Stochastic MABs Ranking

MOAYAD ALNAMMI

Problem Overview (1)

- Given n arms with mean rewards $\mu_1 > \mu_2 \geq \ \dots \geq \mu_n$ where $\mu_{i_*} = \mu_1$
- $\mu_i \in [0,1]$, $\quad \Delta_i = \mu_1 - \mu_i$ for $i = 2, \dots, n$
- **Fixed confidence:** Given confidence δ, find the best arm with probability at least $1 - \delta$. Algorithm satisfies $\sup_{\mu_1,\ldots,\mu_n} P(\hat{i} \neq i_*) \leq \delta$
- **Fixed budget:** Given budget T, do not exceed sample budget and identify best arm with as highest probability possible.
- Paper focuses on fixed confidence setting.
- Summarizes three main strategies: action elimination (AE), upper confidence bound (UCB), lower UCB (LUCB).
- Uses similar framework to prove sample complexity of each.
- Showcases experimental behavior.

Problem Overview (2)

- Given n arms with mean rewards $\mu_1 > \mu_2 \geq \ ... \geq \mu_n$ where $\mu_{i_*} = \mu_1$
- $\mu_i \in [0,1], \quad \Delta_i = \mu_1 - \mu_i$ for $i = 2, \dots, n$
- $X_{i,t}$ is a sample from arm i at time step t. $E[X_{i,t}] = \mu_i$
- $a \leq X_{i,t} \leq b$ with $(b-a) \leq 1$. $(X_{i,t} - \mu_i)$ is a sub-Guassian with $\sigma \leq 0.5$
- $T_i(t)$ denotes the number of samples/pulls from arm i at time t.
- $\hat{\mu}_{i,T_i(t)}$ is empirical mean of arm i at time t.
- Define: $h_t = \arg\max_{i \in [n]} \mu_{i,\hat{T}_i(t)} \qquad \ell_t = \arg\max_{i \in [n] \setminus h_t} \mu_{i,\hat{T}_i(t)} + C_{i,T_i(t)}$
- $C_{i,T_i(t)}$ derived from tail bond, depends on $t, T_i(t), n, \delta$

Definitions and Lemmas (1)

1. SubGaussian RV:

if X is a subGaussian RV with scale parameter σ, then

- $E[X] = 0$
- $E[e^{tX}] \leq \exp\left(\frac{\sigma^2 t^2}{2}\right), \forall t \in R$
- $P(|X| > t) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right), \forall t \in R$

If $a \leq X \leq b$ then take $\sigma = \frac{(b-a)}{2}$ via Hoeffding's: $E[e^{tX}] \leq \exp\left(\frac{(b-a)^2 t^2}{8}\right), \forall t \in R$

Definitions and Lemmas (2)

2. Finite LIL Bound Lemma: see [10]

Let $X_1, X_2, \ldots$ be i.i.d $subGaus(\sigma^2)$. For any $\epsilon \in (0,1)$ and $\delta \in (0, \frac{\log(1+\epsilon)}{e})$ then $\forall t \geq 1$:

$$P\left(\sum_{s=1}^{t} X_s \leq (1+\sqrt{\varepsilon})\sqrt{2\sigma^2(1+\varepsilon)t\log\left(\frac{\log((1+\varepsilon)t)}{\delta}\right)}\right) \geq 1 - \frac{2+\varepsilon}{\varepsilon}\left(\frac{\delta}{\log(1+\varepsilon)}\right)^{1+\varepsilon}$$

Definitions and Lemmas (3)

3. Restated Finite LIL Bound Lemma:

For arm i with mean μ_i, let $X_1, X_2, \ldots$ be i.i.d draws from arm i. We assumed that $(X_s - \mu_i)$ is $subGaus(\sigma^2)$ with $\sigma \leq 0.5$. For any $\epsilon \in (0,1)$ and $\delta \in (0, \frac{\log(1+\epsilon)}{e})$ then $\forall t \geq 1$:

$$P\Big(\big|\frac{1}{t}\sum_{s=1}^{t} X_s - \mu_i\big| \leq U(t,\delta)\Big) \geq 1 - \frac{2+\epsilon}{\epsilon/2}\Big(\frac{\delta}{\log(1+\epsilon)}\Big)^{1+\epsilon}$$

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}})$$

Definitions and Lemmas (4)

4. Apply Lemma:

$$
\begin{aligned}
P\left(\bigcap_{i=1}^{n} |\hat{\mu}_{i,T_i(t)} - \mu_i| \le U(T_i(t), \delta/n)\right) &\ge \sum_{i=1}^{n} P(\ldots) - n + 1 \\
&\ge n\left(1 - \tfrac{2+\epsilon}{\epsilon/2}\left(\tfrac{\delta/n}{\log(1+\epsilon)}\right)^{1+\epsilon}\right) - n + 1 \\
&= 1 - \tfrac{2+\epsilon}{\epsilon/2}\left(\tfrac{1}{\log(1+\epsilon)}\right)^{1+\epsilon} \delta(\delta/n)^{\epsilon} \\
&\ge 1 - \tfrac{2+\epsilon}{\epsilon/2}\left(\tfrac{1}{\log(1+\epsilon)}\right)^{1+\epsilon} \delta
\end{aligned}
$$

Definitions and Lemmas (4)

5. Useful Inequality: see (1) in [10]

For $t \geq 1, \epsilon \in (0,1), c > 0, 0 < \delta \leq 1$:

$$c \leq \frac{1}{t} \log\left(\frac{\log((1+\epsilon)t)}{\delta}\right) \implies t \leq \frac{1}{c} \log\left(\frac{2\log((1+\epsilon)/(c\delta))}{\delta}\right) \tag{1}$$

