}\left(\exp(\frac{\lambda z}{h\left(R\left(y\right)\right)k}) - 1\right) f_Y\left(y\right) f_E\left(z\right) \mathrm{d}y\mathrm{d}z && \text{By Proposition 3} && (59)\\
\leq& \frac{k}{2}\int_E \frac{\lambda}{Lk} z\left(\exp(\frac{\lambda}{Lk}z) - 1\right) f_E\left(z\right)\mathrm{d}z && && (60)\\
=& \frac{k}{2}\int_0^\infty \frac{\lambda}{Lk} z\left(\exp(\frac{\lambda}{Lk}z) - 1\right)\exp(-z)\mathrm{d}z && && (61)\\
\leq& \frac{\lambda^2 v^r}{2(1 - c^r\lambda)}, && \text{With } v^r = \frac{2}{kL^2}, c^r = \frac{2}{kL} && (62)
\end{aligned}
$$

The last step is because for $0 \leq \mu \leq \frac{1}{2}$, $\int_0^\infty \mu z\left(\exp(\mu z) - 1\right)\exp(-z)\mathrm{d}z = \frac{\mu^2(2-\mu)}{(1-\mu)^2} \leq \frac{2\mu^2}{1-2\mu}$. where we let $\mu = \frac{\lambda}{Lk}$.

From Eq. (6) to Eq. (7), we convert the bound of logarithmic moment generating function to the tail bound by using the Cramér-Chernoff method (Boucheron et al., 2013). Markov's inequality implies, for $\lambda \geq 0$,

$$
\mathbb{P}(Z_k \geq \gamma) \leq \exp(-\lambda\gamma)\mathbb{E}[\exp(\lambda Z_k)]. \tag{63}
$$

To choose λ to minimise the upper bound, one can introduce $\psi^*_{Z_k}(\gamma) = \sup_{\lambda \geq 0}\left(\lambda\gamma - \psi_{Z_k}(\lambda)\right)$. Then we get $\mathbb{P}(Z_k \geq \gamma) \leq \exp\left(-\psi^*_{Z_k}(\gamma)\right)$. Set $h_1(u) := 1 + u - \sqrt{1 + 2u}$ for $u > 0$, we have

$$
\psi^*_{Z_k}(t) = \sup_{\lambda \in (0, 1/c^r)}(\gamma\lambda - \frac{\lambda^2 v^r}{2(1 - c^r\lambda)}) = \frac{v^r}{(c^r)^2} h_1(\frac{c^r\gamma}{v^r}) \tag{64}
$$

Since h_1 is an increasing function from $(0, \infty)$ to $(0, \infty)$ with inverse function $h_1^{-1}(u) = u + \sqrt{2u}$ for $u > 0$, we have $\psi^{*-1}(u) = \sqrt{2v^r u} + c^r u$. Eq. (7) is thus proved. □

Lemma 2 (Left Tail Concentration Bounds for Order Statistics)**.** *Define* $v^l := \frac{2(n-k+1)}{(k-1)^2L^2}$. *Under Assumption 1, for all* $\lambda \geq 0$*, and all* $k \in (1, n] \wedge \mathbb{N}^*$*, we have*

$$\log \mathbb{E}[\exp\left(\lambda\left(\mathbb{E}[X_{(k)}] - X_{(k)}\right)\right)] \leq \frac{\lambda^2 v^l}{2}. \tag{8}$$

For all $\gamma \geq 0$*, we obtain the concentration inequality*

$$\mathbb{P}\left(\mathbb{E}[X_{(k)}] - X_{(k)} \geq \sqrt{2v^l\gamma}\right) \leq \exp(-\gamma). \tag{9}$$

Proof. The proof is similar to the proof of Lemma 1. From Theorem 6, we can represent the spacing as $S_{k-1} = X_{(k-1)} - X_{(k)} \stackrel{d}{=} R\left(Y_{(k)+E_{k-1}/(k-1)}\right) - R\left(Y_{(k)}\right)$, where E_{k-1} is standard exponentially distributed and independent of $Y_{(k)}$. The following proof uses Proposition 3, which requires Assumption 1 hold.

$$\begin{aligned}
\psi_{Z'_k}(\lambda) \leq & \frac{\lambda^2(n-k+1)}{2}\mathbb{E}[S_{k-1}^2] && \text{By Theorem 3} && (65)\\
\leq & \frac{\lambda^2(n-k+1)}{2}\int_Y\int_E\left(\frac{z}{(k-1)\times h\left(R\left(y\right)\right)}\right)^2 f_Y\left(y\right) f_E\left(z\right)\mathrm{d}z\mathrm{d}y && \text{By Proposition 3} && (66)\\
\leq & \frac{\lambda^2(n-k+1)}{2}\int_0^\infty\left(\frac{z}{(k-1)L}\right)^2\exp(-z)\mathrm{d}z && && (67)\\
\leq & \frac{\lambda^2 v^l}{2}. && \text{With } v^l = \frac{2(n-k+1)}{(k-1)^2L^2} && (68)
\end{aligned}$$

Eq. (8) is proved. Follow the Cramér-Chernoff method described above, we can prove Eq. (9). □

The concentration results for order statistics can be of independent interest. For example, one can take this result and derive the concentration for sum of order statistics by applying Hoeffding's inequality (Hoeffding, 1994) or Bernstein's inequality (Bernstein, 1924). Kandasamy et al. (2018) took the results from Boucheron & Thomas (2012) and showed such results, but limited for right tail result for exponential random variables of rank 1 order statistics (i.e. maximum).

Now we convert the concentration results of order statistics to the quantiles, based on the results from Lemma 1 and 2, and the Theorem 1 , which shows connection between expected order statistics and quantiles.

