A Cost-Based Analysis for Risk-Averse Explore-Then-Commit Finite-Time Bandits (Supplemental File)

Ali Yekkehkhany[1], Ebrahim Arian[2], Rakesh Nagi[3], and Ilan Shomorony[4]
Electrical and Computer Engineering[1,4], Industrial and Enterprise Systems Engineering[2,3]
University of Illinois at Urbana-Champaign, {yekkehk2, arian2, nagi, ilans}@illinois.edu

---

Appendix

A. Proof of Theorem 1

Consider the Bernoulli random variables $B_k = \mathbb{1}\{R_k \geq \boldsymbol{R}_{-k}\}$ and their unknown means $p_k = \mathbb{E}[B_k] = \mathbb{P}(R_k \geq \boldsymbol{R}_{-k})$ for $k \in \mathcal{K}$. Possessing N independent observations from the joint rewards of the K arms in the pure exploration phase, the confidence interval derived from Hoeffding's inequality for estimating p_k based on Equation (4) with confidence level $1 - 2e^{-\frac{a^2}{2}}$ has the property that

$$\mathbb{P}\left(p_k \in \left(\hat{p}_k - \frac{a}{2\sqrt{N}}, \hat{p}_k + \frac{a}{2\sqrt{N}}\right)\right) \geq 1 - 2e^{-\frac{a^2}{2}}, \ \forall k \in \mathcal{K}. \tag{26}$$

In order to find a bound on regret, defined in Equation (5) as $r(\Delta p) = \mathbb{P}\left(p_{k^*} - p_{\hat{k}} > \Delta p\right)$, note that

$$\begin{aligned}
&\{p_{k^*} - p_{\hat{k}} > \Delta p\} \\
&\overset{(a)}{\subseteq} \left\{\exists k \in \mathcal{K} \text{ such that } p_k \notin \left(\hat{p}_k - \frac{\Delta p}{2}, \hat{p}_k + \frac{\Delta p}{2}\right)\right\} \\
&\overset{(b)}{\subseteq} \left\{\exists k \in \mathcal{K} \text{ such that } p_k \notin \left(\hat{p}_k - \frac{a}{2\sqrt{N}}, \hat{p}_k + \frac{a}{2\sqrt{N}}\right)\right\},
\end{aligned} \tag{27}$$

where (a) follows from the fact that if the score of the selected arm $\hat{k}$ deviates from the score of the optimal arm k^* by more than Δp, then there should exist an arm whose score is estimated by an error greater than $\frac{\Delta p}{2}$, and (b) is true if $\frac{a}{2\sqrt{N}} \leq \frac{\Delta p}{2}$. By using union bound and Equation (26), the probability of the right-hand side of the above equation can be bounded as follows, which results in the following bound on regret:

$$r(\Delta p) = \mathbb{P}\left(p_{k^*} - p_{\hat{k}} > \Delta p\right) \leq 2Ke^{-\frac{a^2}{2}} = \epsilon_r. \tag{28}$$

The above upper bound on regret is derived under the condition that $\frac{a}{2\sqrt{N}} \leq \frac{\Delta p}{2}$, which by using $a^2 = 2\ln\left(\frac{2K}{\epsilon_r}\right)$ and simple algebraic calculations results in $N \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$. □

B. Proof of Theorem 2

Consider the Bernoulli random variables $B_k^M = \mathbb{1}\{R_k^M \geq \boldsymbol{R}_{-k}^M\}$ and their unknown means $p_k^M = \mathbb{E}[B_k^M] = \mathbb{P}(R_k^M \geq \boldsymbol{R}_{-k}^M)$ for $k \in \mathcal{K}$. Possessing N independent observations from the joint rewards of the K arms in pure exploration, there are exactly $\lfloor \frac{N}{M} \rfloor$ independent samples for estimation of p_k^M. Due to the same reasoning in the proof of Theorem 1, the confidence interval for estimating p_k^M based on Equation (9) or (12) with confidence level $1 - 2e^{-\frac{a^2}{2}}$ has the property that

$$\mathbb{P}\left(p_k^M \in \left(\hat{p}_k^M - \frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}}, \hat{p}_k^M + \frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}}\right)\right) \geq 1 - 2e^{-\frac{a^2}{2}}, \tag{29}$$

for all $k \in \mathcal{K}$.

