STATS 710 – Seq. Dec. Making with mHealth Applications — Sep 27, 2016

Lecture 7: Finishing the proof of the regret bound for Thompson Sampling with Beta Priors and Bernoulli Rewards

Instructors: Susan Murphy and Ambuj Tewari — *Scribe: Yutong Wang*

1 Proof of theorem 1 from lecture 6 continued

We adopt all notations from the previous lecture. Recall that we ended with bounding the expected number of times that a suboptimal arm a is played:

$$\mathbb{E}\left[N_T(a)\right] = \sum_{t=1}^{T} \underbrace{\mathbb{P}\left(A_t = a, E^{\mu}_{a,t}, E^{\theta}_{a,t}\right)}_{S_1} + \underbrace{\mathbb{P}\left(A_t = a, E^{\mu}_{a,t}, \overline{E^{\theta}_{a,t}}\right)}_{S_2} + \underbrace{\mathbb{P}\left(A_t = a, \overline{E^{\mu}_{a,t}}\right)}_{S_3}$$

1.1 Bounding S_1

Last lecture gave a bound for S_1. Summarizing, we had

$$S_1 \le \mathbb{E}\left[\sum_{k=0}^{T-1} \frac{1-p_{a,\tau_{k+1}}}{p_{a,\tau_{k+1}}}\right]$$

and the following lemma

Lemma 1. *Let τ_k be the time at which the arm a is pulled for the k^{th} time. Then*

$$\mathbb{E}\left[\frac{1-p_{a,\tau_{k+1}}}{p_{a,\tau_{k+1}}}\right] \le \begin{cases} \frac{3}{\Delta'_a} & \text{if } k < \frac{8}{\Delta'_a} \\ \Theta\left(e^{-k\Delta'^2_a/2} + \frac{1}{(k+1)\Delta'^2_a} e^{-kD_a} + \frac{1}{e^{k\Delta'^2_a/4}-1}\right) & \text{otherwise} \end{cases}$$

where $\Delta'_a = \mu_0 - y_a$ and $D_a = d(y_a, \mu_0)$.

Since we are not proving lemma 1, we will give some heuristics as to why it is true. Recall that $p_{a,\tau_{k+1}} = \mathbb{P}\left(\theta_{0,\tau_{k+1}} > y_a \mid \mathcal{H}_t\right)$. If $\theta_{0,\tau_{k+1}}$ were to be replaced by the sample mean $\hat{\mu}_{0,k}$, then Chernoff bound would imply that $p_{a,\tau_{k+1}} \le 1 - e^{-kC}$ (nevermind the fact that τ_{k+1} is random). Furthermore, $\frac{e^{-kC}}{1-e^{-kC}} \approx e^{-kC}$ for $k \gg 0$. This gives some heuristics as to why the terms inside the Θ in lemma 1 should have this particular form.

1.2 Bounding S_2

To bound S_2 and S_3, we first prove the following lemmas.

Lemma 2.

$$\sum_{t=1}^{T} \mathbb{P}\left(A_t = a, E^{\mu}_{a,t}, \overline{E^{\theta}_{a,t}}\right) \le L_a(T) + 1$$

where $L_a(T) = \frac{\log T}{d(x_a, y_a)}$.

From [AG12]: "This follows from the observation that $\theta_{a,t}$ is well-concentrated around its mean when $N_t(a)$ is large, that is, larger than $L_a(T)$". We will formally illustrate this in the proof.

Proof. First, we make sure that we have taken enough samples of a:

$$\sum_{t=1}^{T} \mathbb{P}\left(A_t = a, E^{\mu}_{a,t}, \overline{E^{\theta}_{a,t}}\right) \leq L_a(T) + \sum_{t=1}^{T} \mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}}, E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T)\right). \tag{1}$$

For each term in the sum in the RHS above, we condition on the history

$$\mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}}, E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T)\right) = \mathbb{E}\left[\mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}}, E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T) \mid \mathcal{H}_t\right)\right] \tag{2}$$

Using the Bayes rule

$$\begin{aligned}
&\mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}}, E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T) \mid \mathcal{H}_t\right) \\
=&\mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}} \mid E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T), \mathcal{H}_t\right) \mathbb{P}\left(E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T) \mid \mathcal{H}_t\right),
\end{aligned}$$

the fact that $\mathbb{P}\left(E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T) \mid \mathcal{H}_t\right) \leq 1$ and (2), we get

$$\mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}}, E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T)\right) \leq \mathbb{E}\left[\mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}} \mid E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T), \mathcal{H}_t\right)\right].$$

