APPENDIX – SUPPLEMENTARY MATERIAL
Contextual Bandit Algorithms with Supervised Learning Guarantees

Alina Beygelzimer
IBM Research
Hawthorne, NY
beygel@us.ibm.com

John Langford
Yahoo! Research
New York, NY
jl@yahoo-inc.com

Lihong Li
Yahoo! Research
Santa Clara, CA
lihong@yahoo-inc.com

Lev Reyzin
Georgia Institute of Technology
Atlanta, GA
lreyzin@cc.gatech.edu

Robert E. Schapire
Princeton University
Princeton, NJ
schapire@cs.princeton.edu

PROOF OF LEMMA 4

Recall that the estimated reward of expert i is defined as

$$\hat{G}_i \doteq \sum_{t=1}^{T} \hat{y}_i(t).$$

Also

$$\hat{\sigma}_i \doteq \sqrt{KT} + \frac{1}{\sqrt{KT}} \sum_{t=1}^{T} \hat{v}_i(t)$$

and that

$$\hat{U} = \max_i \left(\hat{G}_i + \hat{\sigma}_i \cdot \sqrt{\ln(N/\delta)} \right).$$

Lemma 4. Under the conditions of Theorem 2,

$$\begin{aligned} G_{\text{Exp4.P}} \quad \geq \quad & \left(1 - 2\sqrt{\frac{K \ln N}{T}}\right) \hat{U} - 2\sqrt{KT \ln(N/\delta)} \\ & -\sqrt{KT \ln N} - \ln(N/\delta). \end{aligned}$$

Proof. For the proof, we use $\gamma = \sqrt{\frac{K \ln N}{T}}$.
We have

$$p_j(t) \geq p_{\min} = \sqrt{\frac{\ln N}{KT}}$$

and

$$\hat{r}_j(t) \leq 1/p_{\min}$$

so that

$$\hat{y}_i(t) \leq 1/p_{\min} \qquad \text{and} \qquad \hat{v}_i(t) \leq 1/p_{\min}.$$

Thus,

$$\begin{aligned} \frac{p_{\min}}{2}\left(\hat{y}_i(t) + \sqrt{\frac{\ln(N/\delta)}{KT}} \hat{v}_i(t)\right) \quad & \leq \quad \frac{p_{\min}}{2}(\hat{y}_i(t) + \hat{v}_i(t)) \\ & \leq \quad 1. \end{aligned}$$

Let $\bar{w}_i(t) = w_i(t)/W_t$. We will need the following inequality:

Inequality 1. $\sum_i^N \bar{w}_i(t)\hat{v}_i(t) \leq \frac{K}{1-\gamma}$.

As a corollary, we have

$$\begin{aligned} \sum_i^N \bar{w}_i(t)\hat{v}_i(t)^2 \quad & \leq \quad \sum_i^N \bar{w}_i(t)\hat{v}_i(t)\frac{1}{p_{\min}} \\ & \leq \quad \sqrt{\frac{KT}{\ln N}} \frac{K}{1-\gamma}. \end{aligned}$$

Also, [1] (on p.67) prove the following two inequalities (with a typo). For completeness, the proofs of all three inequalities are given below this proof.

Inequality 2. $\sum_{i=1}^N \bar{w}_i(t)\hat{y}_i(t) \leq \frac{r_{j_t}(t)}{1-\gamma}$.

Inequality 3. $\sum_{i=1}^N \bar{w}_i(t)\hat{y}_i(t)^2 \leq \frac{\hat{r}_{j_t}(t)}{1-\gamma}$.