In order to find a bound on regret, defined in Definition 1 as $r_M(\Delta p) = \mathbb{P}\left(p_{k^*}^M - p_{\hat{k}}^M > \Delta p\right)$, note that

$$\begin{aligned} &\{p_{k^*}^M - p_{\hat{k}}^M > \Delta p\} \\ &\subseteq \left\{\exists k \in \mathcal{K} \text{ s.t. } p_k^M \notin \left(\hat{p}_k^M - \frac{\Delta p}{2}, \hat{p}_k^M + \frac{\Delta p}{2}\right)\right\} \\ &\overset{(a)}{\subseteq} \left\{\exists k \in \mathcal{K} \text{ s.t. } p_k^M \notin \left(\hat{p}_k^M - \frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}}, \hat{p}_k^M + \frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}}\right)\right\}, \end{aligned} \tag{30}$$

where (a) is true if $\frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}} \leq \frac{\Delta p}{2}$. By using union bound and Equation (29), the probability of the right-hand side of the above equation can be bounded as follows, which results in the following bound on regret:

$$r_M(\Delta p) = \mathbb{P}\left(p_{k^*}^M - p_{\hat{k}}^M > \Delta p\right) \leq 2Ke^{-\frac{a^2}{2}} = \epsilon_r. \tag{31}$$

The above upper bound on regret is derived under the condition that $\frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}} \leq \frac{\Delta p}{2}$, which by using $a^2 = 2\ln\left(\frac{2K}{\epsilon_r}\right)$ and simple algebraic calculations results in $\lfloor \frac{N}{M} \rfloor \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$. □

C. Proof of Theorem 3

The maximum deviation that $Cr_l(n, n_e)$ and $Cr_u(n, n_e)$ can have from $Cr(n, p_{k^*})$ is investigated with an associated confidence level. To this end, the maximum deviation of $r^*(n, \hat{p}_l^*(n_e))$ and $r^*(n, \hat{p}_u^*(n_e))$ from $r^*(n, p_{k^*})$ is found with the confidence level. First, the maximum deviation of

$\hat{p}_l^*(n_e)$ and $\hat{p}_u^*(n_e)$ from p_{k^*} with the associated confidence level is derived below. Equation (16) suggests that the following holds with confidence level $1-2e^{-\frac{a^2}{2}}$:

$$\begin{aligned} p_{k^*}-\hat{p}_l^*(n_e) &= p_{k^*}-\max\left\{\hat{p}^*(n_e)-\frac{a}{2\sqrt{n_e}},0.5\right\} \\ &\le p_{k^*}-\hat{p}^*(n_e)+\frac{a}{2\sqrt{n_e}} \le \frac{a}{2\sqrt{n_e}}+\frac{a}{2\sqrt{n_e}}=\frac{a}{\sqrt{n_e}}. \end{aligned} \tag{32}$$

On the other hand,

$$\begin{aligned} & p_{k^*}-\hat{p}_l^*(n_e) \\ &= p_{k^*}-\hat{p}^*(n_e)+\hat{p}^*(n_e)-\max\left\{\hat{p}^*(n_e)-\frac{a}{2\sqrt{n_e}},0.5\right\} \\ &\ge \max\left\{\frac{-a}{2\sqrt{n_e}},0.5-\hat{p}^*(n_e)\right\}+\min\left\{\frac{a}{2\sqrt{n_e}},\hat{p}^*(n_e)-0.5\right\}=0. \end{aligned} \tag{33}$$

The above two equations imply that $0\le p_{k^*}-\hat{p}_l^*(n_e)\le\frac{a}{\sqrt{n_e}}$ with confidence level $1-2e^{-\frac{a^2}{2}}$. Similarly, it can be proved that $0\le\hat{p}_u^*(n_e)-p_{k^*}\le\frac{a}{\sqrt{n_e}}$ with the mentioned confidence level.

