Supplementary Materials for Quantile Bandits for Best Arm Identification with Concentration Inequalities

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

A Supplementary Experiments

A.1 Vaccine Simulation

We consider the problem of identifying optimal strategies for allocating influenza vaccine. Following Libin et al. (2017), we format this problem as an instance of the BAI where each vaccine allocation strategy is an arm. We consider allocating 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), and denote the allocation scheme as a Boolean 5-tuple, with each position corresponds to the respective age group (0 represents no allocation; 1 represents allocation). The reward of a strategy is defined as the proportion of individuals that did not experience symptomatic infection. We generate 1000 rewards for each strategy by simulating the epidemic for 180 days using FluTE [1] (with basic reproduction number $R0 = 1.3$).

The violin plot of reward samples is shown in Figure 4. Note there are also some rewards close to 1, which is because that the pathogen is not able to establish an epidemic for certain simulation runs (Libin et al., 2017) and does not reflect the efficacy of the vaccine. We call those rewards as "outliers". The $0.5-$quantile (i.e. the median) is more robust to those outliers compared with the mean. We consider the task of identifying $m = 1$ (optimal arm: $< 0, 1, 0, 0, 0 >$, i.e. only allocation to 5-18 years old group) and $m = 3$ (optimal arms: $< 0, 1, 0, 0, 0 >, < 1, 1, 0, 0, 0 >$ and $< 0, 1, 1, 0, 0 >$) arms with highest $0.5-$quantile values,

out of totally $K = 32$ arms with fixed budget.

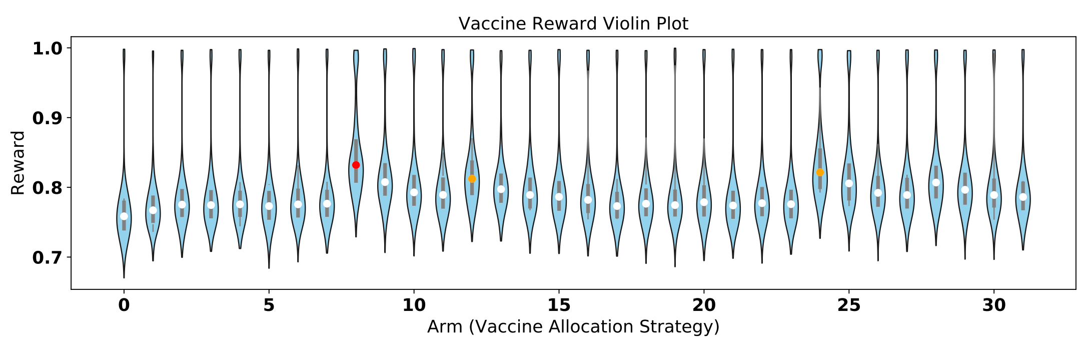

Figure 4: Vaccine Reward Violin Plot. The X-axis represents arms (vaccine allocation strategies), Y-axis represents rewards, which is the proportion of individuals that did not experience symptomatic infection. The circle in each violin represents the median, where the red one is the highest and the orange ones are the second and the third highest. The black line in each violin shows the range of 0.25-quantile to 0.75-quantile.

The performance for BAI task is shown in Figure 5. We compare our algorithm with the quantile-based algorithm only, since $0.5-$quantile introduces different optimality order of arms compared with the mean. Although estimated hazard rates from the reward samples are not strictly under our IHR assumption (Assumption 2), empirical evidence shows Q-SAR is still the best for multiple best arms identification (when $m = 3$). This leaves the performance analysis for Q-SAR without IHR assumption to future work.

A.2 Stochastic Simulation

We provide more details about the environments setting in Section 6. We consider two distributions which satisfy our assumptions: absolute Gaussian distribution (Definition 3), and exponential distribution (Definition 4).

[1] https://github.com/dlchao/FluTE

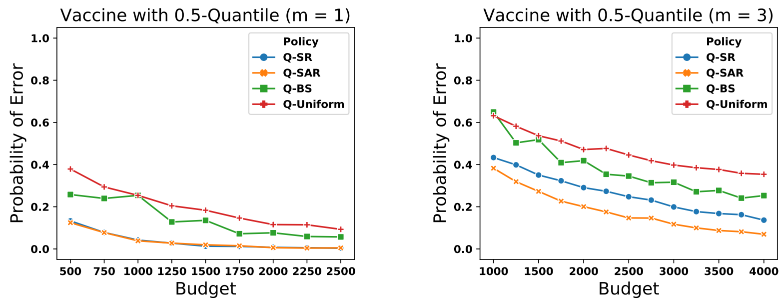

Figure 5: Vaccine BAI Experiments (with 5000 independent runs).

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 the p.d.f $\tilde{\phi}$ and c.d.f $\tilde{\Phi}$,*

$$\tilde{\phi} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x+\mu)^2}{2\sigma^2}}, \tag{21}$$

$$\tilde{\Phi} = \frac{1}{2}\left[\operatorname{erf}\left(\frac{x+\mu}{\sigma\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right], \tag{22}$$

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 $|\mathcal{N}(\mu, \sigma^2)|$.

