Contextual Multi-Armed Bandits — Appendix

Tyler Lu
tl@cs.toronto.edu
Department of Computer Science
University of Toronto
10 King's College Road,
M5S 3G4 Toronto, ON, Canada

Dávid Pál
dpal@cs.ualberta.ca
Department of Computing Science
University of Alberta
T6G 2E8 Edmonton, AB, Canada

Martin Pál
mpal@google.com
Google, Inc.
76 9th Avenue, 4th Floor
New York, NY 10011, USA

A Proof of Lemma 7

Think of $v(x_0)$ being uniformly randomly chosen from Y_0 and let $\mathbf{E}$ denote the expectation with respect to both the random choice of $v(x_0)$ and the payoffs $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_M$. Clearly, the Bayes optimal payoff is

$$\begin{aligned} \mathbf{E}\left[\sum_{t=1}^{M} \sup_{y'_t \in Y} \mu_v(x_0, y'_t)\right] &= M\,\mathbf{E}\left[\sup_{y \in Y} \mu_v(x_0, y)\right] \\ &= M\,\mathbf{E}\left[\mu_v(x_0, v(x_0))\right] \\ &= M(1/2 + r). \end{aligned}$$

The non-trivial part is to upper bound the payoff of A. First, we partition the ads space Y by forming a Voronoi diagram with sites in Y_0. That is, we consider the partition $P = \{S_y \ : \ y \in Y_0\}$ where $S_y \subseteq Y$ is the set of ads which are closer to $y \in Y_0$ than to any other $y' \in Y_0$. We break ties arbitrarily, but we ensure that P is a partition of Y. Note that since Y_0 is $2r$-separated S_y contains an open ball of radius r centered at y. Also note that for any $y' \in S_y$ the highest payoff $\mu_v(x_0, y)$ is achieved at the Voronoi site y regardless of v. For $y \in Y_0$ let n_y be the random variable denoting the number of times the algorithm displays an ad from S_y.

Now, let for $y \in Y_0$ denote by $\mathbf{E}_y$ the conditional expectation $\mathbf{E}[\cdot \mid v(x_0) = y]$. The expected payoff of A can be bounded as

$$\begin{aligned} \mathbf{E}\left[\sum_{t=1}^{M} \mu_v(x_0, y_t)\right] &= \frac{1}{|Y_0|} \sum_{y \in Y_0} \mathbf{E}_y\left[\sum_{t=1}^{M} \mu_v(x_0, y_t)\right] \\ &\le \frac{1}{|Y_0|} \sum_{y \in Y_0} \mathbf{E}_y\left[\sum_{y' \in Y_0} n_{y'}\right] \\ &= \frac{1}{|Y_0|} \sum_{y \in Y_0} \mathbf{E}_y\left[M/2 + r n_y\right] \\ &= M/2 + \frac{r}{|Y_0|} \sum_{y \in Y_0} \mathbf{E}_y\, n_y \end{aligned}$$

Hence,

$$\mathcal{R}_{x_0} \ge r\left(M - \frac{1}{|Y_0|} \sum_{y \in Y_0} \mathbf{E}_y\, n_y\right) \tag{1}$$

and the proof reduces to bounding $\mathbf{E}_y\, n_y$ from above. We do this by comparing the behavior of A on an "completely noisy" independent instance μ' for which $\mu'(x_0, y) = 1/2$ and the payoffs $\hat{\mu}'_1, \hat{\mu}'_2, \ldots, \hat{\mu}'_M$ are i.i.d. Bernoulli random variables with parameter $1/2$ and are independent from $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_M$, $y_1, y_2, \ldots, y_M$ and $v(x_0)$. We denote by $y'_1, y'_2, \ldots, y_M$ the random variables denoting the ads displayed on μ'. For $y \in Y_0$ let n'_y be a random variable denoting the number of times algorithm A displays an ad from S_y for the noisy instance μ'.

For fixed $y \in Y_0$ we define two probability distributions, q and q', over $\{0,1\}^M$ as follows. For any $B = (b_1, b_2, \ldots, b_M) \in \{0,1\}^M$ let

$$\begin{aligned} &q'(B) = 2^{-M} = \\ &\qquad \Pr[\hat{\mu}'_1 = b_1, \hat{\mu}'_2 = b_2, \ldots, \hat{\mu}'_M = b_M \mid v(x_0) = y] \end{aligned}$$

and

$$q(B) = \Pr[\hat{\mu}_1 = b_1, \hat{\mu}_2 = b_2, \ldots, \hat{\mu}_M = b_M \mid v(x_0) = y].$$

