CS229T/STATS231: Statistical Learning Theory

Lecturer: Tengyu Ma
Scribe: Chris Lin, Anand Avati

Lecture 20
December 05, 2018

1 Review and Overview

In the last lecture, we introduced the general bandit problem in the Bayesian setting. Recall the following problem setup.

- Let Θ be the model parameter space and θ^* the ground truth model parameter. Furthermore, let Q be the prior distribution of θ^*, i.e., $\theta^* \sim Q$. Unless otherwise stated, we assume Θ to be a finite set.
- Let $\mathcal{A}$ be the action space. Unless otherwise stated, we assume $\mathcal{A}$ to be a finite set.
- Let $D(a, \theta)$ be the distribution of loss for $a \in \mathcal{A}$ and $\theta \in \Theta$.
- At iteration t, if action A_t is played, we observe the loss $L_{A_t,\theta^*} \sim D(A_t, \theta^*)$. For notational ease, we define $L_t \triangleq L_{A_t,\theta^*}$. The uppercase notation of A_t emphasizes that A_t is a random variable.
- The optimal action is a function $a^* : \Theta \to \mathcal{A}$ defined as

  $$a^*(\theta) = \operatorname*{argmin}_{a \in \mathcal{A}} \mathop{\mathbb{E}}_{L \sim D(a,\theta)}[L].$$

  With this, we also define the optimal action at the ground truth parameter

  $$A^* = a^*(\theta^*).$$

- Let $A_1, ..., A_T$ be the actions played at iterations $1, ..., T$. Define the regret as

  $$\text{regret} = \sum_{t=1}^{T} \mathbb{E}\left[L_t - L_{A^*,\theta^*}\right]$$

  where the expectation is taken over all the random variables. In the Bayesian setting, the ground truth parameter θ^*, the actions $A_1, ..., A_T$, and the losses $L_{A^*,\theta^*}, L_1, ..., L_T$ are all random variables.

We also described Thompson sampling for the general bandit problem in the Bayesian setting. At iteration t, the algorithm proceeds as:

0. Define f_{t-1} as the collection of random variables observed so far. That is

   $$f_{t-1} = \{A_1, L_1, ..., A_{t-1}, L_{t-1}\}.$$

1. Compute

   $$P_t(\theta) = \Pr(\theta^* = \theta | f_{t-1})$$

   which is the posterior distribution of $\theta^* | f_{t-1}$.

2. Sample θ_t from P_t.

3. Play the action $a^*(\theta_t)$.

Our goal is to provide a bound on the regret of Thompson sampling. To this end, we introduced some background on information theory. Let $\mathcal{X}, \mathcal{Y}$ be finite sets, and X, Y be random variables over $\mathcal{X}, \mathcal{Y}$, respectively. We have the following definitions.

- The entropy of X is

$$H(X) = -\sum_{x \in \mathcal{X}} \Pr(X = x) \log \Pr(X = x).$$

- The conditional entropy of $X|Y$ is

$$H(X|Y) = \sum_{y \in \mathcal{Y}} H(X|Y = y) \Pr(Y = y)$$

- The mutual information of X and Y is

$$I(X;Y) = H(X) - H(X|Y)$$

We also have the following facts.

Fact 1. $0 \leq H(X) \leq \log |X|$.

Fact 2. *(Properties of entropy and mutual information)*

- $H(X|Y) = H((X,Y)) - H(Y)$.
- $I(X;Y) = I(Y;X) = H(X) + H(Y) - H((X,Y))$.
- $I(X;Y) \geq 0$, *which is equivalent to* $H(X|Y) \leq H(X)$.
- $I(X;Y) \leq H(X)$, *which is equivalent to* $H(X|Y) \geq 0$.
- $I(X;Y) = 0$ *if and only if* X *and* Y *are independent.*

In this lecture, we will finish our background introduction to information theory (Section 2), provide a bound on the regret of Thompson sampling and apply the bound to specific examples (Section 3), as well as summarize the course (Section 4).

2 Information Theory (Continued from Lecture 19)

We have the following definition for and a fact about conditional mutual information.