Theorem 2 (Two-side Concentration Inequality for Quantiles)**.** *Recall* $v^r = \frac{2}{kL^2}$, $v^l = \frac{2(n-k+1)}{(k-1)^2L^2}$, $c^r = \frac{2}{kL}$, $w_n = \frac{b}{n}$. *For quantile level* $\tau \in (0, 1)$*, let rank* $k = \lfloor n(1-\tau)\rfloor$*. Under Assumption 1 and 2, we have*

$$\mathbb{P}\left(\hat{Q}_n^\tau - Q^\tau \geq \sqrt{2v^r\gamma} + c^r\gamma + w_n\right) \leq \exp(-\gamma).$$
$$\mathbb{P}\left(Q^\tau - \hat{Q}_n^\tau \geq \sqrt{2v^l\gamma} + w_n\right) \leq \exp(-\gamma).$$

Proof. Denote the confidence interval for the right tail bound of order statistics as $d^r_{k,\gamma} = \sqrt{2v^r\gamma} + c^r\gamma$. From Lemma 1, we have $\mathbb{P}\left(X_{(k)} - \mathbb{E}[X_{(k)}] \geq d^r_{k,\gamma}\right) \leq \exp(-\gamma)$. With $k = \lfloor n(1-\tau)\rfloor$, we have $\hat{Q}_n^\tau = X_{(k)}$ and from Theorem 1, we have $\mathbb{E}[X_{(k)}] \leq Q^\tau + w_n$. With probability at least $1 - \exp(-\gamma)$, the following event holds

$$X_{(k)} - \mathbb{E}[X_{(k)}] < d^r_{k,\gamma} \Rightarrow X_{(k)} < \mathbb{E}[X_{(k)}] + d^r_{k,\gamma} \leq Q^\tau + w_n + d^r_{k,\gamma} \Rightarrow \hat{Q}_n^\tau - Q^\tau < w_n + d^r_{k,\gamma}, \tag{69}$$

from which we have $\mathbb{P}(\hat{Q}_n^\tau - Q^\tau \geq w_n + d^r_{k,\gamma}) \leq \exp(-\gamma)$.

Denote the confidence interval for the right tail bound of order statistics as $d^l_{k,\gamma} = \sqrt{2v^l\gamma}$. From Lemma 2, we have $\mathbb{P}\left(\mathbb{E}[X_{(k)}] - X_{(k)} \geq d^l_{k,\gamma}\right) \leq \exp(-\gamma)$. With $k = \lfloor n(1-\tau)\rfloor$, we have $\hat{Q}_n^\tau = X_{(k)}$ and from Theorem 1, we have $\mathbb{E}[X_{(k)}] \geq Q^\tau - w_n$. With probability at least $1 - \exp(-\gamma)$, the following event holds

$$\mathbb{E}[X_{(k)}] - X_{(k)} < d^l_{k,\gamma} \Rightarrow -X_{(k)} < -\mathbb{E}[X_{(k)}] + d^l_{k,\gamma} \leq -(Q^\tau - w_n) + d^l_{k,\gamma} \Rightarrow Q^\tau - \hat{Q}_n^\tau < w_n + d^l_{k,\gamma}, \tag{70}$$

from which we have $\mathbb{P}(Q^\tau - \hat{Q}_n^\tau \geq w_n + d^l_{k,\gamma}) \leq \exp(-\gamma)$. This concludes the proof. □

In ths following, we show the representations for the concentration results.

Corollary 2 (Representation of Concentration inequalities for Order Statistics)**.** *For $\epsilon > 0$, the concentration inequalities for order statistics in Lemma 1 and 2 can be represented as*

$$\mathbb{P}\left(X_{(k)} - \mathbb{E}[X_{(k)}] \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2(v^r + c^r\epsilon)}\right), \tag{71}$$

$$\mathbb{P}\left(\mathbb{E}[X_{(k)}] - X_{(k)} \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2v^l}\right). \tag{72}$$

Proof. Eq. (72) follows by setting $\epsilon = \sqrt{2v^l\gamma}$. We now show the case for Eq. (71). Recall from the proof of Lemma 1 Eq. (64), $h_1(u) = 1 + u - \sqrt{1 + 2u}$. Follow the elementary inequality

$$h_1(u) \geq \frac{u^2}{2(1+u)} \quad u > 0. \tag{73}$$

Lemma 1 implies $\psi^*_{Z_k}(t) \geq \frac{t^2}{2(v+ct)}$, so the statement Eq. (71) follows from Chernoff's inequality. □

Recall that for Q-SAR, we are interested in events of small probability, that is for large values of γ in Theorem 2. In the corollary below, we focus on such events of small probability by considering $\gamma \geq 1$ (i.e. error less than $\frac{1}{e} \approx 0.37$), which allows a simpler expression.

Corollary 1 (Representation of Concentration inequalities for Quantiles)**.** *For $\epsilon > 0$, v^r, v^l, c^r, w_n stay the same as stated in Theorem 2. With $\gamma \geq 1$, Theorem 2 can be represented as*

$$\mathbb{P}\left(\hat{Q}^\tau_n - Q^\tau \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2(v^r + (c^r + w_n)\epsilon)}\right),$$
$$\mathbb{P}\left(Q^\tau - \hat{Q}^\tau_n \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2(v^l + w_n\epsilon)}\right).$$

Proof. With $\gamma \geq 1$, we have $\gamma w_n \geq w_n$, then with probability at least $1 - \exp(-\gamma)$, we have

$$\hat{Q}^\tau_n - Q^\tau \leq \sqrt{2v^r\gamma} + c^r\gamma + w_n \leq \sqrt{2v^r\gamma} + c^r\gamma + w_n\gamma.$$

That is, we have

$$\mathbb{P}\left(\hat{Q}^\tau_n - Q^\tau \geq \sqrt{2v^r\gamma} + (c^r + w_n)\gamma\right) \leq \exp(-\gamma). \tag{74}$$

Similarly, one can prove the other side. Then the similar reasoning as shown in proof of Corollary 2 concludes the proof. □

C. Bandits Proof

In this section, we provide the proof for the bandit theoritical result (Section C.2), with a re-expression of the concentration result (Section C.1).