Definitions and Lemmas (4)

6. Useful Inequality: see (2) in [10]

For $t \geq 1, s \geq 3, \epsilon \in (0,1), c \in (0,1], 0 < \delta \leq 1$:

$$\frac{1}{t} \log\left(\frac{\log((1+\epsilon)t)}{\delta}\right) \geq \frac{c}{s} \log\left(\frac{\log((1+\epsilon)s)}{\delta}\right)$$

$$\Longrightarrow t \leq \frac{s}{c} \frac{\log\left(2\log\left(\frac{1}{c\delta}\right)/\delta\right)}{\log(1/\delta)} \qquad (2)$$

Algorithms

General Strategy	Algorithm	Sample Complexity	Year
Action Elimination (AE)	Successive elimination	$O(\sum_{i\neq i_*} \Delta_i^{-2} \log(n\Delta^{-2}))$ $\Omega(\sum_{i\neq i_*} \Delta_i^{-2})$	2002 [4] 2004 [5]
	PRISM	$O(\sum_{i\neq i_*} \Delta_i^{-2} \log\log(\sum_{j\neq i_*} \Delta_j^{-2}))$ or $O(\sum_{i\neq i_*} \Delta_i^{-2} \log(\Delta_i^{-2}))$	2013 [8]
	*Exp-gap elimination	$O(\sum_{i\neq i_*} \Delta_i^{-2} \log\log(\Delta_i^{-2}))$	2013 [9]
Upper confidence bounds (UCB)	*Lil' UCB	$O(\sum_{i\neq i_*} \Delta_i^{-2} \log\log(\Delta_i^{-2}))$ $\Omega(\sum_{i\neq i_*} \Delta_i^{-2} \log\log(\Delta_i^{-2}))$	Late 2013 [10]
Lower UCB (LUCB)	LUCB	$O(\sum_{i\neq i_*} \Delta_i^{-2} \log(\sum_{j\neq i_*} \Delta_j^{-2}))$	2012 [7] m-best arms

Action Elimination Strategy

1. Let $\Omega_1 = [1, 2, \dots, n]$, t=1
2. While $|\Omega_t| > 1$:
3. Sample from each arm $i \in \Omega_t$, r_t times
4. Compute reference arm $a = argmax_{i \in [n]}\ \hat{\mu}_{i,T_i(t)} + C_{i,T_i(t)}$
5. Update $\Omega_{t+1} = \left\{ i \in \Omega_t : \underbrace{\hat{\mu}_{a,T_a(t)} - C_{a,T_a(t)} < \hat{\mu}_{i,T_i(t)} + C_{i,T_i(t)}}_{\text{Arm eliminated when UCB <= reference arm's LCB}} \right\}$
6. t=t+1
7. Return last $i \in \Omega_t$

PRISM – Jamieson et. al (2013) https://arxiv.org/abs/1306.3917

Input δ. Let $A_1 = \{0, 1, \ldots, n\}$, $n_\ell = \ell 2^\ell$, and $\varepsilon_\ell = \sqrt{\frac{\log(1/\delta)}{2^\ell}}$.

For each phase $\ell = 1, 2, \ldots,$

(1) Let i_ℓ be the output of Median Elimination [10] run on A_ℓ with accuracy $\left(\varepsilon_\ell, \delta^\ell\right)$.

(2) For each arm $i \in A_\ell$, sample n_ℓ times arm i and let $\widehat{\mu}_i(\ell)$ be the corresponding average.

(3) Let

$$A_{\ell+1} = \{i \in A_\ell : \widehat{\mu}_i(\ell) \geq \widehat{\mu}_{i_\ell} - 2\varepsilon_\ell\}.$$

Stop when A_ℓ contains a unique element $\widehat{i}$ and output $\widehat{i}$.

Figure 2: PRISM algorithm for the best arm identification problem.

$$O\left(\log(1/\delta)\left[\mathbf{H}\log(\log(1/\delta)) + \sum_{i=1}^{n} \Delta_i^{-2}\log_2(\Delta_i^{-2})\right]\right)$$

Conservative PRISM: $O\left(\mathbf{H}\log\left(\frac{\log(\mathbf{H})}{\delta}\right)\right)$

AE Termination (1)

$$P(\bigcap_{i=1}^{n} |\mu_{i,\hat{T}_i(t)} - \mu_i| \leq U(T_i(t), \delta/n)) \geq 1 - \frac{2+\epsilon}{\epsilon/2}\left(\frac{1}{\log(1+\epsilon)}\right)^{1+\epsilon}\delta$$

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

1. Let $r_k = 1$ for $k = 1, 2, \ldots$. So $T_i(k) = k$ for $i \in \Omega_k$.
2. Let $C_{i,k} = 2U(k, \frac{\delta}{n})$ and $a = argmax_{i\in\Omega_k} \hat{\mu}_{i,T_i(t)}$
3. At epoch k, if $i_* \in \Omega_k$ then:

$$\begin{aligned}\hat{\mu}_{a,k} - \hat{\mu}_{i_*,k} &= (\hat{\mu}_{a,k} - \mu_a) + (\mu_{i_*} - \hat{\mu}_{i_*,k}) - \Delta_a \\ &\leq U(T_a(k), \delta/n) + U(T_{i_*}(k), \delta/n) - \Delta_a \\ &= 2U(k, \delta/n) - \Delta_a < 2U(k, \delta/n) = C_{a,k} + C_{i_*,k}\end{aligned}$$