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

$$f_{Exp}(x, \theta) = \theta e^{-\theta x}, \tag{23}$$

$$F_{Exp}(x, \theta) = 1 - e^{-\theta x}, \tag{24}$$

We denote the exponential distribution with θ as $Exp(\theta)$.

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.

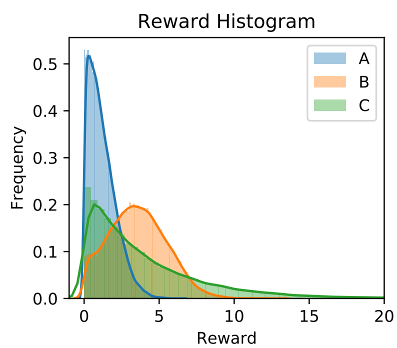

Figure 6: Simulated Arm Rewards Histogram.

A.3 Q-SR

We provide justifications for the design choice of our proposed Q-SR algorithm shown in Algorithm 2. 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 $n_1 = \left\lceil \frac{1}{\overline{\log}(K)} \frac{N-K}{K} \right\rceil$ times, one is pulled $n_2 = \left\lceil \frac{1}{\overline{\log}(K)} \frac{N-K}{K-1} \right\rceil$ times, …, $m+1$ is pulled $n_{K-m} = \left\lceil \frac{1}{\overline{\log}(K)} \frac{N-K}{K+1-(K-m)} \right\rceil$ times, then

$$n_1 + \dots + (m+1)n_{K-m} \leq K + \frac{N-K}{\overline{\log}(K)} \left(\frac{m}{m+1} + \sum_{p=1}^{K-m} \frac{1}{K+1-p} \right) = N. \tag{25}$$

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 future work.

B Concentration Inequality Proof

This section shows the proof of the concentration results shown in Section 5. 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. (13)) 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 $\lambda \in \mathbb{R}$,*

$$\begin{aligned} \operatorname{Ent}[\exp(\lambda W)] &= \lambda \mathbb{E}[W \exp(\lambda W)] - \mathbb{E}[\exp(\lambda W)] \log \mathbb{E}[\exp(\lambda W)] && (26) \\ &\leq \sum_{i=1}^{n} \mathbb{E}[\exp(\lambda W) \phi(-\lambda(W - W_i))] && (27) \end{aligned}$$

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 $k \in [1, n)$,*

$$\operatorname{Ent}\left[\exp(\lambda X_{(k)})\right] \leq k \mathbb{E}\left[\exp(\lambda X_{(k+1)}) \zeta(\lambda S_k)\right]. \tag{17}$$

For $k \in (1, n]$, $\operatorname{Ent}\left[\exp(-\lambda X_{(k)})\right] \leq$

$$(n - k + 1)\mathbb{E}\left[\exp(-\lambda X_{(k)}) \phi(-\lambda S_{k-1})\right]. \tag{18}$$

Proof. We prove the upper bound based on Theorem 5. We first prove Eq. (17). 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

$$\begin{aligned} &\operatorname{Ent}\left[\exp(\lambda X_{(k)})\right] \\ \leq &\mathbb{E}\Big[\sum_{i=1}^{n} \exp(\lambda X_{(k)}) \phi\big(-\lambda(X_{(k)} - X_{(k+1)} \mathbb{I}(X_i \geq X_{(k)}) - X_{(k)} \mathbb{I}(X_i < X_{(k)}))\big)\Big] && \text{Theorem 5} && (28) \\ = &\mathbb{E}\Big[\exp(\lambda X_{(k)}) \phi\big(-\lambda(X_{(k)} - X_{(k+1)})\big) \sum_{i=1}^{n} \mathbb{I}(X_i \geq X_{(k)})\Big] && && (29) \\ = &k \mathbb{E}\left[\exp(\lambda X_{(k)}) \phi(-\lambda S_k)\right] && && (30) \\ = &k \mathbb{E}\left[\exp(\lambda X_{(k+1)}) \exp(\lambda S_k) \phi(-\lambda S_k)\right] && && (31) \\ = &k \mathbb{E}\left[\exp(\lambda X_{(k+1)}) \zeta(\lambda S_k)\right] && \zeta(x) = \exp(x)\phi(-x) && (32) \end{aligned}$$

Similarly, for the proof of Eq. (18), 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 $\operatorname{Ent}\left[\exp(-\lambda X_{(k)})\right] \leq (n - k + 1)\mathbb{E}\left[\exp(-\lambda X_{(k)}) \phi(-\lambda S_{k-1})\right]$.

$\square$

Compared with the proof in Boucheron and 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 $\mathrm{Ent}\left[\exp(\lambda X_{(k)})\right]$ and $\mathrm{Ent}\left[\exp(-\lambda X_{(k)})\right]$, which allows us to derive two-sided concentration bounds instead of one-sided boundS.