C.1. Re-expression of Concentration Results

The proof of Q-SAR error bound uses the concentration results for quantiles. We first further derive the result shown in Corollary 1 to show direct dependency on the number of samples n. Recall the rank $k = \lfloor n(1-\tau) \rfloor$ with quantile level $\tau \in (0, 1)$, which can be re-expressed as

$$\frac{k}{1-\tau} \leq n \leq \frac{k+1}{1-\tau} \tag{75}$$

$$n(1-\tau) - 1 \leq k \leq n(1-\tau). \tag{76}$$

We show the representation of concentration depending on n in the following.

Lemma 3. *For $\epsilon > 0$, recall n denotes the number of samples, b is a constant depending on the density about τ-quantile ($0 < \tau < 1$), L is the lower bound of hazard rate. With $n \geq \frac{4}{1-\tau}$, Corollary 1 can be represented as*

$$\mathbb{P}\left(\hat{Q}_n^\tau - Q^\tau \geq \epsilon\right) \leq \exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(\frac{8}{3} + \frac{4}{3}(2L + b(1-\tau)L^2)\epsilon)}\right),$$

$$\mathbb{P}\left(Q^\tau - \hat{Q}_n^\tau \geq \epsilon\right) \leq \exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(\frac{4(1+\tau)}{1-\tau} + b(1-\tau)L^2\epsilon)}\right).$$

Combining the two bounds together, we have the two-sided bound shown in the following,

$$\mathbb{P}\left(|\hat{Q}_n^\tau - Q^\tau| \geq \epsilon\right) \leq 2\exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(\alpha + \beta\epsilon)}\right),$$

where $\alpha = \frac{4(1+\tau)}{1-\tau}, \beta = \frac{4}{3}(2L + b(1-\tau)L^2)$.

Proof. By assuming $n \geq \frac{4}{1-\tau}$, we have

$$n(1-\tau) - 1 = n(1 - (\tau + \frac{1}{n})) \geq \frac{3}{4}n(1-\tau). \tag{77}$$

$$n(1-\tau) - 2 = n(1 - (\tau + \frac{2}{n})) \geq \frac{1}{2}n(1-\tau). \tag{78}$$

Recall $v^r = \frac{2}{kL^2}, v^l = \frac{2(n-k+1)}{(k-1)^2L^2}, c^r = \frac{2}{kL}, w_n = \frac{b}{n}$, we have

$$\begin{aligned}
\mathbb{P}\left(\hat{Q}_n^\tau - Q^\tau \geq \epsilon\right) &\leq \exp\left(-\frac{\epsilon^2}{2(v^r + (c^r + w_n)\epsilon)}\right) && (79)\\
&= \exp\left(-\frac{\epsilon^2}{2(\frac{2}{kL^2} + (\frac{2}{kL} + \frac{b}{n})\epsilon)}\right) && (80)\\
&\leq \exp\left(-\frac{\epsilon^2}{2(\frac{2}{kL^2} + (\frac{2}{kL} + \frac{b(1-\tau)}{k})\epsilon)}\right) & n \geq \frac{k}{1-\tau} & \quad (81)\\
&= \exp\left(-\frac{kL^2\epsilon^2}{2(2 + (2L + b(1-\tau)L^2)\epsilon)}\right) && (82)\\
&\leq \exp\left(-\frac{\frac{3}{4}n(1-\tau)L^2\epsilon^2}{2(2 + (2L + b(1-\tau)L^2)\epsilon)}\right) & k \geq \frac{3}{4}n(1-\tau) & \quad (83)\\
&= \exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(\frac{8}{3} + \frac{4}{3}(2L + b(1-\tau)L^2)\epsilon)}\right). && (84)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{P}\left(Q^\tau - \hat{Q}_n^\tau \geq \epsilon\right) &\leq \exp\left(-\frac{\epsilon^2}{2(v^l + w_n\epsilon)}\right) && (85)\\
&= \exp\left(-\frac{\epsilon^2}{2(\frac{2(n-k+1)}{(k-1)^2L^2} + \frac{b}{n}\epsilon)}\right) && (86)\\
&\leq \exp\left(-\frac{\epsilon^2}{2(\frac{(1+\tau)}{1/4n(1-\tau)^2L^2} + \frac{b}{n}\epsilon)}\right) & k - 1 \geq \frac{n(1-\tau)}{2} & \quad (87)\\
&= \exp\left(-\frac{1/4n(1-\tau)^2L^2{\epsilon}^2}{2((1+\tau) + 1/4b(1-\tau)^2L^2\epsilon)}\right) && (88)\\
&= \exp\left(-\frac{n(1-\tau)L^2{\epsilon}^2}{2(\frac{4(1+\tau)}{1-\tau} + b(1-\tau)L^2\epsilon)}\right). && (89)
\end{aligned}$$

Then let $\alpha = \max\{\frac{8}{3}, \frac{4(1+\tau)}{1-\tau}\} = \frac{4(1+\tau)}{1-\tau}, \beta = \max\{\frac{4}{3}(2L + b(1-\tau)L^2), b(1-\tau)L^2\} = \frac{4}{3}(2L + b(1-\tau)L^2))$, we have

$$\begin{aligned}\mathbb{P}\left(|\hat{Q}_n^\tau - Q^\tau| \geq \epsilon\right) &= \mathbb{P}\left(\hat{Q}_n^\tau - Q^\tau \geq \epsilon\right) + \mathbb{P}\left(Q^\tau - \hat{Q}_n^\tau \geq \epsilon\right) \\ &\leq 2\exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(\alpha+\beta\epsilon)}\right).\end{aligned}$$