4. Thus $i_* \in \Omega_{k+1}$ since $\Omega_{k+1} = \{i \in \Omega_k\colon\ \hat{\mu}_{a,k} - \hat{\mu}_{i,k} < C_{a,k} + C_{i,k}\}$
5. Induction $\forall k \geq 1, i_* \in \Omega_k$.
6. If AE terminates then last arm is i_*(with prob. at least …)

AE Sample Bound (1)

$$P(\bigcap_{i=1}^{n} |\mu_{i,\hat{T}_i(t)} - \mu_i| \leq U(T_i(t), \delta/n)) \geq 1 - \frac{2+\epsilon}{\epsilon/2}\left(\frac{1}{\log(1+\epsilon)}\right)^{1+\epsilon}\delta$$

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

1. $\Omega_{k+1} = \{i \in \Omega_k\colon\ \hat{\mu}_{a,k} - \widehat{\mu}_{i,k} < C_{a,k} + C_{i,k}\}$
2. At epoch k, for arm $i \in \Omega_k$:

$$\hat{\mu}_{a,k} - \hat{\mu}_{i,k} \geq \hat{\mu}_{i_*,k} - \hat{\mu}_{i,k} + \Delta_i - \Delta_i$$
$$\geq -2U(k,\delta/n) + \Delta_i$$

3. Arm $i \notin \Omega_{k+1}$ if $\hat{\mu}_{a,k} - \widehat{\mu}_{i,k} \geq C_{a,k} + C_{i,k} = 2U(k,\delta/n)$
4. Arm i guaranteed to be thrown out when LB in (2) exceeds UB in condition.

$$-2U(k,\delta/n) + \Delta_i \geq 2U(k,\delta/n)$$

5. I.e worst case: arm i in play as long as: $\Delta_i/4 < U(k,\delta/n)$
6. Solve for k:

$$\Delta_i/4 < (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log\left(\frac{\log((1+\epsilon)k)}{\delta/n}\right)}{2k}}$$

AE Sample Bound (2)

$$P(\bigcap_{i=1}^{n} |\mu_{i,\hat{T}_i(t)} - \mu_i| \le U(T_i(t), \delta/n)) \ge 1 - \frac{2+\epsilon}{\epsilon/2}\left(\frac{1}{\log(1+\epsilon)}\right)^{1+\epsilon}\delta$$

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

$$\Delta_i/4 < (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log\left(\frac{\log((1+\epsilon)k)}{\delta/n}\right)}{2k}}$$

$$\implies \frac{\Delta_i^2}{\gamma} < \frac{1}{k}\log\left(\frac{\log((1+\epsilon)k)}{\delta/n}\right)$$

where $\gamma = 8((1+\sqrt{\epsilon})^2(1+\epsilon))$

$$\implies k < \frac{\gamma}{\Delta_i^2}\log\left(\frac{2\log(\gamma\Delta_i^{-2}(1+\epsilon)(n/\delta))}{\delta/n}\right)$$

Using (1) with $t = k, \delta = \delta/n, c = \frac{\Delta_i^2}{\gamma}$

$$\le \frac{\gamma}{\Delta_i^2}\log\left(\frac{2^2\log(\gamma\Delta_i^{-2}(1+\epsilon))^2}{\delta^2/n^2}\right)$$

since $\gamma > 8$ and $\frac{n}{\delta}\log\left(\frac{n}{\delta}\right) \le \frac{n^2}{\delta^2}$

$$= \frac{2\gamma}{\Delta_i^2}\log\left(\frac{2\log(\gamma\Delta_i^{-2}(1+\epsilon))}{\delta/n}\right)$$

AE Sample Bound (3)

$$P(\bigcap_{i=1}^n |\mu_{i,\hat{T}_i(t)} - \mu_i| \le U(T_i(t), \delta/n)) \ge 1 - \tfrac{2+\epsilon}{\epsilon/2}\left(\tfrac{1}{\log(1+\epsilon)}\right)^{1+\epsilon}\delta$$

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

1. Sum over suboptimal arm bounds:

$$\sum_{i \neq i_*} k_i < \sum_{i \neq i_*} \frac{2\gamma}{\Delta_i^2} \log\left(\frac{2\log(\gamma\Delta_i^{-2}(1+\epsilon))}{\delta/n}\right)$$

2. Account for optimal arm: $k_{i_*} < \max_{i \neq i_*} \frac{2\gamma}{\Delta_i^2} \log\left(\frac{2\log(\gamma\Delta_i^{-2}(1+\epsilon))}{\delta/n}\right)$

$$< \sum_{i \neq i_*} \frac{2\gamma}{\Delta_i^2} \log\left(\frac{2\log(\gamma\Delta_i^{-2}(1+\epsilon))}{\delta/n}\right)$$

3. Complexity: $O\left(\sum_{i \neq i_*} \Delta_i^{-2} \log\left(\frac{n\log(\Delta_i^{-2})}{\delta}\right)\right)$

4. Can't remove log(n) term due to choice of reference arm. PRISM and exp-gap use median elimination, but pays for it in constants.