Note that the sequence of payoffs received by the algorithm uniquely determines its behavior and hence for any $y \in Y_0$,

$$\begin{aligned} &\mathbf{E}_y[n_y \mid \hat{\mu}_1 = b_1, \hat{\mu}_2 = b_2, \ldots, \hat{\mu}_M = b_M] \\ &\qquad = \mathbf{E}[n'_y \mid \hat{\mu}'_1 = b_1, \hat{\mu}'_2 = b_2, \ldots, \hat{\mu}'_M = b_M] \end{aligned}$$

Consider, for any $y \in Y_0$,

$$
\begin{aligned}
&\mathbf{E}\, n'_y - \mathbf{E}_y\, n_y = \\
&\sum_{B \in \{0,1\}^M} q(B)\, \mathbf{E}_y[n_y \mid \hat{\mu}_1 = b_1, \ldots, \hat{\mu}_M = b_M] \\
&\qquad - \sum_{B \in \{0,1\}^M} q'(B)\, \mathbf{E}[n'_y \mid \hat{\mu}'_1 = b_1, \ldots, \hat{\mu}'_M = b_M] \\
&= \sum_{B \in \{0,1\}^M} (q(B) - q'(B))\, \mathbf{E}_y[n_y \mid \\
&\qquad\qquad \hat{\mu}_1 = b_1, \ldots, \hat{\mu}_M = b_M] \\
&\leq \sum_{\substack{B \in \{0,1\}^M \\ q(B) > q'(B)}} (q(B) - q'(B))\, \mathbf{E}_y[n_y \mid \\
&\qquad\qquad \hat{\mu}_1 = b_1, \ldots, \hat{\mu}_M = b_M] \\
&\leq M \sum_{\substack{B \in \{0,1\}^M \\ q(B) > q'(B)}} (q(B) - q'(B)) \\
&= \frac{M}{2} \sum_{B \in \{0,1\}^M} |q(B) - q'(B)| \qquad (2)
\end{aligned}
$$

where the last inequality follows from that $n_y \leq M$. The last expression is $M/2$ times the so-called *total variation* (or L_1) distance between the distributions q, q'. It may be bounded by Pinsker's inequality [Cover and Thomas, 2006, Lemma 11.6.1] which states that

$$\sum_{B \in \{0,1\}^M} |q(B) - q'(B)| \leq \sqrt{2D(q' \| q)}\,, \qquad (3)$$

where

$$D(q' \| q) = \sum_{B \in \{0,1\}^m} q'(B) \ln \left(\frac{q'(B)}{q(B)} \right)$$

is the Kullback-Leibler divergence of the distributions q' and q.

We use the chain rule to compute $D(q' \| q)$. First, for a sequence $B = (b_1, b_2, \ldots, b_{t-1}) \in \{0,1\}^{t-1}$, $1 \leq t \leq M$, and $b \in \{0,1\}$ we denote by

$$
\begin{aligned}
q_t(b|B) = \Pr[&\hat{\mu}_t = b \mid \\
&\hat{\mu}_1 = b_1, \ldots, \hat{\mu}_{t-1} = b_{t-1}, v(x_0) = y]
\end{aligned}
$$

and

$$
\begin{aligned}
q'_t(b|B) = \Pr[&\hat{\mu}'_t = b \mid \\
&\hat{\mu}'_1 = b_1, \ldots, \hat{\mu}'_{t-1} = b_{t-1}, v(x_0) = y]
\end{aligned}
$$

the conditional distributions of t-th payoffs $\hat{\mu}_t$ and $\hat{\mu}'_t$. Note that the event $\hat{\mu}_1 = b_1, \hat{\mu}_2 = b_2, \ldots, \hat{\mu}_{t-1} = b_{t-1}$ on which we are conditioning, is determined by B and in turn this event determines the ad y_t that A displays in t-round on the instances μ_v. We write y_t as $y_t(B)$ to stress this dependence. Hence, by the chain rule

$$
\begin{aligned}
D(q' \| q) &= \sum_{t=1}^{M} \frac{1}{2^{t-1}} \sum_{B \in \{0,1\}^{t-1}} D(q'_t(\cdot|B) \| q_t(\cdot|B)) \\
&= \sum_{t=1}^{M} \frac{1}{2^{t-1}} \left(\sum_{\substack{B \in \{0,1\}^{t-1} \\ y_t(B) \in S_{v(x_0)}}} D(q'_t(\cdot|B) \| q_t(\cdot|B)) \right. \\
&\qquad \left. + \sum_{\substack{B \in \{0,1\}^{t-1} \\ y_t(B) \notin S_y}} D(q'_t(\cdot|B) \| q_t(\cdot|B)) \right)
\end{aligned}
$$

where we have split the inner sum into two cases: (i) the ad $y_t(B)$ lies near the "correct" ad y, that is, $y_t(B) \in S_y$ and (ii) the ad y_t does not lie near the "correct" ad, that is, $y_t(B) \notin S_y$.