□

Remark 2 (Lower bound of sample size)**.** *In Lemma 3, we make an assumption about the lower bound of sample size, i.e. $n \geq \frac{4}{1-\tau}$. Note that along with the left inequality of Equation (75), the lower bound of sample size can be expressed as $n \geq \frac{\max\{k,4\}}{1-\tau}$. When $k \geq 4$, we have $n \geq \frac{k}{1-\tau} = \frac{\lfloor n(1-\tau)\rfloor}{1-\tau}$, which holds for all $n \geq 1$. This implies that instead of making an assumption about sample size n, we could equivalently make an assumption about the rank ($k \geq 4$) or the quantile level ($\tau \leq 1 - \frac{4}{n}$ with $n \geq 4$).*
Note the constant 4 in $n \geq \frac{4}{1-\tau}$ is chosen to have a simpler expression for the concentration bounds, one can choose any constant bigger than 2 (such that the term $n(1-\tau) - 2$ is valid in Equation (78)).

We show a variant of Lemma 3 where we remove the lower bound assumption of the number of samples. The derived concentration bounds have a constant term, which does not influence the convergence rate in terms of n.

Lemma 4. *For $\epsilon > 0$, recall n denotes the number of samples, b is a constant depending on the density about τ-quantile ($0 < \tau < 1$), L is the lower bound of hazard rate. Corollary 1 can be represented as*

$$\begin{aligned}\mathbb{P}\left(\hat{Q}_n^\tau - Q^\tau \geq \epsilon\right) &\leq \exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(2+(2L+b(1-\tau)L^2)\epsilon)} + \frac{L^2\epsilon^2}{2(2+(2L+b(1-\tau)L^2)\epsilon)}\right), \\ \mathbb{P}\left(Q^\tau - \hat{Q}_n^\tau \geq \epsilon\right) &\leq \exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(\frac{2(2+\tau)}{1-\tau} + b(1-\tau)L^2\epsilon)} + \frac{L^2\epsilon^2}{2(\frac{1}{2}\frac{\tau}{1-\tau} + \frac{1}{4}b(1-\tau)L^2\epsilon)}\right).\end{aligned}$$

Combining the two bounds together, we have the two-sided bound shown in the following,

$$\mathbb{P}\left(|\hat{Q}_n^\tau - Q^\tau| \geq \epsilon\right) \leq 2\exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(\widetilde{\alpha}+\widetilde{\beta}\epsilon)} + \frac{L^2\epsilon^2}{2(\widetilde{\alpha}+\widetilde{\beta}\epsilon)}\right),$$

where $\widetilde{\alpha} = 2\frac{\tau+2}{1-\tau}, \widetilde{\beta} = 2L + b(1-\tau)L^2$.

Proof. From Eq. (82), we have

$$\begin{aligned}\mathbb{P}\left(\hat{Q}_n^\tau - Q^\tau \geq \epsilon\right) &\leq \exp\left(-\frac{kL^2\epsilon^2}{2(2+(2L+b(1-\tau)L^2)\epsilon)}\right) && (90)\\ &\leq \exp\left(-\frac{(n(1-\tau)-1)L^2\epsilon^2}{2(2+(2L+b(1-\tau)L^2)\epsilon)}\right) \qquad k \geq n(1-\tau)-1 && (91)\\ &= \exp\left(-\frac{(n(1-\tau))L^2\epsilon^2}{2(2+(2L+b(1-\tau)L^2)\epsilon)} + \frac{L^2\epsilon^2}{2(2+(2L+b(1-\tau)L^2)\epsilon)}\right). && (92)\end{aligned}$$

From Eq. (86), we have

$$\mathbb{P}\left(Q^\tau - \hat{Q}_n^\tau \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2(\frac{2(n-k+1)}{(k-1)^2L^2} + \frac{b}{n}\epsilon)}\right) \tag{93}$$

$$\leq \exp\left(-\frac{L^2\epsilon^2}{2(\frac{2(n\tau+2)}{(n(1-\tau)-2)^2} + \frac{bL^2}{n}\epsilon)}\right) \qquad k-1 \geq n(1-\tau)-2 \tag{94}$$

$$\leq \exp\left(-\frac{L^2\epsilon^2}{2(\frac{2(n\tau+2)}{(n(1-\tau))(n(1-\tau)-4)} + \frac{bL^2}{n}\epsilon)}\right) \tag{95}$$

$$= \exp\left(-\frac{(n(1-\tau))(n(1-\tau)-4)L^2\epsilon^2}{2(2(n\tau+2) + n(1-\tau)^2 bL^2\epsilon)}\right) \tag{96}$$

$$= \exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(2\frac{\tau+2/n}{1-\tau} + (1-\tau)bL^2\epsilon)} + \frac{L^2\epsilon^2}{2(\frac{1}{2}\frac{(n\tau+2)}{n(1-\tau)} + \frac{1}{4}(1-\tau)bL^2\epsilon)}\right) \tag{97}$$

$$\leq \exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(2\frac{\tau+2}{1-\tau} + b(1-\tau)L^2\epsilon)} + \frac{L^2\epsilon^2}{2(\frac{1}{2}\frac{\tau}{1-\tau} + \frac{1}{4}b(1-\tau)L^2\epsilon)}\right). \qquad n \geq 1 \to 1/n \leq 1 \tag{98}$$