UCB Strategy

1. Let $h_\mathrm{t} = argmax_{i\in[n]}\ \hat{\mu}_{i,T_i(t)}$ and $\ell_\mathrm{t} = argmax_{i\in[n]\backslash h_t}\ \hat{\mu}_{i,T_i(t)} + C_{i,T_i(t)}$
2. Sample from each arm $i \in \Omega$, 1 time. t=n+1
3. while $\hat{\mu}_{h_t,T_{h_t}(t)} - C_{h_t,T_{h_t}(t)} < \hat{\mu}_{\ell_t,T_{\ell_t}(t)} + C_{\ell_t,T_{\ell_t}(t)}$
4. Sample from $argmax_{i\in[n]}\ \hat{\mu}_{i,T_i(t)} + C_{i,T_i(t)}$
5. t=t+1
6. output h_t

7. Stop when $\exists i \in [n]\colon T_i(t) > \alpha \sum_{j\neq i} T_j(t)$ output $argmax_i\, T_i(t)$
8. Intuition of stop condition 2: There is an arm that was sampled relatively more than the other arms. This means that this arm had consistently highest UCB.

Lil'UCB – Jamieson et. al (2013) http://proceedings.mlr.press/v35/jamieson14.pdf

lil' UCB

input: confidence $\delta > 0$, algorithm parameters $\varepsilon, \lambda, \beta > 0$
initialize: sample each of the n arms once, set $T_i(t) = 1$ for all i and set $t = n$
while $T_i(t) < 1 + \lambda \sum_{j \neq i} T_j(t)$ for all i

sample arm

$$I_t = \underset{i \in \{1,\ldots,n\}}{\text{argmax}} \left\{ \widehat{\mu}_{i,T_i(t)} + (1+\beta)(1+\sqrt{\varepsilon})\sqrt{\frac{2\sigma^2(1+\varepsilon)\log\left(\frac{\log((1+\varepsilon)T_i(t))}{\delta}\right)}{T_i(t)}} \right\}.$$

set $T_i(t+1) = T_i(t) + 1$ if $I_t = i$, otherwise set $T_i(t+1) = T_i(t)$.
else stop and output $\arg\max_{i \in \{1,\ldots,n\}} T_i(t)$

Figure 1: The lil' UCB algorithm.

$$\mathbf{H}_1 = \sum_{i \neq i^*} \frac{1}{\Delta_i^2} \quad \text{and} \quad \mathbf{H}_3 = \sum_{i \neq i^*} \frac{\log\log_+(1/\Delta_i^2)}{\Delta_i^2}$$

$$c_1 \mathbf{H}_1 \log(1/\delta) + c_3 \mathbf{H}_3$$

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

UCB Termination (1)

Stop condition: $\exists i \in [n] : T_i(t) > \alpha \sum_{j\neq i} T_j(t)$

1. Let $C_{i,t} = (1+\beta)U(T_i(t), \delta/n)$
2. Let $\alpha = \left(\frac{2+\beta}{\beta}\right)^2 \left(1 + \frac{\log(2\log((\frac{2+\beta}{\beta})^2 n/\delta))}{\log(n/\delta)}\right)$
3. At time t, if arm $i \neq i_*$ is sampled: $i = \arg\max_{i\in[n]} \hat{\mu}_{i,T_i(t)} + (1+\beta)U(T_i(t),\delta/n)$

$$\begin{aligned}\mu_i + (2+\beta)U(T_i(t),\delta/n) \geq \hat{\mu}_{i,T_i(t)} + (1+\beta)U(T_i(t),\delta/n) &\geq \hat{\mu}_{i_*,T_{i_*}(t)} + (1+\beta)U(T_{i_*}(t),\delta/n) \\ &\geq \mu_{i_*} + \beta U(T_{i_*}(t),\delta/n)\end{aligned}$$

$$\implies (2+\beta)U(T_i(t),\delta/n) \geq \beta U(T_{i_*}(t),\delta/n) \text{ since } \mu_{i_*} > \mu_i$$

$$\begin{aligned}\implies T_i(t) &\leq T_{i_*}(t)\frac{(2+\beta)^2}{\beta^2}\frac{\log\left(2\log\left(\frac{n(2+\beta)^2}{\delta\beta^2}\right)/(\delta/n)\right)}{\log(n/\delta)} \quad \text{using (2) } t = T_i(t), s = T_{i_*}(t), \delta = \delta/n, c = \frac{\beta^2}{(2+\beta)^2} \\ &= \alpha T_{i_*}(t)\end{aligned}$$

$$\implies T_i(t) \leq \alpha \sum_{j\neq i} T_j(t)$$

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

UCB Termination (2)

Stop condition: $\exists i \in [n] : T_i(t) > \alpha \sum_{j \neq i} T_j(t)$

$$\implies T_i(t) \leq \alpha \sum_{j \neq i} T_j(t)$$

1. From stop condition, UCB won’t terminate on suboptimal arm (with probability at least …).
2. Note: if we sample i_* at time t, then only $T_{i_*}(t)$ increases and the above holds for the remaining suboptimal arms.