Definition 2.1. *Let* X, Y, Z *be random variables. The conditional mutual information of* X, Y *given* Z *is*

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z).$$

Fact 3. *(Chain rule for conditional mutual information)*
Let $X, Y_1, ..., Y_n$ *be random variables. Then*

$$I(X;(Y_1,...,Y_n)) = I(X;Y_1) + I(X;Y_2|Y_1) + \cdots + I(X;Y_n|Y_1,...,Y_{n-1}).$$

Proof. By applying the definition of conditional mutual information to each term on the RHS, we obtain

$$\begin{aligned}
&I(X;Y_1) + I(X;Y_2|Y_1) + \cdots + I(X;Y_n|Y_1,...,Y_{n-1}) \\
&= [H(X) - H(X|Y_1)] + [H(X|Y_1) - H(X|Y_1,Y_2)]) + \cdots \\
&+ [H(X|Y_1,...,Y_{n-1}) - H(X|Y_1,...,Y_n)] \\
&= H(X) - H(X|Y_1,...,Y_n) \\
&= I(X;(Y_1,...,Y_n))
\end{aligned}$$

where the second equality comes from recognizing a telescoping sum, and the last equality from the definition of mutual information. □

It would be useful to relate mutual information to other distribution distance measures. We have the relationship between mutual information and KL divergence in the following fact.

Fact 4. *Let X, Y be random variables, and P_Y, P_X their respective marginal distributions. Then*

$$I(X;Y) = \mathop{\mathbb{E}}_{y\sim P_Y}[KL(P_{X|Y=y}||P_X)]$$

Proof. See Appendix C of [1]. □

Finally, we finish our background introduction to information theory with a theorem on sequence of random variables.

Theorem 2.1. *(Data processing inequality)*
Suppose the sequence of random variables $X \to Y \to Z$ is a Markov chain. Then

$$I(X;Y) \geq I(X;Z).$$

Proof. By applying the chain rule for conditional mutual information in different orders, we get

$$\begin{aligned}
I(X;(Y,Z)) &= I(X;Y) + I(X;Z|Y) &(1)\\
&= I(X;Z) + I(X;Y|Z). &(2)
\end{aligned}$$

By the definition of a Markov chain, we note that X and Z are independent conditioned on Y. Therefore $I(X;Z|Y) = 0$. By the properties of mutual information (Fact 2), we know that $I(X;Y|Z) \geq 0$. With these, comparing (1) and (2), we get

$$I(X;Y) \geq I(X;Z).$$

□

Remark 2.1.1. *This theorem is not necessary for proving the results in this lecture. It is discussed here for some more exposure to information theory.*

Remark 2.1.2. *For a particular joint distribution of (X,Y), we can have $I(X;Y) \geq cI(X;Z)$ where $c > 1$ is a constant. This is called the strong data processing inequality. The stronger bound is a result of knowing more about the random variable distribution.*

3 Regret Bound for Thompson Sampling

Armed with the above background on information theory, we now proceed to discuss the main goal of this lecture: providing a regret bound for Thompson sampling. The problem setup, notations, and definitions from Section 1 are carried forward in this section.

In this section, we proceed in the following way. First, we define the notion of information ratio, which can be related to the regret such that bounding the information ratio implies bounding the regret. Second, we provide a bound for the information ratio, thereby bounding the regret. Finally, we apply these results to bound the regret for the multi-armed bandit problem and linear bandit problem in the Bayesian setting.

We begin with defining the information ratio.

Definition 3.1. *(Information ratio)*
With the setup in Section 1, the information ratio at iteration t is

$$\Gamma_t = \frac{[\mathbb{E}[L_t - L_{A^*,\theta^*}|f_{t-1}]]^2}{I(A^*;(A_t,L_t)|f_{t-1})}.$$

Remark 3.0.1. *The numerator is the square of the regret term at iteration t, which is a random variable. The denominator is the conditional mutual information between the optimal action A^* and (A_t, L_t), conditioned on the previously observed actions and losses. Note that the denominator is a scalar.*

Remark 3.0.2. *Intuitively, the denominator is the amount of information gained about A^* from observing A_t and L_t. Therefore, a small information ratio is desirable. With a small information ratio, if an algorithm suffers a lot of regret, it also gains more information about the optimal action.*

With the notion of information ratio, we now discuss the main result of this lecture.

Theorem 3.1. *If for all $t = 1, ..., T$, $\Gamma_t \leq \Gamma$ almost surely for some constant Γ, then*

$$\text{regret} \leq \sqrt{\Gamma \cdot H(A^*) \cdot T}.$$

Proof.

$$\begin{aligned}
\text{regret} &= \sum_{t=1}^{T} \mathbb{E}[L_t - L_{A^*,\theta^*}] \\
&= \sum_{t=1}^{T} \mathop{\mathbb{E}}_{f_{t-1}} \Big[\mathbb{E}[L_t - L_{A^*,\theta^*}|f_{t-1}] \Big] \text{ (law of total expectation)} \\
&\leq \sum_{t=1}^{T} \mathop{\mathbb{E}}_{f_{t-1}} \Big[\Gamma^{\frac{1}{2}} \cdot I(A^*;(A_t,L_t)|f_{t-1})^{\frac{1}{2}} \Big] \text{ (definition of } \Gamma_t \text{, and } \Gamma_t \leq \Gamma) \\
&\leq \Gamma^{\frac{1}{2}} \sum_{t=1}^{T} \Big[\mathbb{E}[I(A^*;(A_t,L_t)|f_{t-1})] \Big]^{\frac{1}{2}} \text{ (Cauchy-Schwarz inequality)} \\
&\leq \Gamma^{\frac{1}{2}} \cdot T^{\frac{1}{2}} \left(\sum_{t=1}^{T} \mathbb{E}[I(A^*;(A_t,L_t)|f_{t-1})] \right)^{\frac{1}{2}} \text{ (Cauchy-Schwarz inequality)}
\end{aligned}$$