The second inner sum in the last expression evaluates to zero, since when $y_t(B) \notin S_y$, $q_t(\cdot|B) = q'_t(\cdot|B) = 1/2$ are the same Bernoulli distribution and thus we have $D(q'_t(\cdot|B) \| q_t(\cdot|B)) = 0$. The terms of the first inner sum can be bounded if we realize that $q_t(\cdot|B)$ is a Bernoulli distribution with parameter $1/2 + s$ where $s = \max\{0, r - L_Y(y_t, y)\} \leq r$ and $q'_t(\cdot|B)$ is a Bernoulli distribution with parameter $1/2$. Hence, for B for which $y_t \in S_y$

$$
\begin{aligned}
D(q'_t(\cdot|B) \| q_t(\cdot|B)) &= \frac{1}{2} \ln \frac{1/2}{1/2 + s} + \frac{1}{2} \ln \frac{1/2}{1/2 - s} \\
&= -\frac{1}{2} \ln(1 - 4s^2) \\
&\leq 8 \ln(4/3) s^2 \\
&\leq 8 \ln(4/3) r^2\,,
\end{aligned}
$$

where used the inequality $-\ln(1 - x) \leq 4 \ln(4/3) x$ for $x \in [0, 1/4]$ which can be proved by checking it for the left and the right end point of the interval and using the convexity of logarithm. We can guarantee that $r \in [0, 1/4]$ by picking T_0 big enough.

$$D(q' \| q) \leq 8r^2 \ln \left(\frac{4}{3} \right) \sum_{t=1}^{M} \frac{1}{2^{t-1}} \sum_{B \in \{0,1\}^{t-1}} \mathbf{1}\{y_t(B) \in S_y\} \qquad (4)$$

where $\mathbf{1}\{\cdot\}$ is an indicator function.

We combine (2), Pinsker's inequality (3) and the inequality

(4) we have just obtained, and we have

$$\left(\frac{1}{|Y_0|}\sum_{y\in Y_0}\mathbf{E}_y\, n_y\right) - \frac{M}{|Y_0|}$$
$$= \frac{1}{|Y_0|}\sum_{y\in Y_0}\left(\mathbf{E}_y\, n_y - \mathbf{E}\, n'_y\right)$$
$$\leq \frac{M}{2}\frac{1}{|Y_0|}\sum_{y\in Y_0}\sqrt{2D(q\|q')}$$
$$\leq \frac{M}{2}\frac{1}{|Y_0|}\sum_{y\in Y_0}$$
$$\sqrt{16\ln\left(\frac{4}{3}\right)r^2\sum_{t=1}^{M}\frac{1}{2^{t-1}}\sum_{B\in\{0,1\}^{t-1}}\mathbf{1}\{y_t(B)\in S_y\}}$$
$$\leq \frac{M}{2}$$
$$\sqrt{\frac{16\ln\left(\frac{4}{3}\right)r^2}{|Y_0|}\sum_{\substack{t=1\\ y\in Y_0}}^{M}\frac{1}{2^{t-1}}\sum_{B\in\{0,1\}^{t-1}}\mathbf{1}\{y_t(B)\in S_y\}}$$

(by the arithmetic and quadratic mean inequality)

$$= Mr\sqrt{\frac{4\ln(4/3)}{|Y_0|}\sum_{\substack{t=1\\ B\in\{0,1\}^{t-1}\\ y\in Y_0}}^{M}\frac{\mathbf{1}\{y_t(B)\in S_y\}}{2^{t-1}}}$$
$$= Mr\sqrt{\frac{4\ln(4/3)}{|Y_0|M}}$$

where the last equality follows since

$$\sum_{\substack{y\in Y_0\\ B\in\{0,1\}^{t-1}}}\mathbf{1}\{y_t(B)\in S_y\} = 2^{t-1}.$$

Therefore, combining with (1) we have

$$\mathcal{R}_{x_0} \geq r\left(M\left(1-\frac{1}{|Y_0|}\right) - M^{3/2}r\sqrt{\frac{4\ln(4/3)}{|Y_0|}}\right).$$

It can be easily verified that $r = \alpha C\sqrt{|Y_0|/M}$ for some constant C lying in the interval $I = [1/(2\sqrt{cd}), 2/\sqrt{cd}]$ provided T_0 is big enough. Substituting that for r leads to

$$\mathcal{R}_{x_0} \geq \left(\left(1-\frac{1}{|Y_0|}\right)C\alpha - C^2\alpha^2\sqrt{4\ln(4/3)}\right)\sqrt{M|Y_0|}\,.$$

If $\alpha > 0$ is chosen small enough, $|Y_0| \geq 2$ and $\beta = \min_{C\in I}\left(1-\frac{1}{|Y_0|}\right)C\alpha - C^2\alpha^2\sqrt{4\ln(4/3)}$ is positive. This finishes the proof.

References

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Willey & Sons, 2nd edition edition, 2006.