Then let $\widetilde{\alpha} = 2\max\left\{1, \frac{\tau+2}{1-\tau}, \frac{1}{4}\frac{\tau}{1-\tau}\right\} = 2\frac{\tau+2}{1-\tau}$, $\widetilde{\beta} = \max\{2L + b(1-\tau)L^2, b(1-\tau)L^2, \frac{1}{4}b(1-\tau)L^2)\} = 2L + b(1-\tau)L^2$. we have

$$\mathbb{P}\left(|\hat{Q}_n^\tau - Q^\tau| \geq \epsilon\right) = \mathbb{P}\left(\hat{Q}_n^\tau - Q^\tau \geq \epsilon\right) + \mathbb{P}\left(Q^\tau - \hat{Q}_n^\tau \geq \epsilon\right) \leq 2\exp\left(-\frac{n(1-\tau)L^2\epsilon^2}{2(\widetilde{\alpha} + \widetilde{\beta}\epsilon)} + \frac{L^2\epsilon^2}{2(\widetilde{\alpha} + \widetilde{\beta}\epsilon)}\right).$$

□

Remark 3 (Constant term in concentration bound)**.** *Note the constant term* $\frac{L^2\epsilon^2}{2(\widetilde{\alpha}+\widetilde{\beta}\epsilon)}$ *in Lemma 4 is due to the floor operator of the rank* k*, as explained in Eq. (75). This constant term is a bias term coming from estimating the quantile by a single order statistic and is unavoidable without additional assumptions.*
By comparing Lemma 3 and Lemma 4 we observe that one needs to balance between the constant term, convergence rate, and assumptions to be made. For example, Lemma 3 reduces the constant term by assuming a lower bound on the sample size. On one hand, this assumption guarantees there is enough number of samples to have a more accurate estimation; on the other hand, compared with Lemma 4, Lemma 3 has a smaller convergence rate in terms of n *(larger parameters* α, β*).*

C.2. Q-SAR Error Bounds

In this section, we show the proof of Q-SAR error bounds, based on the concentration results we proposed. In Theorem 4, we show the error bound based on Lemma 3 under the assumption of lower bound of budget. In Theorem 8, we release the budget assumption, and show a variant of the error bound based on Lemma 4. The proof technique follows Bubeck et al. (2013).

Theorem 4 (Q-SAR Probability of Error Upper Bound)**.** *For the problem of identifying* m *best arms out of* K *arms, with budget* $N \geq \frac{4}{1-\tau}\overline{\log}(K) + K$*, the probability of error (Definition 2) for Q-SAR satisfies*

$$e_N \leq 2K^2 \exp\left(-\frac{N-K}{\overline{\log}(K)H^\tau}\right),$$

where problem complexity H^τ *is defined in Eq. (23).*

Proof. Recall we order the arms according to optimality as $o_1, \ldots, o_K$ s.t. $Q^\tau_{o_1} \geq \cdots \geq Q^\tau_{o_K}$. The optimal arm set of size m is $\mathcal{S}^*_m = \{o_1, \ldots, o_m\}$. In phase p, there are $K+1-p$ arms inside of the active set $\mathcal{A}_p$, we sort the arms inside of $\mathcal{A}_p$

and denote them as $\ell_1, \ell_2, \ldots, \ell_{K+1-p}$ such that $Q^\tau_{\ell_1} \geq Q^\tau_{\ell_2} \geq \cdots \geq Q^\tau_{\ell_{K+1-p}}$. If the algorithm does not make any error in the first $p-1$ phases (i.e. not reject an arm from optimal set and not accept an arm from non-optimal set), then we have

$$\{\ell_1, \ell_2, \ldots, \ell_{l_p}\} \subseteq \mathcal{S}^*_m, \quad \{\ell_{l_p+1}, \ldots, \ell_{K+1-p}\} \subseteq \mathcal{K} \backslash \mathcal{S}^*_m. \tag{99}$$

Additionally, we sort the arms in $\mathcal{A}_p$ according to the empirical quantiles at phase p as $a_{best}(= a_1), a_2, ..., a_{l_p}, a_{l_p+1}, ..., a_{worst}(= a_{K-p+1})$ such that $\hat{Q}^\tau_{a_{best}, n_p} \geq \hat{Q}^\tau_{a_2, n_p} \geq \cdots \geq \hat{Q}^\tau_{a_{worst}, n_p}$.

Consider an event ξ,

$$\xi = \{\forall i \in \{1, \ldots, K\}, p \in \{1, \ldots, K-1\}, \left| \hat{Q}^\tau_{i, n_p} - Q^\tau_i \right| < \frac{1}{4} \Delta_{(K+1-p)} \}.$$

Recall $\alpha = \frac{4(1+\tau)}{1-\tau}$, and we adapt β to β_i with i indicating the index of arm, that is, $\beta_i = \frac{4}{3}(2L_i + b_i(1-\tau)L_i^2)$. Recall the sample size for phase p is $n_p = \lceil \frac{N-K}{\overline{\log}(K)(K+1-p)} \rceil$. Based on Lemma 3 and the union bound, we derive the upper bound of probability for the complementary event $\bar{\xi}$ as

$$\begin{aligned}
\mathbb{P}(\bar{\xi}) &\leq \sum_{i=1}^{K} \sum_{p=1}^{K-1} \mathbb{P} \left(\left| \hat{Q}^\tau_{i, n_p} - Q^\tau_i \right| \geq \frac{1}{4} \Delta_{(K+1-p)} \right) && \text{union bound} \quad (100)\\
&\leq \sum_{i=1}^{K} \sum_{p=1}^{K-1} 2 \exp \left(- \frac{n_p (1-\tau) L_i^2 (\frac{1}{4}\Delta_{(K+1-p)})^2}{2(\alpha + \beta_i \frac{1}{4} \Delta_{(K+1-p)})} \right) && \text{By Lemma 3} \quad (101)\\
&\leq \sum_{i=1}^{K} \sum_{p=1}^{K-1} 2 \exp \left(- \frac{N-K}{\overline{\log}(K)(K+1-p)} \frac{1}{\frac{8}{1-\tau}(\frac{4\alpha}{L_i^2 \Delta^2_{(K+1-p)}} + \frac{\beta_i}{L_i^2 \Delta_{(K+1-p)}})} \right) && (102)\\
&\leq 2K^2 \exp \left(- \frac{N-K}{\overline{\log}(K) H^\tau} \right). && (103)
\end{aligned}$$

where $H^\tau = \max_{\{i,j \in \mathcal{K}\}} \frac{8j}{1-\tau} (\frac{4\alpha}{L_i^2 \Delta^2_{(j)}} + \frac{\beta_i}{L_i^2 \Delta_{(j)}})$.