We now claim that

$$\mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}} \mid E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T), \mathcal{H}_t\right) \leq \frac{1}{T}. \tag{3}$$

Given the claim, the RHS of (1) $\leq L_a(T) + \sum_{t+1}^{T} 1/T = L_a(T) + 1$, which proves the lemma. So it remains to show that (3) holds. Now,

$$\begin{aligned}
&\mathbb{P}\left(A_t = a, \overline{E^{\theta}_{a,t}} \mid E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T), \mathcal{H}_t\right) && (4)\\
\leq&\mathbb{P}\left(\overline{E^{\theta}_{a,t}} \mid E^{\mu}_{a,t}, N_{t-1}(a) > L_a(T), \mathcal{H}_t\right) && (5)\\
=&\mathbb{P}\left(\theta_{a,t} > y_a \mid \hat{\mu}_{a,t-1} \leq x_a, N_{t-1}(a) > L_a(T), \mathcal{H}_t\right) \quad \text{by definition of } E^{\theta}_{a,t} \text{ and } E^{\mu}_{a,t} && (6)\\
=&\mathbb{P}\left(\theta_{a,t} > y_a \mid \underbrace{S_{a,t-1} \leq x_a(N_{t-1}(a)+1), N_{t-1}(a) > L_a(T), \mathcal{H}_t}_{=:\bigstar}\right) \quad \text{by definition of } \hat{\mu}_{a,t-1} && (7)\\
=&\mathbb{P}\left(Beta(S_{a,t-1}+1, N_{t-1}(a) - S_{a,t-1}+1) > y_a \mid \bigstar\right) && (8)
\end{aligned}$$

where the last equality follows from the definition of $\theta_{a,t}$, and $Beta(\alpha, \beta) > y_a$ denotes the event that a $Beta(\alpha, \beta)$ distributed random variable is greater than y_a. We have a general fact about the Beta distribution:

Lemma 3. *For $\alpha' > \alpha$ and $y \in [0, 1]$, we have*

$$\mathbb{P}\left(Beta(\alpha, C - \alpha) > y\right) \leq \mathbb{P}\left(Beta(\alpha', C - \alpha') > y\right)$$

Proof. Recall the identity

$$1 - \mathrm{CDF}_{Beta(\alpha,\beta)}(y) = \mathrm{CDF}_{Binom(\alpha+\beta-1,y)}(\alpha - 1). \tag{9}$$

Using the identity, we have

$$\begin{aligned}
\mathbb{P}\left(Beta(\alpha, C-\alpha) > y\right) =& 1 - \mathrm{CDF}_{Beta(\alpha,C-\alpha)}(y) \\
=& \mathrm{CDF}_{Binom(C-1,y)}(\alpha - 1) \\
\le& \mathrm{CDF}_{Binom(C-1,y)}(\alpha' - 1) \\
=& 1 - \mathrm{CDF}_{Beta(\alpha',C-\alpha')}(y) = \mathbb{P}\left(Beta(\alpha', C-\alpha') > y\right)
\end{aligned}$$

This proves lemma 3. □

Going back to (8), we have

$$\begin{aligned}
&\mathbb{P}\left(Beta(S_{a,t-1}+1, N_{t-1}(a) - S_{a,t-1}+1) > y_a \mid \bigstar\right) && (10)\\
\le& \mathbb{P}\left(Beta(x_a(N_{t-1}(a)+1)+1, (1-x_a)(N_{t-1}(a)+1)) > y_a \mid \bigstar\right) \quad \because \text{lemma 3}, && (11)\\
=& \mathrm{CDF}_{Binom(N_{t-1}(a)+1,y_a)}(x_a(N_{t-1}(a)+1)) \quad \because \text{identity (9)}. && (12)
\end{aligned}$$

To bound the quantity above, we need the following lemma.