In the following, Lipschitz constant of function $r^*(n,p)$ with respect to p is calculated by differentiating the regret function presented in Equation (14) with respect to p as

$$\begin{aligned} \frac{\partial r^*(n,p)}{\partial p} = & \sum_{i=\lfloor\frac{n}{2}\rfloor+1}^{n}\binom{n}{i}\cdot(1-p)^i\cdot p^{n-i}\cdot\left(\frac{n-i}{p}-\frac{i}{1-p}\right) \\ & +\frac{1}{2}\cdot\binom{n}{\frac{n}{2}}\cdot(1-p)^{\frac{n}{2}}\cdot p^{\frac{n}{2}}\cdot\frac{n}{2}\cdot\left(\frac{1}{p}-\frac{1}{1-p}\right)\cdot\mathbb{1}\{n \text{ is even}\}. \end{aligned} \tag{34}$$

Since $0.5\le p\le 1$, it is easy to verify that $\frac{\partial r^*(n,p)}{\partial p}\le 0$, so $r^*(n,p)$ is decreasing in terms of p. Consider n is an odd number, then $\frac{\partial r^*(n,p)}{\partial p}=\sum_{i=\lfloor\frac{n}{2}\rfloor+1}^{n}\binom{n}{i}\cdot(1-p)^i\cdot p^{n-i}\cdot\left(\frac{n-i}{p}-\frac{i}{1-p}\right)=\sum_{i=\lfloor\frac{n}{2}\rfloor+1}^{n}\binom{n}{i}\cdot(1-p)^i\cdot p^{n-i}\cdot\left(\frac{n\cdot(1-p)-i}{p\cdot(1-p)}\right)$, where $n\cdot(1-p)-i\le\frac{n}{2}-i\le-\frac{1}{2}$ as $0.5\le p\le 1$ and $i\ge\frac{n+1}{2}$, which proves that $\frac{\partial r^*(n,p)}{\partial p}\le 0$. Similarly, it can be proved that $\frac{\partial r^*(n,p)}{\partial p}\le 0$ for the case when n is an even number or one can use the following equation for the derivative. The derivative of $r^*(n,p)$ with respect to p calculated above can be written as follows by algebraic manipulations:

$$\frac{\partial r^*(n,p)}{\partial p}=\begin{cases} -n\binom{n-1}{\frac{n-1}{2}}p^{\frac{n-1}{2}}(1-p)^{\frac{n-1}{2}}, & \text{if } n \text{ is odd,} \\ -(n-1)\binom{n-2}{\frac{n-2}{2}}p^{\frac{n-2}{2}}(1-p)^{\frac{n-2}{2}}, & \text{if } n \text{ is even.} \end{cases} \tag{35}$$

Note that $\frac{\partial r^*(n,p)}{\partial p}=\frac{\partial r^*(n+1,p)}{\partial p}$ when n is an odd number and $p\in[0.5,1]$. On the other hand, it is obvious that $r^*(n,1)=r^*(n+1,1)$, so

$$r^*(n,p)=r^*(n+1,p), \text{ if } n \text{ is odd.} \tag{36}$$

As a result, in terms of regret, it is not worth it to perform even number of experiments since the last experiment does not improve regret.