Now letting $b = \frac{p_{\min}}{2}$ and $c = \frac{p_{\min}\sqrt{\ln(N/\delta)}}{2\sqrt{KT}}$ we have

$$\begin{aligned}
\frac{W_{t+1}}{W_t} &= \sum_{i=1}^{N} \frac{w_i(t+1)}{W_t} \\
&= \sum_{i=1}^{N} \bar{w}_i(t) \exp\left(b\hat{y}_i(t) + c\hat{v}_i(t)\right) \\
&\leq \sum_{i=1}^{N} \bar{w}_i(t)\left[1 + b\hat{y}_i(t) + c\hat{v}_i(t)\right] \qquad (1) \\
&\quad + \sum_{i=1}^{N} \bar{w}_i(t)\left[2b^2\hat{y}_i(t)^2 + 2c^2\hat{v}_i(t)^2\right] \\
&= 1 + b\sum_{i=1}^{N} \bar{w}_i(t)\hat{y}_i(t) + c\sum_{i=1}^{N} \bar{w}_i(t)\hat{v}_i(t) \\
&\quad + 2b^2\sum_{i=1}^{N} \bar{w}_i(t)\hat{y}_i(t)^2 + 2c^2\sum_{i=1}^{N} \bar{w}_i(t)\hat{v}_i(t)^2 \\
&\leq 1 + b\frac{r_{j_t}(t)}{1-\gamma} + c\frac{K}{1-\gamma} + 2b^2\frac{\hat{r}_{j_t}(t)}{1-\gamma} \qquad (2) \\
&\quad + 2c^2\sqrt{\frac{KT}{\ln N}}\frac{K}{1-\gamma}.
\end{aligned}$$

Eq. (1) uses $e^a \leq 1 + a + (e-2)a^2$ for $a \leq 1$, $(a+b)^2 \leq 2a^2 + 2b^2$, and $e - 2 < 1$. Eq. (2) uses inequalities 1 through 3.

Now take logarithms, use the inequality $\ln(1+x) \leq x$, sum both sides over T, and we obtain

$$\begin{aligned}
\ln\left(\frac{W_{T+1}}{W_1}\right) &\leq \frac{b}{1-\gamma}\sum_{t=1}^{T} r_{j_t}(t) + c\frac{KT}{1-\gamma} \\
&\quad + \frac{2b^2}{1-\gamma}\sum_{t=1}^{T} \hat{r}_{j_t}(t) + 2c^2\sqrt{\frac{KT}{\ln N}}\frac{KT}{1-\gamma} \\
&\leq \frac{b}{1-\gamma}G_{\text{Exp4.P}} + c\frac{KT}{1-\gamma} + \frac{2b^2}{1-\gamma}K\hat{U} \\
&\quad + 2c^2\sqrt{\frac{KT}{\ln N}}\frac{KT}{1-\gamma}.
\end{aligned}$$

Here, we used

$$G_{\text{Exp4.P}} = \sum_{t=1}^{T} r_{j_t}(t)$$

and

$$\sum_{t=1}^{T} \hat{r}_{j_t}(t) = K\sum_{t=1}^{T}\frac{1}{K}\sum_{j=1}^{K}\hat{r}_j(t) \leq K\hat{G}_{\text{uniform}} \leq K\hat{U}.$$

because we assumed that the set of experts includes one who always selects each action uniformly at random.

We also have $\ln(W_1) = \ln(N)$ and

$$\begin{aligned}
\ln(W_{T+1}) &\geq \max_i\left(\ln w_i(T+1)\right) \\
&= \max_i\left(b\hat{G}_i + c\sum_{t=1}^{T}\hat{v}_i(t)\right) \\
&= b\hat{U} - b\sqrt{KT\ln(N/\delta)}.
\end{aligned}$$

Combining then gives

$$\begin{gathered}
b\hat{U} - b\sqrt{KT\ln(N/\delta)} - \ln N \\
\leq \\
\tfrac{b}{1-\gamma}G_{\text{Exp4.P}} + c\tfrac{KT}{1-\gamma} + \tfrac{2b^2}{1-\gamma}K\hat{U} + 2c^2\sqrt{\tfrac{KT}{\ln N}}\tfrac{KT}{1-\gamma}.
\end{gathered}$$