UCB Bound (1)

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

Stop condition: $\exists i \in [n] : T_i(t) > \alpha \sum_{j\neq i} T_j(t)$

$$\mu_i + (2+\beta)U(T_i(t),\delta/n) \geq \mu_{i_*} + \beta U(T_{i_*}(t),\delta/n)$$

$$\implies (2+\beta)U(T_i(t),\delta/n) - \beta U(T_{i_*}(t),\delta/n) \geq (\mu_{i_*} - \mu_i) = \Delta_i$$

$$\implies (2+\beta)U(T_i(t),\delta/n) \geq \Delta_i$$

$$\implies T_i(t) \leq 1 + \frac{2\gamma}{\Delta_i^2}\log\left(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/n}\right) \text{using (1) with } t = T_i(t), \delta = \delta/n, c = \frac{\Delta_i^2}{\gamma}$$

$$\gamma = (2+\beta)^2(1+\sqrt{\epsilon})^2(1+\epsilon)/2$$

1. Gives UB on suboptimal pulls. Algorithm stops when:

$$T_{i_*} = t - \sum_{i\neq i_*} T_i(t) > \alpha \sum_{i\neq i_*} T_i(t) \implies t > (1+\alpha)\sum_{i\neq i_*} T_i(t)$$

2. From before: $(1+\alpha)\sum_{i\neq i_*} T_i(t) \leq (1+\alpha)\sum_{i\neq i_*}\left(1 + \frac{2\gamma}{\Delta_i^2}\log\left(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/n}\right)\right)$
3. In other words: $t \leq O\left(\sum_{i\neq i_*} \Delta_i^{-2}\log\left(\frac{n\log(\Delta_i^{-2})}{\delta}\right)\right)$
4. Author remark: $\beta = 1.66$ optimizes bound, but smaller works in practice.

LUCB Strategy

1. Let $h_{\mathrm{t}} = argmax_{i\in[n]}\ \hat{\mu}_{i,T_i(t)}$ and $\ell_{\mathrm{t}} = argmax_{i\in[n]\setminus h_t}\ \hat{\mu}_{i,T_i(t)} + C_{i,T_i(t)}$
2. Sample from each arm $i \in \Omega$, 1 time. t=n+1
3. while $\hat{\mu}_{h_t,T_{h_t}(t)} - C_{h_t,T_{h_t}(t)} < \hat{\mu}_{\ell_t,T_{\ell_t}(t)} + C_{\ell_t,T_{\ell_t}(t)}$
4. Sample from h_t and ℓ_t
5. t=t+1
6. output h_t

Remark: Better exploration than UCB, e.g. 2-arms case.

LUCB Termination (1)

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

Stop condition: $\hat{\mu}_{h_t,T_{h_t}(t)} - C_{h_t,T_{h_t}(t)} \geq \hat{\mu}_{\ell_t,T_{\ell_t}(t)} + C_{\ell_t,T_{\ell_t}(t)}$

$$h_t = \arg\max_{i\in[n]} \hat{\mu}_{i,T_i(t)}$$

$$\ell_t = \arg\max_{i\in[n]\setminus h_t} \hat{\mu}_{i,T_i(t)} + C_{i,T_i(t)}$$

1. Let $C_{i,t} = U(T_i(t), \delta/n)$
2. At time t, if $h_t = i \neq i_*$ then:

$$\hat{\mu}_i - U(T_i(t),\delta/n) \leq \mu_i < \mu_{i_*} \leq \hat{\mu}_{i_*} + U(T_{i_*}(t),\delta/n) \leq \hat{\mu}_\ell + U(T_\ell(t),\delta/n)$$

3. From stop condition, LUCB won’t terminate on suboptimal arm (with probability at least …).

LUCB Bound (1)

$$U(t,\delta) = (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}$$

Stop condition: $\hat{\mu}_{h_t,T_{h_t}(t)} - C_{h_t,T_{h_t}(t)} \geq \hat{\mu}_{\ell_t,T_{\ell_t}(t)} + C_{\ell_t,T_{\ell_t}(t)}$

1. Define: $c = (\mu_1 + \mu_2)/2$
2. Define event: i_* is BAD if $\hat{\mu}_{i_*,T_{i_*}(t)} - U(T_{i_*}(t), \delta/n) < c$.
3. Define event: $i \neq i_*$ is BAD if $\hat{\mu}_{i,T_i(t)} + U(T_i(t), \delta/n) > c$.
4. Claim for all $t \geq 1$:

   $$\cap\{\hat{\mu}_{h_t,T_{h_t}(t)} - C_{h_t,T_{h_t}(t)} < \hat{\mu}_{\ell_t,T_{\ell_t}(t)} + C_{\ell_t,T_{\ell_t}(t)}\} \implies \{h_t \text{ is } BAD\} \cup \{l_t \text{ is } BAD\}$$

5. Proof by contradiction in appendix. $\neg(p \implies q) \equiv (p \wedge \neg q)$
6. If LUCB hasn't terminated yet, then either h_t or ℓ_t is BAD.
7. By contraposition, if both h_t and ℓ_t are NOT BAD, then LUCB has terminated. So when does this happen?

LUCB Bound (2)

Stop condition: $\hat{\mu}_{h_t,T_{h_t}(t)} - C_{h_t,T_{h_t}(t)} \geq \hat{\mu}_{\ell_t,T_{\ell_t}(t)} + C_{\ell_t,T_{\ell_t}(t)}$

$i \neq i_*$ is BAD if $\hat{\mu}_{i,T_i(t)} + U(T_i(t), \delta/n) > c$.

i_* is BAD if $\hat{\mu}_{i_*,T_{i_*}(t)} - U(T_{i_*}(t), \delta/n) < c$.