From Proposition 1, and by normalising entropy as shown in Eq. (15), we derive an upper bound of the logarithmic moment generating function of $Z_k = X_{(k)} - \mathbb{E}[X_{(k)}]$ and $Z'_k = \mathbb{E}[X_{(k)}] - X_{(k)}$. The upper bounds are expressed in terms of the spacings.

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

$$\psi_{Z_k}(\lambda) \leq \lambda \frac{k}{2} \mathbb{E}\left[S_k\left(\exp(\lambda S_k) - 1\right)\right]. \tag{19}$$

For $k \in (1, n]$,

$$\psi_{Z'_k}(\lambda) \leq \frac{\lambda^2(n-k+1)}{2} \mathbb{E}[S_{k-1}^2]. \tag{20}$$

Proof. The proof of Eq. (19) is based on Proposition 1 and follows the same reasoning from Boucheron and Thomas (2012) Theorem 2.9. Note since Eq. (17) holds for $k \in [1, n)$, Eq. (19) can be proved for $k \in [1, n)$ (Boucheron and Thomas (2012) only proved for $k \in [1, n/2]$).

We now prove Eq. (20). 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 $\lambda \geq 0$,

$$\begin{aligned}
\mathrm{Ent}\left[\exp(-\lambda X_{(k)})\right] &\leq (n-k+1)\mathbb{E}\left[\exp(-\lambda X_{(k)})\phi\left(-\lambda S_{k-1}\right)\right] && \text{By Proposition 1} && (33)\\
&\leq (n-k+1)\mathbb{E}\left[\exp(-\lambda X_{(k)})\right]\mathbb{E}\left[\phi\left(-\lambda S_{k-1}\right)\right] && \text{By Proposition 4} && (34)
\end{aligned}$$

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

$$\mathrm{Ent}[\exp(\lambda Z'_k)] \leq (n-k+1)\mathbb{E}[\exp(\lambda Z'_k)]\mathbb{E}[\phi(-\lambda S_{k-1})]. \tag{35}$$

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

$$\frac{\mathrm{Ent}\left[\exp(\lambda Z'_k)\right]}{\lambda^2 \mathbb{E}\left[\exp(\lambda Z'_k)\right]} \leq \frac{n-k+1}{\lambda^2}\mathbb{E}[\phi(-\lambda S_{k-1})] \leq \frac{n-k+1}{2}\mathbb{E}[S_{k-1}^2]. \tag{36}$$

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

$$\int_0^\lambda \frac{\mathrm{Ent}\left[\exp(tZ'_k)\right]}{t^2 \mathbb{E}\left[\exp(tZ'_k)\right]} dt = \int_0^\lambda \frac{\mathbb{E}[tZ'_k] - \log \mathbb{E}[\exp(tZ'_k)]}{t^2} dt = \frac{\log \mathbb{E}[\exp(tZ'_k)]}{t}\Big|_0^\lambda = \frac{1}{\lambda}\log \mathbb{E}[\exp(\lambda Z'_k)]. \tag{37}$$

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

$$\int_0^\lambda \frac{n-k+1}{2}\mathbb{E}[S_{k-1}^2] dt = \frac{\lambda(n-k+1)}{2}\mathbb{E}[S_{k-1}^2]. \tag{38}$$

Combining Eq. (36), (37) and (38), we get

$$\psi_{Z'_k}(\lambda) = \log \mathbb{E}[\exp(\lambda Z'_k)] \leq \frac{\lambda^2(n-k+1)}{2}\mathbb{E}[S_{k-1}^2]. \tag{39}$$

which concludes the proof.

$\square$

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. $F^{\leftarrow}(F_{exp}(t))$.

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 and 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*

$$\left(Y_{(n)},\ldots,Y_{(k)},\ldots,Y_{(1)}\right) \stackrel{d}{=} \left(\frac{E_n}{n},\ldots,\sum_{i=k}^{n}\frac{E_i}{i},\ldots,\sum_{i=1}^{n}\frac{E_i}{i}\right), \tag{40}$$

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

$$\left(X_{(n)}, \ldots, X_{(1)}\right) \stackrel{d}{=} \left(R\left(Y_{(n)}\right), \ldots, R\left(Y_{(1)}\right)\right), \tag{41}$$

where $R(\cdot)$ 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 and Thomas, 2012))**.** *Let F be an absolutely continuous distribution function with hazard rate h, the derivative of R-transform is $R' = 1/h(R)$. Then if the hazard rate h is non-decreasing, 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 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 for proving of Theorem 3.