Note that we have assumed the number of samples of each arm is at least $\frac{4}{1-\tau}$ in Lemma 3. That is, $Kn_1 = K \lceil \frac{N-K}{\overline{\log}(K)K} \rceil \geq \frac{4}{1-\tau}$, which gives $N \geq \frac{4}{1-\tau} \overline{\log}(K) + K$. This means, the bound derived above holds when we have budget N no less than $\frac{4}{1-\tau} \overline{\log}(K) + K$.

We show that on event ξ, Q-SAR algorithm does not make any error by induction on phases. Assume that the algorithm does not make any error on the first $p-1$ phases, i.e. does not reject an arm from optimal set and not accept an arm from non-optimal set. Then in the following, we show the algorithm does not make an error on the p^{th} phase. We discuss in terms of two cases:

Case 1: If an arm ℓ_j is accepted, then $\ell_j \in \mathcal{S}^*_m$.
We prove by contradiction. Assume arm ℓ_j is accepted in phase p, but $\ell_j \notin \mathcal{S}^*_m$, i.e. $Q^\tau_{\ell_j} \leq Q^\tau_{\ell_{l_p+1}} \leq Q^\tau_{o_{m+1}}$. According to Algorithm 1, arm ℓ_j is accepted only if its empirical quantile is the maximum among all active arms in phase p, thus $\hat{Q}^\tau_{\ell_j, n_p} \geq \hat{Q}^\tau_{\ell_1, n_p}$. On event ξ, we have

$$Q^\tau_{\ell_j} + \frac{1}{4}\Delta_{(K+1-p)} > \hat{Q}^\tau_{\ell_j, n_p} \geq \hat{Q}^\tau_{\ell_1, n_p} > Q^\tau_{\ell_1} - \frac{1}{4}\Delta_{(K+1-p)} \tag{104}$$

$$\Rightarrow \Delta_{(K+1-p)} > \frac{1}{2}\Delta_{(K+1-p)} > Q^\tau_{\ell_1} - Q^\tau_{\ell_j} \geq Q^\tau_{\ell_1} - Q^\tau_{o_{m+1}}. \tag{105}$$

Another requirement to accept ℓ_j is $\widehat{\Delta}_{best} > \widehat{\Delta}_{worst}$, that is,

$$\hat{Q}^\tau_{\ell_j, n_p} - \hat{Q}^\tau_{a_{l_p+1}, n_p} > \hat{Q}^\tau_{a_{l_p}, n_p} - \hat{Q}^\tau_{a_{K+1-p}, n_p}. \tag{106}$$

In the following, we will connect Eq. (106) with the corresponding population quantiles on event ξ. We first connect $\hat{Q}^\tau_{a_{K+1-p}, n_p}$ and $Q^\tau_{\ell_{K+1-p}}$. Since $\hat{Q}^\tau_{a_{K+1-p}, n_p}$ is the minimum empirical quantile at phase p,

$$\hat{Q}^\tau_{a_{K+1-p}, n_p} \leq \hat{Q}^\tau_{\ell_{K+1-p}, n_p} < Q^\tau_{\ell_{K+1-p}} + \frac{1}{4}\Delta_{(K+1-p)}. \tag{107}$$

We then connect $\hat{Q}^{\tau}_{a_{l_p+1},n_p}, \hat{Q}^{\tau}_{a_{l_p},n_p}$ to $Q^{\tau}_{o_m}$. On event ξ, for all $i \leq l_p$,

$$\hat{Q}^{\tau}_{\ell_i,n_p} > Q^{\tau}_{\ell_i} - \frac{1}{4}\Delta_{(K+1-p)} \geq Q^{\tau}_{\ell_{l_p}} - \frac{1}{4}\Delta_{(K+1-p)} \geq Q^{\tau}_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}, \tag{108}$$

which means there are l_p arms in active set, i.e. $\{\ell_1, \ell_2, \dots, \ell_{l_p}\}$, having empirical quantiles bigger or equal than $Q^{\tau}_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}$. Additionally, although $j > l_p$, ℓ_j has the maximum empirical quantile, which is bigger than $Q^{\tau}_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}$ as well. So in total there are $l_p + 1$ arms having empirical quantiles bigger or equal to $Q^{\tau}_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}$, i.e.

$$\hat{Q}^{\tau}_{a_{l_p},n_p} \geq \hat{Q}^{\tau}_{a_{l_p+1},n_p} \geq Q^{\tau}_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}. \tag{109}$$

Combine Eq. (106)(107)(109) together, we have

$$(Q^{\tau}_{\ell_j} + \frac{1}{4}\Delta_{(K+1-p)}) - (Q^{\tau}_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}) > (Q^{\tau}_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}) - (Q^{\tau}_{\ell_{K+1-p}} + \frac{1}{4}\Delta_{(K+1-p)}) \tag{110}$$

$$\Rightarrow \Delta_{(K+1-p)} > 2Q^{\tau}_{o_m} - (Q^{\tau}_{\ell_j} + Q^{\tau}_{\ell_{K+1-p}}) > Q^{\tau}_{o_m} - Q^{\tau}_{\ell_{K+1-p}}. \tag{111}$$

From Eq. (105)(111), we have $\Delta_{(K+1-p)} > \max\{Q^{\tau}_{\ell_1} - Q^{\tau}_{o_{m+1}}, Q^{\tau}_{o_m} - Q^{\tau}_{\ell_{K+1-p}}\}$, which contradicts the fact that $\Delta_{(K+1-p)} \leq \max\{Q^{\tau}_{\ell_1} - Q^{\tau}_{o_{m+1}}, Q^{\tau}_{o_m} - Q^{\tau}_{\ell_{K+1-p}}\}$, since at phase p, there are only $p-1$ arms have been accepted or rejected. So we have if an arm ℓ_j is accepted, then $\ell_j \in \mathcal{S}^*_m$, which finishes the proof of Case 1.