$c = (\mu_1 + \mu_2)/2$

1. Define $\tau_i = \min\{k : U(k, \delta/n) \leq \Delta_i/4\}$ for $i \neq i_*$
2. For $i \neq i_*$ and $s \geq \tau_i$:

$$\begin{aligned}
\hat{\mu}_{i,s} + U(s,\delta/n) &\leq \mu_i + 2U(s,\delta/n) \\
&= c + 2U(s,\delta/n) - \tfrac{\mu_{i_*}-\mu_i}{2} + \tfrac{\mu_i-\mu_2}{2} \\
&\leq c + 2U(s,\delta/n) - \tfrac{\Delta_i}{2} \qquad \text{using } \mu_2 \geq \mu_i \implies \mu_i - \mu_2 \leq 0 \\
&\leq c \qquad \text{using } U(s,\delta/n) \leq \Delta_i/4 \text{ for } s \geq \tau_i
\end{aligned}$$

3. So, if $T_i(t) \geq \tau_i$ then $i \neq i_*$ is NOT BAD.
4. For i_*, set $\tau_{i_*} = \tau_2$: $\hat{\mu}_{i_*,s} - U(s,\delta/n) \geq \mu_{i_*} - 2U(s,\delta/n) = c - 2U(s,\delta/n) + \frac{\Delta_2}{2}$

$$\geq c \qquad \text{using } U(s,\delta/n) \leq \Delta_2/4 \text{ for } s \geq \tau_2$$

5. So, if $T_{i_*}(t) \geq \tau_2$ then i_* is NOT BAD.

LUCB Bound (3)

Stop condition: $\hat{\mu}_{h_t,T_{h_t}(t)} - C_{h_t,T_{h_t}(t)} \geq \hat{\mu}_{\ell_t,T_{\ell_t}(t)} + C_{\ell_t,T_{\ell_t}(t)}$

$$\min\{k : \Delta_i/4 \geq U(k,\delta/n)\} \leq \frac{2\gamma}{\Delta_i^2}\log\left(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/n}\right)$$

$$\gamma = (2+\beta)^2(1+\sqrt{\epsilon})^2(1+\epsilon)/2$$

1. We want both h_t and ℓ_t NOT BAD for termination.
2. Guaranteed when all $i \neq i_*$ are NOT BAD ($T_i(t) \geq \tau_i$).

$$T_{rounds} = \sum_{t=1}^{\infty} \mathbb{1}\{h_t \text{ is } BAD \text{ or } \ell_t \text{ is } BAD\} \leq \sum_{t=1}^{\infty}\sum_{i=1}^{n} \mathbb{1}\{\{h_t = i \text{ or } \ell_t = i\} \cap \{i \text{ is } BAD\}\}$$

$$\leq \sum_{t=1}^{\infty}\sum_{i=1}^{n} \mathbb{1}\{\{h_t = i \text{ or } \ell_t = i\} \cap \{T_i(t) \leq \tau_i\}\}$$

τ_i times until $T_i(t) > \tau_i$ for each i

$$\leq \sum_{i=1}^{n} \tau_i \quad \leq \sum_{i=1}^{n} \frac{2\gamma}{\Delta_i^2}\log\left(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/n}\right)$$

3. Note we sample 2 per round. Sample complexity:

$$O\left(\sum_{i\neq i_*} \Delta_i^{-2}\log\left(\frac{n\log(\Delta_i^{-2})}{\delta}\right)\right)$$

4. Author remarks: not clear how to remove log(n) term with this approach.

Recap of Analysis

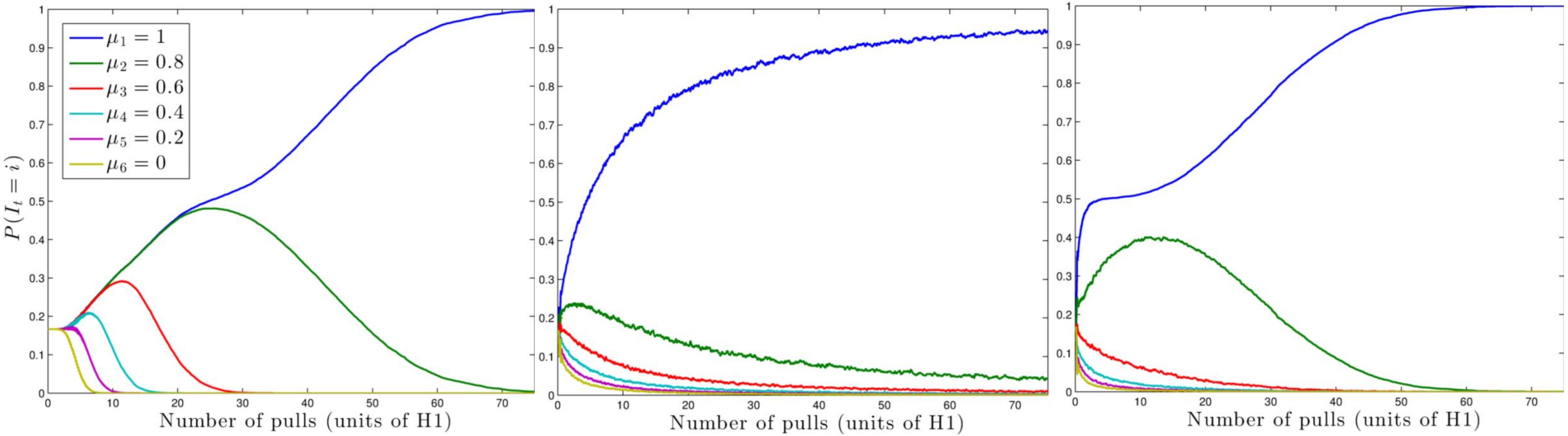
1. Three strategies have similar sample complexities: log(n) term can be negligible if n is small (becomes close to optimal complexity).
2. Using LiL Lemma gave simple proofs and similar complexities.
3. LUCB complexity improves on result of [7].