Theorem 7 (Harris' inequality (Boucheron et al., 2013))**.** *Let $X_1, \ldots, X_n$ be independent real-valued random variables and define the random vector $X = (X_1, \ldots, X_n)$ 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,*

$$\mathbb{E}\left[f_1(X_{(k)})f_2\left(S_{k-1}\right)\right] \leq \mathbb{E}\left[f_1(X_{(k)})\right]\mathbb{E}\left[f_2\left(S_{k-1}\right)\right]. \tag{42}$$

Proof. From Definition 5 and Theorem 6, let $Y_{(1)}, \ldots, Y_{(n)}$ 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.

$$\begin{aligned}
\mathbb{E}\left[f_1\left(X_{(k)}\right)f_2\left(S_{k-1}\right)\right] =& \mathbb{E}\left[f_1(R(Y_{(k)}))f_2\left(R(Y_{(k-1)}) - R(Y_{(k)})\right)\right] && (43)\\
=& \mathbb{E}\left[\mathbb{E}\left[f_1(R(Y_{(k)}))f_2\left(R(E_{k-1} + Y_{(k)}) - R(Y_{(k)})\right)|Y_{(k)}\right]\right] && (44)\\
=& \mathbb{E}\left[f_1(R(Y_{(k)}))\mathbb{E}\left[f_2\left(R(E_{k-1} + Y_{(k)}) - R(Y_{(k)})\right)|Y_{(k)}\right]\right]. && (45)
\end{aligned}$$

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.

$$\mathbb{E}\left[f_2\left(R(E_{k-1} + Y_{(k)}) - R(Y_{(k)})\right)|Y_{(k)}\right] = \int_0^{\infty} e^{-x} f_2(R(\frac{x}{k-1} + Y_{(k)}) - R(Y_{(k)}))dx. \tag{46}$$

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,

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

□

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. *Let $S_k = X_{(k)} - X_{(k+1)}$ be the k rank order statistics spacing, and L be the lower bound of hazard rate of random variable X. For any $k \in [1, n)$, the expectation of spacing S_k can be bounded under Assumption 2, $\mathbb{E}[S_k] \leq \frac{1}{kL}$.*

Proof. We show the upper bound the expectations of the k^{th} spacing of order statistics, assuming the lower bound hazard rate is L,

$$\begin{aligned}
\mathbb{E}[S_k] &= \mathbb{E}[X_{(k)} - X_{(k+1)}] \\
&= \mathbb{E}[R\left(Y_{(k+1)} + \frac{E_k}{k}\right) - R\left(Y_{(k+1)}\right)] && \text{By Theorem 6} && (50)\\
&= \int_Y \int_E \left(R\left(y+\frac{z}{k}\right) - R\left(y\right)\right) f_Y\left(y\right) f_E\left(z\right) \mathrm{d}z\mathrm{d}y && && (51)\\
&\leq \int_Y \int_E \frac{z}{k \times h\left(R\left(y\right)\right)} f_Y\left(y\right) f_E\left(z\right) \mathrm{d}z\mathrm{d}y && \text{By Proposition 3} && (52)\\
&\leq \int_E \frac{z}{kL} f_E\left(z\right) \mathrm{d}z = \frac{1}{kL} && L \text{ is the lower bound of the hazard rate.} && (53)
\end{aligned}$$

□

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 $v^r = \frac{2}{kL^2}$, $c^r = \frac{2}{kL}$, where L is lower bound of hazard rate. Under Assumption 1 and 2 , for all $\lambda \in [0, 1/c^r)$, and all $k \in [1, n)$, we have*

$$\log \mathbb{E}[\exp\left(\lambda\left(X_{(k)} - \mathbb{E}[X_{(k)}]\right)\right)] \leq \frac{\lambda^2 v^r}{2(1-c^r\lambda)}. \tag{7}$$

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

$$\mathbb{P}\left(X_{(k)} - \mathbb{E}[X_{(k)}] \geq \sqrt{2v^r\gamma} + c^r\gamma\right) \leq \exp(-\gamma). \tag{8}$$

Proof. We first prove Eq. (7). 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 $Y_{(k+1)}$.

$$\psi_{Z_k}(\lambda) \leq \lambda \frac{k}{2} \mathbb{E}\left[S_k\left(\exp(\lambda S_k) - 1\right)\right] \qquad \text{By Theorem 3} \tag{54}$$

$$\leq \lambda \frac{k}{2} \int_E \int_Y \frac{z}{h\left(R\left(y\right)\right)k}\left(\exp(\frac{\lambda z}{h\left(R\left(y\right)\right)k}) - 1\right) f_Y\left(y\right) f_E\left(z\right) \mathrm{d}y\mathrm{d}z \qquad \text{By Proposition 3} \tag{55}$$

$$\leq \frac{k}{2} \int_E \frac{\lambda}{Lk} z\left(\exp(\frac{\lambda}{Lk} z) - 1\right) f_E\left(z\right) \mathrm{d}z \tag{56}$$

$$= \frac{k}{2} \int_0^\infty \frac{\lambda}{Lk} z\left(\exp(\frac{\lambda}{Lk} z) - 1\right) \exp(-z) \mathrm{d}z \tag{57}$$

$$\leq \frac{\lambda^2 v^r}{2(1 - c^r\lambda)}, \qquad \text{With } v^r = \frac{2}{kL^2}, c^r = \frac{2}{kL} \tag{58}$$

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. (7) to Eq. (8), 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 $\lambda \geq 0$,

$$\mathbb{P}(Z_k \geq \varepsilon) \leq \exp(-\lambda\varepsilon)\mathbb{E}[\exp(\lambda Z_k)]. \tag{59}$$