Solving for $G_{\text{Exp4.P}}$ now gives

$$\begin{aligned}
G_{\text{Exp4.P}} &\geq \left(1-\gamma-2bK\right)\hat{U} - \left(\frac{1-\gamma}{b}\right)\ln N \\
&\quad - (1-\gamma)\sqrt{KT\ln(N/\delta)} - \frac{c}{b}KT \\
&\quad - 2\frac{c^2}{b}\sqrt{\frac{KT}{\ln N}}KT \\
&\geq \left(1-\gamma-2bK\right)\hat{U} - \sqrt{KT\ln(N/\delta)} \qquad (3) \\
&\quad - \frac{1}{b}\ln N - \frac{c}{b}KT - 2\frac{c^2}{b}\sqrt{\frac{KT}{\ln N}}KT \\
&= \left(1 - 2\sqrt{\frac{K\ln N}{T}}\right)\hat{U} - \ln(N/\delta) \qquad (4) \\
&\quad - 2\sqrt{KT\ln N} - \sqrt{KT\ln(N/\delta)},
\end{aligned}$$

using $\gamma > 0$ in Eq. (3) and plugging in the definition of γ, b, c in Eq. (4). □

We prove Inequalities 1 through 3 below.

Let $\bar{w}_i(t) = w_i(t)/W_t$.

Inequality 1. $\sum_i^N \bar{w}_i(t)\hat{v}_i(t) \leq \frac{K}{1-\gamma}$.

Proof.

$$\begin{aligned}
\sum_i^N \bar{w}_i(t)\hat{v}_i(t) &= \sum_i^N \bar{w}_i(t)\sum_j^K \frac{\xi_j^i(t)}{p_j(t)} \\
&= \sum_{j=1}^{K}\frac{1}{p_j(t)}\sum_i^N \bar{w}_i(t)\xi_j^i(t) \\
&= \sum_{j=1}^{K}\frac{1}{p_j(t)}\left(\frac{p_j(t) - p_{\min}}{1-\gamma}\right) \\
&\leq \sum_{j=1}^{K}\frac{1}{1-\gamma} \\
&= \frac{K}{1-\gamma}.
\end{aligned}$$

□

Inequality 2. $\sum_{i=1}^{N} \bar{w}_i(t)\hat{y}_i(t) \le \frac{r_{j_t}(t)}{1-\gamma}$.

Proof.

$$\begin{aligned}
\sum_{i=1}^{N} \bar{w}_i(t)\hat{y}_i(t) &= \sum_{i=1}^{N} \bar{w}_i(t) \left(\sum_{j=1}^{K} \xi_j^i(t)\hat{r}_j(t) \right) \\
&= \sum_{j=1}^{K} \left(\sum_{i=1}^{N} \bar{w}_i(t)\xi_j^i(t) \right) \hat{r}_j(t) \\
&= \sum_{j=1}^{K} \left(\frac{p_j(t) - p_{\min}}{1-\gamma} \right) \hat{r}_j(t) \\
&\le \frac{r_{j_t}(t)}{1-\gamma}.
\end{aligned}$$

□

Inequality 3. $\sum_{i=1}^{N} \bar{w}_i(t)\hat{y}_i(t)^2 \le \frac{\hat{r}_{j_t}(t)}{1-\gamma}$.

Proof.

$$\begin{aligned}
\sum_{i=1}^{N} \bar{w}_i(t)\hat{y}_i(t)^2 &= \sum_{i=1}^{N} \bar{w}_i(t) \left(\sum_{j=1}^{K} \xi_j^i(t)\hat{r}_j(t) \right)^2 \\
&= \sum_{i=1}^{N} \bar{w}_i(t) \left(\xi_{j_t}^i(t)\hat{r}_{j_t}(t) \right)^2 \\
&\le \left(\frac{p_{j_t}(t)}{1-\gamma} \right) \hat{r}_{j_t}(t)^2 \\
&\le \frac{\hat{r}_{j_t}(t)}{1-\gamma}.
\end{aligned}$$

□

References

[1] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal of Computing*, 32(1):48–77, 2002.