Case 2: If an arm ℓ_j is rejected, the $\ell_j \notin \mathcal{S}^*_m$.

The proof of Case 2 is similar to the proof of Case 1. We prove by contradiction. Assume arm ℓ_j is rejected in phase p, but $\ell_j \in \mathcal{S}^*_m$, i.e. $Q^{\tau}_{\ell_j} \geq Q^{\tau}_{\ell_{l_p}} \geq Q^{\tau}_{o_m}$. According to Algorithm 1, arm ℓ_j is rejected only if its empirical quantile is the minimum among all active arms in phase p, thus $\hat{Q}^{\tau}_{\ell_j,n_p} \leq \hat{Q}^{\tau}_{\ell_{K+1-p},n_p}$. On event ξ, we have

$$Q^{\tau}_{\ell_j} - \frac{1}{4}\Delta_{(K+1-p)} < \hat{Q}^{\tau}_{\ell_j,n_p} \leq \hat{Q}^{\tau}_{\ell_{K+1-p},n_p} < Q^{\tau}_{\ell_{K+1-p}} + \frac{1}{4}\Delta_{(K+1-p)} \tag{112}$$

$$\Rightarrow \Delta_{(K+1-p)} > \frac{1}{2}\Delta_{(K+1-p)} > Q^{\tau}_{\ell_j} - Q^{\tau}_{\ell_{K+1-p}} \geq Q^{\tau}_{o_m} - Q^{\tau}_{\ell_{K+1-p}}. \tag{113}$$

Another requirement to accept ℓ_j is $\widehat{\Delta}_{best} \leq \widehat{\Delta}_{worst}$, i.e.

$$\hat{Q}^{\tau}_{a_1,n_p} - \hat{Q}^{\tau}_{a_{l_p+1},n_p} \leq \hat{Q}^{\tau}_{a_{l_p},n_p} - \hat{Q}^{\tau}_{\ell_j,n_p}. \tag{114}$$

In the following, we will connect Eq. (114) with the corresponding population quantiles on event ξ. We first connect $\hat{Q}^{\tau}_{a_1,n_p}$ and $Q^{\tau}_{\ell_1}$. Since $\hat{Q}^{\tau}_{a_1,n_p}$ is the maximum empirical quantile at phase p,

$$\hat{Q}^{\tau}_{a_1,n_p} \geq \hat{Q}^{\tau}_{\ell_1,n_p} > Q^{\tau}_{\ell_1} - \frac{1}{4}\Delta_{(K+1-p)}. \tag{115}$$

We then connect $\hat{Q}^{\tau}_{a_{l_p+1},n_p}, \hat{Q}^{\tau}_{a_{l_p},n_p}$ to $Q^{\tau}_{o_{m+1}}$. On event ξ, for all $i \geq l_p + 1$,

$$\hat{Q}^{\tau}_{\ell_i,n_p} < Q^{\tau}_{\ell_i} + \frac{1}{4}\Delta_{(K+1-p)} \leq Q^{\tau}_{\ell_{l_p+1}} + \frac{1}{4}\Delta_{(K+1-p)} \leq Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}, \tag{116}$$

Additionally, although $j < l_p + 1$, ℓ_j has the minimum empirical quantile, which is smaller than $Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}$ as well. So that,

$$\hat{Q}^{\tau}_{a_{l_p+1},n_p} \leq \hat{Q}^{\tau}_{a_{l_p},n_p} \leq Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}. \tag{117}$$

Combining Eq. (114), (115) and (117) together, we have

$$(Q^{\tau}_{\ell_1} - \frac{1}{4}\Delta_{(K+1-p)}) - (Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}) \leq (Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}) - (Q^{\tau}_{\ell_j} - \frac{1}{4}\Delta_{(K+1-p)}) \tag{118}$$

$$\Rightarrow \Delta_{(K+1-p)} \geq (Q^{\tau}_{\ell_j} + Q^{\tau}_{\ell_1}) - 2Q^{\tau}_{o_{m+1}} > Q^{\tau}_{\ell_1} - Q^{\tau}_{o_{m+1}}. \tag{119}$$

From Eq. (113)(119), we have $\Delta_{(K+1-p)} > \max\{Q^\tau_{\ell_1} - Q^\tau_{o_{m+1}}, Q^\tau_{o_m} - Q^\tau_{\ell_{K+1-p}}\}$, which contradicts the fact that $\Delta_{(K+1-p)} \leq \max\{Q^\tau_{\ell_1} - Q^\tau_{o_{m+1}}, Q^\tau_{o_m} - Q^\tau_{\ell_{K+1-p}}\}$, since at phase p, there are only $p-1$ arms have been accepted or rejected. So we have if an arm ℓ_j is rejected, then $\ell_j \notin \mathcal{S}^*_m$, which finishes the proof of Case 2. □

Now we show a variant of Theorem 4, using the result of Lemma 4. Define a slightly different problem complexity $\widetilde{H}^\tau$, which has the same form of H^τ in Eq. (23) with smaller parameters $\widetilde{\alpha}$ and $\widetilde{\beta}$.

$$\widetilde{H}^\tau := \max_{i,j \in \mathcal{K}} \frac{8j}{1-\tau}\left(\frac{4\widetilde{\alpha}}{L_i^2 \Delta_{(j)}^2} + \frac{\widetilde{\beta}_i}{L_i^2 \Delta_{(j)}}\right), \tag{120}$$

where $\widetilde{\alpha} = 2\frac{\tau+2}{1-\tau}, \widetilde{\beta} = 2L_i + (1-\tau) b_i L_i^2$.