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

$$\psi^*_{Z_k}(t) = \sup_{\lambda \in (0, 1/c^r)}(\varepsilon\lambda - \frac{\lambda^2 v^r}{2(1 - c^r\lambda)}) = \frac{v^r}{(c^r)^2} h_1(\frac{c^r\varepsilon}{v^r}) \tag{60}$$

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. (8) is thus proved. □

Lemma 2 (Left Tail Concentration Bounds for Order Statistics)**.** *Define* $v^l = \frac{2(n-k+1)}{(k-1)^2 L^2}$, *where* L *is lower bound of hazard rate. Under Assumption 1 and 2, for all* $\lambda \geq 0$, *and all* $k \in (1, n]$, *we have*

$$\log \mathbb{E}[\exp\left(\lambda\left(\mathbb{E}[X_{(k)}] - X_{(k)}\right)\right)] \leq \frac{\lambda^2 v^l}{2}. \tag{9}$$

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

$$\mathbb{P}\left(\mathbb{E}[X_{(k)}] - X_{(k)} \geq \sqrt{2v^l\gamma}\right) \leq \exp(-\gamma). \tag{10}$$

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 $Y_{(k)}$.

$$\psi_{Z'_k}(\lambda) \leq \frac{\lambda^2(n-k+1)}{2}\mathbb{E}[S^2_{k-1}]. \qquad \text{By Theorem 3} \tag{61}$$

$$\leq \frac{\lambda^2(n-k+1)}{2} \int_Y \int_E \left(\frac{z}{(k-1) \times h\left(R\left(y\right)\right)}\right)^2 f_Y\left(y\right) f_E\left(z\right) \mathrm{d}z\mathrm{d}y \qquad \text{By Proposition 3} \tag{62}$$

$$\leq \frac{\lambda^2(n-k+1)}{2} \int_0^\infty \left(\frac{z}{(k-1)L}\right)^2 \exp(-z)\mathrm{d}z \tag{63}$$

$$\leq \frac{\lambda^2 v^l}{2} \qquad \text{With } v^l = \frac{2(n-k+1)}{(k-1)^2 L^2}. \tag{64}$$

Eq. (9) is proved. Follow the Cramér-Chernoff method described above, we can prove Eq. (10). □

The 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. took the results from Boucheron and 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 4 , which shows connection between expected order statistics and quantiles.

Theorem 2 (Two-side Concentration Inequality for Quantiles)**.** *Define* $v^r = \frac{2}{kL^2}$, $v^l = \frac{2(n-k+1)}{(k-1)^2 L^2}$, $c^r = \frac{2}{kL}, w_n = \frac{b}{n}$, *where* L *is lower bound of hazard rate,* $b > 0$ *is a constant. For quantile level* $\tau \in (0,1)$, *define rank* $k = \lfloor n(1-\tau) \rfloor$. *Under Assumption 1 and 2, for* $n \geq \frac{2}{1-\tau}$ *and* $\gamma \geq 0$, *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 4, we have $\mathbb{E}[X_{(k)}] \leq Q^\tau + w_n$. With probability at least $1 - \exp(-\gamma)$, the following event holds

$$X_{(k)} - \mathbb{E}[X_{(k)}] < d^r_{k,\gamma} \Rightarrow X_{(k)} < \mathbb{E}[X_{(k)}] + d^r_{k,\gamma} \leq Q^\tau + w_n + d^r_{k,\gamma} \Rightarrow \hat{Q}_n^\tau - Q^\tau < w_n + d^r_{k,\gamma} \tag{65}$$

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 4, we have $\mathbb{E}[X_{(k)}] \geq Q^\tau - w_n$. With probability at least $1 - \exp(-\gamma)$, the following event holds

$$\mathbb{E}[X_{(k)}] - X_{(k)} < d^l_{k,\gamma} \Rightarrow -X_{(k)} < -\mathbb{E}[X_{(k)}] + d^l_{k,\gamma} \leq -(Q^\tau - w_n) + d^l_{k,\gamma} \Rightarrow Q^\tau - \hat{Q}_n^\tau < w_n + d^l_{k,\gamma} \tag{66}$$

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*

$$\mathbb{P}\left(X_{(k)} - \mathbb{E}[X_{(k)}] \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2(v^r + c^r\epsilon)}\right) \tag{67}$$

$$\mathbb{P}\left(\mathbb{E}[X_{(k)}] - X_{(k)} \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2v^l}\right) \tag{68}$$

Proof. Eq. (68) follows by setting $\epsilon = \sqrt{2v^l\gamma}$. We now show the case for Eq. (67). Recall from the proof of Lemma 1 Eq. (60), $h_1(u) = 1 + u - \sqrt{1+2u}$. Follow the elementary inequality

$$h_1(u) \geq \frac{u^2}{2(1+u)}, \quad u > 0 \tag{69}$$

Lemma 1 implies $\psi^*_{Z_k}(t) \geq \frac{t^2}{2(v+ct)}$, so the statement Eq. (67) follows from Chernoff's inequality. ☐