It is easy to verify that $\frac{\partial r^*(n,p)}{\partial p}\Big|_{p=0.5}$ can get arbitrarily large by increasing n. Hence, it is assumed that $p_{k^*} \in [0.5+\epsilon_p, 1]$, where ϵ_p can be any small number in the interval $(0, 0.5]$. In the following, the logarithm in base two of $\left|\frac{\partial r^*(n,p)}{\partial p}\right|$ is taken when n is an odd number, and as mentioned earlier, when n is even, the answer is the same as for $n-1$ which is an odd number.

$$\log_2\left|\frac{\partial r^*(n,p)}{\partial p}\right| = \log_2 n + \log_2 \frac{(n-1)!}{\left((\frac{n-1}{2})!\right)^2} + \frac{n-1}{2}\Big(\log_2 p + \log_2(1-p)\Big)$$
$$\overset{(a)}{\leq} \log_2 n + \Big[(n-\frac{1}{2})\log_2(n-1) - (n-1)\log_2 e + \log_2 e - 2\Big(\frac{n}{2}\log_2\frac{n-1}{2} - \frac{n-1}{2}\log_2 e + \frac{1}{2}\log_2 2\pi\Big)\Big]$$
$$-(n-1)(1+\delta_p) \leq \frac{1}{2}\log_2(n+2) - \delta_p(n-1), \tag{37}$$

where (a) follows by Stirling's approximation, $(n-1)! \leq (n-1)^{n-\frac{1}{2}}e^{-n+2}$ and $\left(\frac{n-1}{2}\right)! \geq \sqrt{2\pi}\left(\frac{n-1}{2}\right)^{\frac{n}{2}} e^{-(\frac{n-1}{2})}$, and defining $\delta_p = \frac{1}{2}\left(-2 - \log_2(0.5+\epsilon_p) - \log_2(0.5-\epsilon_p)\right) > 0$. As a result,

$$\left|\frac{\partial r^*(n,p)}{\partial p}\right| \leq \sqrt{n+2}\cdot 2^{-\delta_p(n-1)}, \quad \lim_{n\to\infty}\left|\frac{\partial r^*(n,p)}{\partial p}\right| = 0. \tag{38}$$

Also note that $\left|\frac{\partial r^*(n,p)}{\partial p}\right|$ given by Equation (35) is finite for any given n, so Equation (38) suggests that $\left|\frac{\partial r^*(n,p)}{\partial p}\right|$ is finite for any $n \in \{1,2,3,\dots\}$ and any $p \in [0.5+\epsilon_p, 1]$.

Equations (32), (33), (38), and the fact that $r^*(n,p)$ is decreasing in terms of p result in the following equation for any $n \in \{1,2,3,\dots\}$ with confidence level $1-2e^{-\frac{a^2}{2}}$:

$$0 \leq Cr\left(n, p_{k^*}\right) - Cr_l\left(n, n_e\right) = \alpha\cdot\left[r^*\left(n, p_{k^*}\right) - r^*\left(n, \hat{p}_u^*(n_e)\right)\right] \leq \frac{a\cdot\alpha\cdot\sqrt{n+2}\cdot 2^{-\delta_p\cdot(n-1)}}{\sqrt{n_e}}. \tag{39}$$

The above equation is true when n is odd, but recall that $r^*(n,p) = r^*(n+1,p)$ for an odd number n. In order to come up with a unified formula for $Cr\left(n, p_{k^*}\right) - Cr_l\left(n, n_e\right)$ for even and odd numbers n, define $\Delta Cr(n, n_e)$ as

$$\Delta Cr(n, n_e) \triangleq \frac{a\cdot\alpha\cdot\sqrt{n+2}\cdot 2^{-\delta_p\cdot(n-2)}}{\sqrt{n_e}}, \tag{40}$$

where $\lim_{n_e\to\infty} \Delta Cr(n, n_e) = 0,\ \forall n \in \{1,2,3,\cdots\}$. The same bounds can be found for $Cr_u(n, n_e) - Cr(n, p_{k^*})$, so

$$\begin{aligned} 0 &\leq Cr(n, p_{k^*}) - Cr_l(n, n_e) \leq \Delta Cr(n, n_e), \\ 0 &\leq Cr_u(n, n_e) - Cr(n, p_{k^*}) \leq \Delta Cr(n, n_e). \end{aligned} \tag{41}$$