Lemma 4 (KL-divergence version of Chernoff-Hoeffding)**.** *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables. Let $p_i = \mathbb{E}[X_i]$, $X = \frac{1}{n}\sum_{i=1}^n X_i$, $\mu = \mathbb{E}[X]$. Then*

$$\begin{aligned}
\mathbb{P}\left(X \ge \mu + \lambda\right) \le \exp\left(-nd(\mu+\lambda, \mu)\right), \quad \forall 0 < \lambda \le 1 - \mu. \\
\mathbb{P}\left(X \le \mu - \lambda\right) \le \exp\left(-nd(\mu-\lambda, \mu)\right), \quad \forall 0 < \lambda < \mu
\end{aligned}$$

where $d(a,b) = a\log(\frac{a}{b}) + (1-a)\log(\frac{1-a}{1-b})$.

See the supplementary material of [AG13]. Now continuing from (12)

$$\begin{aligned}
\mathrm{CDF}_{Binom(N_{t-1}(a)+1,y_a)}(x_a(N_{t-1}(a)+1)) \le& e^{-(N_{t-1}(a)+1)d(x_a,y_a)} \quad \because \text{lemma 4} \\
\le& e^{-L_a(T)d(x_a,y_a)} \quad \because N_{t-1}(a) > L_a(T) \text{ by } \bigstar \\
=& \frac{1}{T} \quad \because \text{definition of } L_a(T).
\end{aligned}$$

This proves (3) which concludes the proof of lemma 2. □

1.3 Bounding S_3

Lemma 5.

$$\sum_{t=1}^{T} \mathbb{P}\left(A_t = a, \overline{E^{\mu}_{a,t}}\right) \le 1 + \frac{1}{d(x_a, y_a)}.$$

Proof. Define τ_k as the time index at which action a is taken for the k-th time and $\tau_0 = 0$.

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{P}\left(A_t = a, \overline{E^\mu_{a,t}}\right) &\leq \mathbb{E}\left[\sum_{k=0}^{T-1} \sum_{t=\tau_k+1}^{\tau_{k+1}} \mathbb{1}(A_t = a)\mathbb{1}(\overline{E^\mu_{a,t}})\right] \quad \because \tau_T \geq T \\
&= \mathbb{E}\left[\sum_{k=0}^{T-1} \mathbb{1}(\overline{E^\mu_{a,\tau_{k+1}}})\right] \\
&\leq 1 + \mathbb{E}\left[\sum_{k=1}^{T-1} \mathbb{1}(\overline{E^\mu_{a,\tau_{k+1}}})\right] \quad \text{(Shifting the start of the summation)} \\
&\leq 1 + \sum_{k=1}^{T-1} e^{-kd(x_a,\mu_a)} \quad \because \text{lemma 4 and } \overline{E^\mu_{a,\tau_{k+1}}} = \{\hat{\mu}_{a,\tau_{k+1}} > x_a\} \\
&\leq 1 + \frac{e^{-d(x_a,\mu_a)}}{1 - e^{-d(x_a,\mu_a)}} \quad \because \text{geometric series} \\
&\leq 1 + \frac{1}{d(x_a,\mu_a)} \quad \because \frac{e^{-x}}{1-e^{-x}} \leq \frac{1}{x} \iff e^x \geq 1 + x
\end{aligned}$$

□

1.4 Putting it all together

$$\begin{aligned}
\mathbb{E}\left[N_T(a)\right] \leq \frac{24}{\Delta'^2_a} + \sum_{j=0}^{T-1} \Theta\left(e^{-j\Delta'^2_a/2} + \frac{1}{(j+1)\Delta'^2_a} e^{-jD_a} + \frac{1}{e^{j\Delta'^2_a/4} - 1}\right) \quad \because \text{lemma 1} \\
+ \frac{\log T}{d(x_a, y_a)} + 1 \quad \because \text{lemma 2} \\
+ \frac{1}{d(x_a, y_a)} + 1 \quad \because \text{lemma 5} \\
\leq (1+\epsilon)^2 \frac{\log T}{d(\mu_a, \mu_0)} + O\left(\frac{1}{\epsilon^2}\right) \quad \because d(x_a, y_a) = \frac{d(\mu_a, \mu_0)}{(1+\epsilon)^2} \text{ by construction.}
\end{aligned}$$

where the $O\left(\frac{1}{\epsilon^2}\right)$ term hides the "constants" that depend on the distribution.

References

[AG12] Shipra Agrawal and Navin Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. In *COLT*, pages 39–1, 2012.

[AG13] Shipra Agrawal and Navin Goyal. Further optimal regret bounds for Thompson sampling. In *AISTATS*, pages 99–107, 2013.