Corollary 1 (Representation of Concentration inequalities for Quantiles)**.** *For* $\epsilon > w_n$, *Theorem 2 can be represented as*

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

Proof. Observe that from the concentration inequalities of order statistics shown in Lemma 1 and 2 to the concentration inequalities of quantiles shown in Theorem 2, the confidence intervals change from $d^r_{k,\gamma}, d^l_{k,\gamma}$ to $d^r_{k,\gamma} + w_n, d^l_{k,\gamma} + w_n$. With similar reasoning as shown in proof of Corollary 2, and we let $\epsilon = d^r_{k,\gamma} + w_n$, we have $\gamma = \frac{(\epsilon - w_n)^2}{2(v^r + c^r(\epsilon - w_n))}$. Similar for the left hand side. $\square$

C Bandits Proof

We show the proof of Theorem 1. The proof technique follows Bubeck et al. (2013), and based on the concentration inequality proposed in Section 5.

Theorem 1 (Q-SAR Probability of Error Upper Bound). *The probability of error (Definition 1) for Q-SAR satisfies*

$$e_N \leq 2K^2 \exp\left(-\frac{N-K}{16\overline{\log}(K)H^{Q-SAR}} + C\right),$$

where C is the constant $(\frac{1}{4(1-\tau)} + \frac{b}{2\Delta_{(1)}})\left(L_C^2 \Delta^2_{(K)} \max\{\frac{1-\tau}{4+L_C\Delta_{(K)}}, \frac{(1-\tau)^2}{8}\}\right)$, with $L_C = \max_{i\in\mathcal{K}} L_i$, and b is a positive constant.

Proof. Recall we order the arms according to optimality as $o_1, \dots, o_K$ s.t. $Q^\tau_{o_1} \geq \dots \geq Q^\tau_{o_K}$. The optimal arm set of size m is $\mathcal{S}^*_m = \{o_1, \dots, o_m\}$. 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 $\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{K+1-p}$ such that $Q^\tau_{\bar{a}_1} \geq Q^\tau_{\bar{a}_2} \geq \dots \geq Q^\tau_{\bar{a}_{K+1-p}}$. 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

$$\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{l_p}\} \subseteq \mathcal{S}^*_m, \quad \{\bar{a}_{l_p+1}, \dots, \bar{a}_{K+1-p}\} \subseteq \mathcal{K}\backslash\mathcal{S}^*_m \tag{70}$$

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 $\hat{Q}^\tau_{a_{best},n_p} \geq \hat{Q}^\tau_{a_2,n_p} \geq \dots \geq \hat{Q}^\tau_{a_{worst},n_p}$.

Consider a event ξ,

$$\xi = \{\forall i \in \{1, \dots, K\}, p \in \{1, \dots, K-1\}, \left|\hat{Q}^\tau_{i,n_p} - Q^\tau_i\right| < \frac{1}{4}\Delta_{(K+1-p)}\} \tag{71}$$

Based on Corollary 1 and the union bound, with $A_{i,p} = \frac{(1-\tau)\Delta^2_{(K+1-p)}L_i^2}{16(4+\Delta_{(K+1-p)}L_i)}$, $B_{i,p} = \frac{(1-\tau)^2\Delta^2_{(K+1-p)}L_i^2}{16\times 8}$, we derive the upper bound of probability for the complementary event $\bar{\xi}$ as

$$\mathbb{P}(\bar{\xi}) \leq \sum_{i=1}^{K}\sum_{p=1}^{K-1} \mathbb{P}\left(\left|\hat{Q}^\tau_{i,n_p} - Q^\tau_i\right| \geq \frac{1}{4}\Delta_{(K+1-p)}\right) \quad \text{union bound} \tag{72}$$

$$\leq \sum_{i=1}^{K}\sum_{p=1}^{K-1} \left(\mathbb{P}\left(\hat{Q}^\tau_{i,n_p} - Q^\tau_i \geq \frac{1}{4}\Delta_{(K+1-p)}\right) + \mathbb{P}\left(Q^\tau_i - \hat{Q}^\tau_{i,n_p} \geq \frac{1}{4}\Delta_{(K+1-p)}\right)\right) \tag{73}$$

$$\leq \sum_{i=1}^{K}\sum_{p=1}^{K-1} \exp\left(-\frac{(\frac{1}{4}\Delta_{(K+1-p)} - w_n)^2}{2(v^r + c^r(\frac{1}{4}\Delta_{(K+1-p)} - w_n))}\right) + \exp\left(-\frac{(\frac{1}{4}\Delta_{(K+1-p)} - w_n)^2}{2v^l}\right) \quad \text{By Corollary1} \tag{74}$$

$$\leq \sum_{i=1}^{K}\sum_{p=1}^{K-1} \exp\left(-A_{i,p} \times n_p + (\frac{1}{1-\tau} + \frac{8b}{\Delta_{(K+1-p)}})A_{i,p}\right) + \tag{75}$$