The upper bound in Equation (18) with confidence level $1-2e^{-\frac{a^2}{2}}$ is proved as follows. Equation (41) results in the following for any $n \in \{1,2,3,\dots\}$:

$$Cr\left(n, \hat{p}^*(n_e)\right) - \Delta Cr(n, n_e) \leq Cr\left(n, p_{k^*}\right) \leq Cr\left(n, \hat{p}^*(n_e)\right) + \Delta Cr(n, n_e). \tag{42}$$

Taking minimum with respect to n from all sides of the above inequality results in

$$\begin{aligned} &Cr\left(\hat{N}^*(n_e), \hat{p}^*(n_e)\right) - \max_n\left\{\Delta Cr(n, n_e)\right\} \\ \leq &Cr\left(N^*, p_{k^*}\right) \leq Cr\left(\hat{N}^*(n_e), \hat{p}^*(n_e)\right) + \max_n\left\{\Delta Cr(n, n_e)\right\}. \end{aligned} \tag{43}$$

Using Equations (42) and (43) concludes as

$$
\begin{aligned}
&Cr\left(\hat{N}^*(n_e), p_{k^*}\right) - Cr\left(N^*, p_{k^*}\right)\\
&\leq \max_n \{\Delta Cr(n, n_e)\} + \Delta Cr(\hat{N}^*(n_e), n_e) \leq \frac{D_p}{2\sqrt{n_e}} + \Delta Cr(\hat{N}^*(n_e), n_e),
\end{aligned}
\tag{44}
$$

where $D_p = \frac{a \cdot \alpha \cdot 2^{\left(4\delta_p + 1 - \frac{1}{2\ln 2}\right)}}{\sqrt{2\delta_p \ln 2}}$ is a constant that is derived as follows. For a given n_e, the function $\Delta Cr(n, n_e)$ is increasing in terms of n when $n < \frac{1}{2\delta_p \ln 2} - 2$ and is decreasing when $n > \frac{1}{2\delta_p \ln 2} - 2$. Hence, $\max_n \Delta Cr(n, n_e) \leq \Delta Cr(\frac{1}{2\delta_p \ln 2} - 2, n_e) = \frac{a \cdot \alpha \cdot 2^{\left(4\delta_p - \frac{1}{2\ln 2}\right)}}{\sqrt{2\delta_p n_e \ln 2}}$.

In the following, the upper bound in Equation (19) with confidence level $1 - 2e^{-\frac{a^2}{2}}$ is derived as

$$
\begin{aligned}
&\max_{n \in \mathcal{I}(n_e)} \Big(Cr(n, p_{k^*}) - Cr(N^*, p_{k^*})\Big)\\
&\overset{(a)}{\leq} \max_{n \in \mathcal{I}(n_e)} \Big(Cr_l(n, n_e) - Cr(N^*, p_{k^*}) + \Delta Cr(n, n_e)\Big)\\
&\overset{(b)}{=} \max_{n \in \mathcal{I}(n_e)} \Big(\underbrace{Cr_l(n, n_e) - Cr_u(N_u^*, n_e)}_{\text{it is non-positive due to Equation (17)}} + \\
&\qquad\qquad Cr_u(N_u^*, n_e) - Cr(N^*, p_{k^*}) + \Delta Cr(n, n_e)\Big)\\
&\overset{(c)}{\leq} \max_{n \in \mathcal{I}(n_e)} \Big(Cr_u(N^*, n_e) - Cr(N^*, p_{k^*}) + \Delta Cr(n, n_e)\Big)\\
&\overset{(d)}{\leq} \max_{n \in \mathcal{I}(n_e)} 2\Delta Cr(n, n_e) \leq \max_n 2\Delta Cr(n, n_e) \leq \frac{D_p}{\sqrt{n_e}},
\end{aligned}
\tag{45}
$$

where (a) follows by Equation (41), (b) is true by subtracting and adding the term $Cr_u(N_u^*, n_e)$, (c) uses the fact that $N_u^* = \arg\min_n Cr_u(n, n_e)$, so $Cr_u(N_u^*, n_e) \leq Cr_u(N^*, n_e)$, and (d) again follows by Equation (41). $\square$