Algorithms

General Strategy	Algorithm	Sample Complexity	Year
Action Elimination (AE)	Successive elimination	$O(\sum_{i \neq i_*} \Delta_i^{-2} \log(n\Delta^{-2}))$ $\Omega(\sum_{i \neq i_*} \Delta_i^{-2})$	2002 [4] 2004 [5]
	PRISM	$O(\sum_{i \neq i_*} \Delta_i^{-2} \log\log(\sum_{j \neq i_*} \Delta_j^{-2}))$ or $O(\sum_{i \neq i_*} \Delta_i^{-2} \log(\Delta_i^{-2}))$	2013 [8]
	*Exp-gap elimination	$O(\sum_{i \neq i_*} \Delta_i^{-2} \log\log(\Delta_i^{-2}))$	2013 [9]
Upper confidence bounds (UCB)	*Lil' UCB	$O(\sum_{i \neq i_*} \Delta_i^{-2} \log\log(\Delta_i^{-2}))$ $\Omega(\sum_{i \neq i_*} \Delta_i^{-2} \log\log(\Delta_i^{-2}))$	Late 2013 [10]
Lower UCB (LUCB)	LUCB	$O(\sum_{i \neq i_*} \Delta_i^{-2} \log(\sum_{j \neq i_*} \Delta_j^{-2}))$	2012 [7] m-best arms

Experimental: Qualitative Behavior (1)

1. Setup: $n = 6$ arms, means $= \{1, 0.8, 0.6, 0.4, 0.2, 0\}$, $X_{i,s} \sim \mathcal{N}(\mu_i, 0.25), \delta = 0.1, \epsilon = 0.01$



(a) Action Elimination Sampling (b) UCB Sampling (c) LUCB Sampling

AE: drops arms from the running over time in increasing order.

UCB/LUCB identify best arm early on.

Experimental: Stopping Time Behavior (1)

1. Define **LIL Stopping** (LS) Criteria: $\hat{\mu}_{h_t,T_{h_t}(t)} - C_{h_t,T_{h_t}(t)} > \hat{\mu}_{\ell_t,T_{\ell_t}(t)} + C_{\ell_t,T_{\ell_t}(t)}$ where $C_{i,T_i(t)} = U(T_i(t), \delta/n)$. Apply to any algorithm, then outputs best arm with probability $\geq 1 - \frac{2+\epsilon}{\epsilon/2}\left(\frac{1}{\log(1+\epsilon)}\right)^{1+\epsilon}\delta$

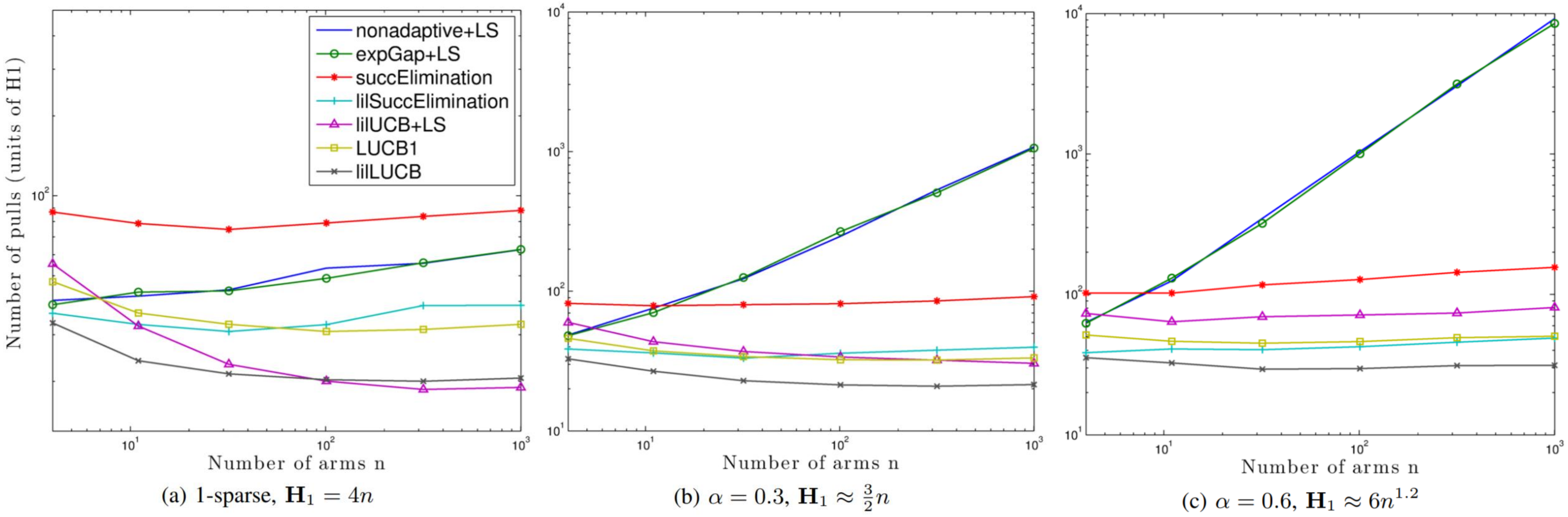
2. Algorithms:
   1. Nonadaptive+LS: randomly permute the arms, then sample in order until LS met.
   2. ***Exp-Gap Elimination (+LS):** AE that uses median elimination.
   3. Successive Elimination: AE with $C_{i,k} = \sqrt{\log(\pi^2/3nk^2/\delta)/k}$
   4. Lil'successive Elimination: AE algorithm in section 2.
   5. ***Lil'UCB (+LS):** UCB with $\beta = 1,\ \alpha = 9,\ \delta = \left(\frac{\nu\epsilon}{5(2+\epsilon)}\right)^{1/(1+\epsilon)}$ where ν is confidence.
   6. LUCB1: LUCB with $C_{i,T_i(t)}$ as in ref [12].
   7. Lil'LUCB: LUCB algorithm in section 2.