$$\exp\left(-B_{i,p} \times n_p + (\frac{4}{1-\tau} + \frac{8b}{\Delta_{(K+1-p)}})B_{i,p}\right) \tag{76}$$

$$\leq \sum_{i=1}^{K}\sum_{p=1}^{K-1} 2\exp\left(-\min\{A_{i,p}, B_{i,p}\} \times n_p + (\frac{4}{1-\tau} + \frac{8b}{\Delta_{(K+1-p)}})\max\{A_{i,p}, B_{i,p}\}\right) \tag{77}$$

$$\leq 2K^2 \exp(-\frac{N-K}{16\overline{\log}(K)H^{Q-SAR}} + C). \tag{78}$$

where $H^{Q-SAR} = \max_{i \in \{1,\dots,K\}} \frac{i}{L_H^2 \Delta_{(i)}^2 \min\{\frac{1-\tau}{4+L_H\Delta_{(i)}}, \frac{(1-\tau)^2}{8}\}}$ and C is a constant such that $C = (\frac{1}{4(1-\tau)} + \frac{b}{2\Delta_{(1)}}) \left(L_C^2 \Delta_{(K)}^2 \max\{\frac{1-\tau}{4+L_C\Delta_{(K)}}, \frac{(1-\tau)^2}{8}\} \right)$.

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 $\bar{a}_j$ is accepted, then $\bar{a}_j \in \mathcal{S}_m^*$.
We prove by contradiction. Assume arm $\bar{a}_j$ is accepted in phase p, but $\bar{a}_j \notin \mathcal{S}_m^*$, i.e. $Q^\tau_{\bar{a}_j} \leq Q^\tau_{\bar{a}_{l_p+1}} \leq Q^\tau_{o_{m+1}}$. According to Algorithm 1, arm $\bar{a}_j$ is accepted only if its empirical quantile is the maximum among all active arms in phase p, thus $\hat{Q}^\tau_{\bar{a}_j,n_p} \geq \hat{Q}^\tau_{\bar{a}_1,n_p}$. On event ξ, we have

$$Q^\tau_{\bar{a}_j} + \frac{1}{4}\Delta_{(K+1-p)} > \hat{Q}^\tau_{\bar{a}_j,n_p} \geq \hat{Q}^\tau_{\bar{a}_1,n_p} > Q^\tau_{\bar{a}_1} - \frac{1}{4}\Delta_{(K+1-p)} \tag{79}$$

$$\Rightarrow \Delta_{(K+1-p)} > \frac{1}{2}\Delta_{(K+1-p)} > Q^\tau_{\bar{a}_1} - Q^\tau_{\bar{a}_j} \geq Q^\tau_{\bar{a}_1} - Q^\tau_{o_{m+1}}. \tag{80}$$

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

$$\hat{Q}^\tau_{\bar{a}_j,n_p} - \hat{Q}^\tau_{a_{l_p+1},n_p} > \hat{Q}^\tau_{a_{l_p},n_p} - \hat{Q}^\tau_{a_{K+1-p},n_p}. \tag{81}$$

In the following, we will connect Eq. (81) with the corresponding population quantiles on event ξ. We first connect $\hat{Q}^\tau_{a_{K+1-p},n_p}$ and $Q^\tau_{\bar{a}_{K+1-p}}$. Since $\hat{Q}^\tau_{a_{K+1-p},n_p}$ is the minimum empirical quantile at phase p,

$$\hat{Q}^\tau_{a_{K+1-p},n_p} \leq \hat{Q}^\tau_{\bar{a}_{K+1-p},n_p} < Q^\tau_{\bar{a}_{K+1-p}} + \frac{1}{4}\Delta_{(K+1-p)}. \tag{82}$$

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 $i \leq l_p$,

$$\hat{Q}^\tau_{\bar{a}_i,n_p} > Q^\tau_{\bar{a}_i} - \frac{1}{4}\Delta_{(K+1-p)} \geq Q^\tau_{\bar{a}_{l_p}} - \frac{1}{4}\Delta_{(K+1-p)} \geq Q^\tau_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}, \tag{83}$$

which means there are l_p arms in active set, i.e. $\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{l_p}\}$, having empirical quantiles bigger or equal than $Q^\tau_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}$. Additionally, although $j > l_p$, $\bar{a}_j$ 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.

$$\hat{Q}^\tau_{a_{l_p},n_p} \geq \hat{Q}^\tau_{a_{l_p+1},n_p} \geq Q^\tau_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}. \tag{84}$$

Combine Eq. (81)(82)(84) together, we have

$$(Q^\tau_{\bar{a}_j} + \frac{1}{4}\Delta_{(K+1-p)}) - (Q^\tau_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}) > (Q^\tau_{o_m} - \frac{1}{4}\Delta_{(K+1-p)}) - (Q^\tau_{\bar{a}_{K+1-p}} + \frac{1}{4}\Delta_{(K+1-p)}) \tag{85}$$

$$\Rightarrow \Delta_{(K+1-p)} > 2Q^\tau_{o_m} - (Q^\tau_{\bar{a}_j} + Q^\tau_{\bar{a}_{K+1-p}}) > Q^\tau_{o_m} - Q^\tau_{\bar{a}_{K+1-p}}. \tag{86}$$