$$= \Gamma^{\frac{1}{2}} \cdot T^{\frac{1}{2}} \left(\sum_{t=1}^{T} I(A^*; (A_t, L_t) | f_{t-1}) \right)^{\frac{1}{2}} \text{ (the mutual information is a scalar)}$$
$$= \Gamma^{\frac{1}{2}} \cdot T^{\frac{1}{2}} \cdot I(A^*; (A_1, L_1, ..., A_T, L_T))^{\frac{1}{2}} \text{ (chain rule)}$$
$$\leq \Gamma^{\frac{1}{2}} \cdot T^{\frac{1}{2}} \cdot H(A^*)^{\frac{1}{2}} \text{ (Fact 2)}$$

□

This theorem shows that the regret can be bounded based on a bound on the information ratio. We will then discuss two lemmas that lead to a useful bound on the information ratio. For the remaining exposition, we simplify the notations in the following way: let p_t be the distribution of $\theta^*|f_{t-1}$, and q_t be the distribution of $A^*|f_{t-1}$.

It is worthy to mention an insight here. Recall that in Thompson sampling, the algorithm computes p_t, draws θ_t from p_t, and sets $A_t = a^*(\theta_t)$ (all of these are conditioned on f_{t-1}). Therefore, $A_t|f_{t-1}$ has the same distribution as q_t. It follows that $A_t|f_{t-1}$ and $A^*|f_{t-1}$ are identically and independently distributed.

We now proceed to the lemmas.

Lemma 3.2. *Assume $L_{a,\theta} \in [0,1]$ almost surely for all $a \in \mathcal{A}, \theta \in \Theta$. Suppose A^* follows the distribution q, and $A \sim q$ is independent from A^*. Let $L \sim D(A, \theta^*)$ where $\theta^* \sim p$. Define the matrix $M \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{A}|}$ as*

$$M_{a,a'} = \sqrt{q(a)q(a')} \cdot \Big[\mathbb{E}[L_{a,\theta^*} | A^* = a'] - \mathbb{E}[L_{a,\theta^*}] \Big].$$

Then

$$I(A^*; (A, L)) \geq \|M\|_F^2 .$$

Note: In class we had defined M with absolute value, as

$$M_{a,a'} = \sqrt{q(a)q(a')} \cdot \Big| \mathbb{E}[L_{a,\theta^*} | A^* = a'] - \mathbb{E}[L_{a,\theta^*}] \Big|.$$

This change in the lecture notes is necessary for the proof of Lemma 3.3. Also, in the original paper M is defined without absolute value [1].

Proof.

$$\begin{aligned}
I(A^*; (A, L)) &= I(A^*; L|A) + I(A^*; A) \text{ (chain rule)} \\
&= I(A^*; L|A) \text{ (A^* and A are independent)} \\
&= \sum_{a \in \mathcal{A}} q(a) I(A^*; L|A = a) \text{ (definition)} \\
&= \sum_{a \in \mathcal{A}} q(a) I(A^*; L_{a,\theta^*}) \text{ (A^* and A are independent; } L_{A,\theta^*} \sim D(A, \theta^*)) \\
&= \sum_{a \in \mathcal{A}} q(a) \sum_{a' \in \mathcal{A}} q(a') KL\Big(P_{L_{a,\theta^*}|A^*=a'} || P_{L_{a,\theta^*}} \Big) \text{ (Fact 4)} \\
&\geq \sum_{a \in \mathcal{A}} q(a) \sum_{a' \in \mathcal{A}} q(a') TV\Big(P_{L_{a,\theta^*}|A^*=a'}, P_{L_{a,\theta^*}} \Big)^2 \text{ (Pinsker inequality).}
\end{aligned}$$