3. Complexity order: Exp-Gap=lil'UCB < lil'SE=lil'LUCB < LUCB1 < SE

Experimental: Stopping Time Behavior (2)

1. Three problems:
   1. 1-sparse with $\mu_1 = 0.25$ and $\mu_i = 0$. $H_1 = 4n$ hardness.
   2. $\alpha = 0.3$ scenario with $\mu_0 = 1$ and $\mu_i = 1 - (i/n)^\alpha$. $H_1 \approx 1.5n$ hardness.
   3. $\alpha = 0.6$ scenario with $\mu_0 = 1$ and $\mu_i = 1 - (i/n)^\alpha$. $H_1 \approx 6n^{1.2}$ hardness. (superlinear)
2. Run each algorithm 50 times on each problem with increasing n.

Experimental: Stopping Time Behavior (2)



- Exp-Gap similar to Non-Adap. due to constants in sample complexity. *See ref[9].*
- Vanilla vs. LiL versions (SE and LUCB): LiL versions better than vanilla.
- Lil'UCB+LS good for large sparse problems. But lilLUCB best overall.
- n needs to be large enough to justify lil'UCB+LS.

Main Takeaways

- Sampling strategies: AE, UCB, LUCB.
- Using LiL Lemma gave simple proofs and similar complexities.
- In practice, need to account for constants in algorithms.

2. Finite LIL Bound Lemma: see [10]

Let $X_1, X_2, \ldots$ be i.i.d $subGaus(\sigma^2)$. For any $\epsilon \in (0,1)$ and $\delta \in (0, \frac{\log(1+\epsilon)}{e})$ then $\forall t \geq 1$:

$$P\left(\sum_{s=1}^{t} X_s \leq (1+\sqrt{\varepsilon})\sqrt{2\sigma^2(1+\varepsilon)t\log\left(\frac{\log((1+\varepsilon)t)}{\delta}\right)}\right) \geq 1 - \frac{2+\varepsilon}{\varepsilon}\left(\frac{\delta}{\log(1+\varepsilon)}\right)^{1+\varepsilon}$$

Lemma: Let $X_1, X_2, \ldots$ be i.i.d zero-mean sub-Gaussian RVs with scale parameter $\sigma > 0$ and let $\delta \in (0,1)$. Then with probability at least $1 - 4\delta^2$, for all $t \geq 1$:

$$\sum_{s=1}^{t} X_s \leq 4\sigma\sqrt{t\log(\log_2(2t)/\delta)}$$

Proof: Assume $\sigma = 1$ and let $S_t = \sum_{s=1}^{t} X_s$. Recall sub-Gaussian tail bound:

$$P(\bigcup_{t=1}^{m} S_t \geq x) = P(\max_{t=1}^{m} S_t \geq x) \leq e^{-\frac{1}{2}x^2/m}$$

Now we want to show Lemma holds for all $t \geq 1$. So consider $t = 2^k$ for $k \geq 0$:

$$\begin{aligned}
P\left(\bigcup_{k\geq 0} S_{2^k} \geq 4\sqrt{2^k \log(\log_2(2^{k+1})/\delta)}\right) &\leq \sum_{k\geq 0} e^{-2\log(\log_2(2^{k+1})/\delta)} \\
&= \sum_{k\geq 0} \frac{\delta^2}{(k+1)^2} \\
&= \sum_{k\geq 0} \frac{\delta^2\pi^2}{6} \\
&\leq 2\delta^2
\end{aligned}$$

Now we look at the gaps:

$$P\left(\bigcup_{t=2^k+1}^{2^{k+1}} S_t - S_{2^k} \geq 4\sqrt{t\log(\log_2(2t)/\delta)}\right) \leq P\left(\bigcup_{t=1}^{2^k} S_t \geq 4\sqrt{2^k\log(\log_2(2^{k+1})/\delta)}\right)$$

$$= P\left(\max_{t=1}^{2^k} S_t \geq 4\sqrt{2^k\log(\log_2(2^{k+1})/\delta)}\right)$$

$$\leq e^{-2\log(\log_2(2^{k+1})/\delta)}$$

$$= \frac{\delta^2}{(k+1)^2}$$

$$\implies \sum_{k\geq 0} P\left(\bigcup_{t=2^k+1}^{2^{k+1}} S_t - S_{2^k} \geq 4\sqrt{t\log(\log_2(2t)/\delta)}\right) \leq \sum_{k\geq 0}\frac{\delta^2}{(k+1)^2}$$

$$\leq 2\delta^2$$

Adding both:

$$P\left(\bigcup_{t\geq 1} S_t \geq 4\sqrt{t\log(\log_2(2t)/\delta)}\right) \leq P\left(\bigcup_{k\geq 0} S_{2^k} \geq 4\sqrt{2^k\log(\log_2(2^{k+1})/\delta)}\right) +$$

$$\sum_{k\geq 0} P\left(\bigcup_{t=2^k+1}^{2^{k+1}} S_t - S_{2^k} \geq 4\sqrt{t\log(\log_2(2t)/\delta)}\right)$$

$$\leq 2\delta^2 + 2\delta^2 = 4\delta^2$$