Theorem 8 (Q-SAR Probability of Error Upper Bound Variant)**.** *For the problem of identifying m best arms out of K arms, the probability of error (Definition 2) for Q-SAR satisfies*

$$e_N \leq 2K^2 \exp\left(-\frac{N-K}{\overline{\log}(K)\widetilde{H}^\tau} + C\right),$$

where problem complexity variant $\widetilde{H}^\tau$ is defined in Eq. (120), and constant $C = \max_{\{i,j\in\mathcal{K}\}} \frac{L_i^2 \Delta_{(j)}^2}{8(4\widetilde{\alpha} + \widetilde{\beta}_i \Delta_{(j)})}$.

Proof. The only difference of the proof of Theorem 4 is we derive the bound of of $\mathbb{P}(\bar{\xi})$ (See Eq. (24) for the definition of event ξ) based on a Lemma 4, and we do not have the lower bound assumption made for budget.

$$\begin{aligned}
\mathbb{P}(\bar{\xi}) &\leq \sum_{i=1}^{K}\sum_{p=1}^{K-1} \mathbb{P}\left(\left|\hat{Q}^\tau_{i,n_p} - Q^\tau_i\right| \geq \frac{1}{4}\Delta_{(K+1-p)}\right) && \text{union bound} \quad (121)\\
&\leq \sum_{i=1}^{K}\sum_{p=1}^{K-1} \exp\left(-\frac{n_p(1-\tau)L_i^2(\frac{1}{4}\Delta_{(K+1-p)})^2}{2(\widetilde{\alpha} + \widetilde{\beta}_i\frac{1}{4}\Delta_{(K+1-p)})} + \frac{L_i^2(\frac{1}{4}\Delta_{(K+1-p)})^2}{2(\widetilde{\alpha} + \widetilde{\beta}_i\frac{1}{4}\Delta_{(K+1-p)})}\right) && \text{By Lemma 4} \quad (122)\\
&\leq 2K^2 \exp\left(-\frac{N-K}{\overline{\log}(K)\widetilde{H}^\tau} + C\right), && (123)
\end{aligned}$$

where $C = \max_{\{i,j\in\mathcal{K}\}} \frac{L_i^2 \Delta_{(j)}^2}{8(4\widetilde{\alpha} + \widetilde{\beta}_i \Delta_{(j)})}$.

Then we conclude the proof by following the same reasoning in the proof of Theorem 4. □

D. Discussion

Quantile Estimation Complexity: This paper focuses on how quantiles provide a different way to summarise the distribution of each arm. Quantiles are interesting and useful summary statistics for risk-averse decision-making, but estimating quantiles may be more expensive than estimating the mean. We provide the time complexity of our algorithms (for K arms) in the following. Estimating quantiles needs binary search in each round when we get new samples. For Q-SAR, in each phase $p \in [1, K-1]$, the time complexity is $\mathcal{O}(K\log(n_p - n_{p-1}) + K\log K) = \mathcal{O}(K\log(N/K^2) + K\log K)$. Combining for all $K-1$ phases, the time complexity is $\mathcal{O}(K^2\log(N/K^2) + K^2\log K)$. For space complexity, one needs to save the samples for each arm and also updates information (quantiles) for each arm, so the space complexity is $\mathcal{O}(N+K)$ for both algorithms. One could save time and space for estimating quantiles by using online algorithms. For example, instead of performing binary search from scratch, one can retain an estimate of the quantile and update the estimate given the new samples. This is the key idea of online algorithms such as stochastic gradient descent. Such approaches (and their analysis) is beyond the scope of this paper.

Understanding IHR Distributions: The hazard rate of random variables provide an useful way to think about real phenomena. For example, let the random variable X denote the age of a car when it has a serious engine problem for the first time. One would expect the hazard rate increases over time. If the random variable X denotes the time before you win a lottery, then the hazard rate would be approximately constant.
Some examples of general distributions with IHR include:

• Gamma distribution with two parameters $\lambda > 0, \alpha > 0$, with p.d.f. $f(x) = \frac{\lambda^\alpha x^{\alpha-1} \exp\{-\lambda x\}}{\Gamma(\alpha)}$ for $x > 0$, where the gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp\{-x\} dx$. When $\alpha \geq 1$, the hazard rate is non-decreasing. The case $\alpha = 1$ corresponds to the exponential distribution (Definition 4) and the hazard rate is constant.
• Weibull Distribution with two parameters $\lambda > 0, p > 0$, with p.d.f. $f(x) = p\lambda^p x^{p-1} \exp\{-(\lambda x)^p\}$ for $x > 0$. When $p \geq 1$, the hazard rate is non-decreasing. The Weibull distribution reduces to the exponential distribution (Definition 4) when $p = 1$.
• Absolute Gaussian distribution (Definition 3). The lower bound of hazard rate for the centered Absolute Gaussian distribution is $\frac{1}{\sigma\sqrt{2\pi}}$.

Recall that the IHR assumption allows us to consider distributions of unbounded rewards. It does so by constraining the tails of the density. The random variable with IHR is light-tailed, i.e. having tails the same as or lighter than an exponential distribution. Light-tailed distributions include a wide range of distributions, including sub-gamma and sub-Gaussian distributions.

Estimate Lower Bound of Hazard Rate and Concentration Inequality In practice, one can estimate the lower bound of hazard rate L by estimating the p.d.f. and c.d.f. around 0 (with non-negative support and IHR assumption). So one can design a UCB-type of algorithms by adaptively estimating L. But in practice, introducing new variables to estimate would influence the stability of the algorithm.