From Eq. (80)(86), we have $\Delta_{(K+1-p)} > \max\{Q^\tau_{\bar{a}_1} - Q^\tau_{o_{m+1}}, Q^\tau_{o_m} - Q^\tau_{\bar{a}_{K+1-p}}\}$, which contradicts the fact that $\Delta_{(K+1-p)} \leq \max\{Q^\tau_{\bar{a}_1} - Q^\tau_{o_{m+1}}, Q^\tau_{o_m} - Q^\tau_{\bar{a}_{K+1-p}}\}$, since at phase p, there are only $p-1$ arms have been accepted or rejected. So we have if an arm $\bar{a}_j$ is accepted, then $\bar{a}_j \in \mathcal{S}_m^*$, which finishes the proof of Case 1.

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

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

$$Q^\tau_{\bar{a}_j} - \frac{1}{4}\Delta_{(K+1-p)} < \hat{Q}^\tau_{\bar{a}_j,n_p} \leq \hat{Q}^\tau_{\bar{a}_{K+1-p},n_p} < Q^\tau_{\bar{a}_{K+1-p}} + \frac{1}{4}\Delta_{(K+1-p)} \tag{87}$$

$$\Rightarrow \Delta_{(K+1-p)} > \frac{1}{2}\Delta_{(K+1-p)} > Q^\tau_{\bar{a}_j} - Q^\tau_{\bar{a}_{K+1-p}} \geq Q^\tau_{o_m} - Q^\tau_{\bar{a}_{K+1-p}}. \tag{88}$$

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

$$\hat{Q}^{\tau}_{a_1,n_p} - \hat{Q}^{\tau}_{a_{l_p+1},n_p} \leq \hat{Q}^{\tau}_{a_{l_p},n_p} - \hat{Q}^{\tau}_{\bar{a}_j,n_p}. \tag{89}$$

In the following, we will connect Eq. (89) with the corresponding population quantiles on event ξ. We first connect $\hat{Q}^{\tau}_{a_1,n_p}$ and $Q^{\tau}_{\bar{a}_1}$. Since $\hat{Q}^{\tau}_{a_1,n_p}$ is the maximum empirical quantile at phase p,

$$\hat{Q}^{\tau}_{a_1,n_p} \geq \hat{Q}^{\tau}_{\bar{a}_1,n_p} > Q^{\tau}_{\bar{a}_1} - \frac{1}{4}\Delta_{(K+1-p)}. \tag{90}$$

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 $i \geq l_p + 1$,

$$\hat{Q}^{\tau}_{\bar{a}_i,n_p} < Q^{\tau}_{\bar{a}_i} + \frac{1}{4}\Delta_{(K+1-p)} \leq Q^{\tau}_{\bar{a}_{l_p+1}} + \frac{1}{4}\Delta_{(K+1-p)} \leq Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}, \tag{91}$$

Additionally, although $j < l_p + 1$, $\bar{a}_j$ 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,

$$\hat{Q}^{\tau}_{a_{l_p+1},n_p} \leq \hat{Q}^{\tau}_{a_{l_p},n_p} \leq Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}. \tag{92}$$

Combining Eq. (89), (90) and (92) together, we have

$$(Q^{\tau}_{\bar{a}_1} - \frac{1}{4}\Delta_{(K+1-p)}) - (Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}) \leq (Q^{\tau}_{o_{m+1}} + \frac{1}{4}\Delta_{(K+1-p)}) - (Q^{\tau}_{\bar{a}_j} - \frac{1}{4}\Delta_{(K+1-p)}) \tag{93}$$

$$\Rightarrow \Delta_{(K+1-p)} \geq (Q^{\tau}_{\bar{a}_j} + Q^{\tau}_{\bar{a}_1}) - 2Q^{\tau}_{o_{m+1}} > Q^{\tau}_{\bar{a}_1} - Q^{\tau}_{o_{m+1}}. \tag{94}$$

From Eq. (88)(94), we have $\Delta_{(K+1-p)} > \max\{Q^{\tau}_{\bar{a}_1} - Q^{\tau}_{o_{m+1}}, Q^{\tau}_{o_m} - Q^{\tau}_{\bar{a}_{K+1-p}}\}$, which contradicts the fact that $\Delta_{(K+1-p)} \leq \max\{Q^{\tau}_{\bar{a}_1} - Q^{\tau}_{o_{m+1}}, Q^{\tau}_{o_m} - Q^{\tau}_{\bar{a}_{K+1-p}}\}$, since at phase p, there are only $p-1$ arms have been accepted or rejected. So we have if an arm $\bar{a}_j$ is rejected, then $\bar{a}_j \notin \mathcal{S}^*_m$, which finishes the proof of Case 2. $\square$