Recall that $TV(p,q) = \inf_{0\le f\le 1} |\mathbb{E}_p f - \mathbb{E}_q f|$. Taking $f = \mathbb{E}[L_{a,\theta^*}] \in [0,1]$ in

$$TV\Big(P_{L_{a,\theta^*}|A^*=a'}, P_{L_{a,\theta^*}}\Big)$$

gives us a lower bound on the TV. It follows that

$$\begin{aligned} I(A^*;(A,L)) &\ge \sum_{a\in\mathcal{A}} q(a) \sum_{a'\in\mathcal{A}} q(a') \Big[\mathbb{E}[L_{a,\theta^*}|A^*=a] - \mathbb{E}[L_{a,\theta^*}]\Big]^2 \\ &= \|M\|_F^2 \end{aligned}$$

□

Remark 3.2.1. *The setting in this lemma applies to each iteration of Thompson sampling, with the assumption that the loss is bounded between 0 and 1.*

Lemma 3.3. *In the same setting as Lemma 3.2,*

$$\mathbb{E}[L_{A,\theta^*} - L_{A^*,\theta^*}] = -trace(M).$$

Proof. (Omitted in class).

Recall that $M_{a,a'} = \sqrt{q(a)q(a')} \cdot \Big[\mathbb{E}[L_{a,\theta^*}|A^*=a'] - \mathbb{E}[L_{a,\theta^*}]\Big]$.

$$\begin{aligned} trace(M) &= \sum_{a\in\mathcal{A}} M_{a,a} \\ &= \sum_{a\in\mathcal{A}} \sqrt{q(a)q(a)} \cdot \Big[\mathbb{E}[L_{a,\theta^*}|A^*=a] - \mathbb{E}[L_{a,\theta^*}]\Big] \\ &= \sum_{a\in\mathcal{A}} q(a) \cdot \Big[\mathbb{E}[L_{a,\theta^*}|A^*=a] - \mathbb{E}[L_{a,\theta^*}]\Big] \\ &= \mathop{\mathbb{E}}_{a\sim q}\big[\mathbb{E}[L_{a,\theta^*}|A^*=a]\big] - \mathop{\mathbb{E}}_{a\sim q}\big[\mathbb{E}[L_{a,\theta^*}]\big] \\ &= \mathbb{E}[L_{A^*,\theta^*} - L_{A,\theta^*}] \end{aligned}$$

□

Corollary 3.3.1. *In the same setting as Lemma 3.2, the information ratio is upper bounded by $rank(M)$.*

Proof. We have the information ratio

$$\frac{[\mathbb{E}[L_{A,\theta^*} - L_{A^*,\theta^*}]]^2}{I(A^*;(A,L))} \le \frac{trace(M)^2}{\|M\|_F^2} \le rank(M)$$

where the first inequality uses Lemmas 3.2 and 3.3. □

We now apply the results to bound the regret of specific problems.

Corollary 3.3.2. *Assume that the loss is bounded between 0 and 1. For the multi-armed bandit problem, we have the regret bound*

$$\text{regret} \le \sqrt{T \cdot |\mathcal{A}| \cdot \log|\mathcal{A}|}.$$

Proof. The information ratio is upper bounded by $rank(M)$, which is bounded by $|\mathcal{A}|$. By Fact 1, $H(A^*) \leq \log |\mathcal{A}|$. Applying these to Theorem 3.1 gives the desired bound. □

Corollary 3.3.3. *Assume that the loss is bounded between 0 and 1. For the linear bandit problem with $\theta \in \mathbb{R}^d$, $a \in \mathbb{R}^d$ such that $\|\theta\|_2 \leq 1$, $\|a\|_2 \leq 1$, we have the regret bound*

$$\text{regret} \leq d\sqrt{T \log T}.$$

Proof. We have $rank(M) \leq d$. Consider the ϵ-cover of the L_2 unit ball. Setting $\epsilon = 1/O(T)$, we have

$$\begin{aligned} |\mathcal{A}| &\leq O(T)^d \\ \Rightarrow H(A^*) &\leq d \log T. \end{aligned}$$

Applying these to Theorem 3.1 gives the desired bound. □

4 Course Summary and Outlook

This concludes the CS229T/STATS231 course content. Let us quickly and briefly recap the entire course.

- We started with the **asymptotic properties of the Maximum Likelihood Estimator**, and saw convergence at the rate of $O(1/n)$ in the well specified case.
- Then we moved on to **uniform convergence**. There we covered the non-asymptotic cases (finite sample), and relaxed the well-specified assumption. The penalty paid for this is that the convergence is only at the rate of $O(1/\sqrt{n})$. We covered how to get uniform convergence results using **Hoeffding inequality** in
  - the **finite hypothesis** space (via union bound),
  - as well as in the **infinite hypothesis** space with
    - discretization/ϵ-net covers,
    - Rademacher complexity,
    - and VC-dimensions.
- We also covered **Margin theory**, with a focus on its application to 2-layer feed-forward neural networks.
- Then we moved on to the statistical theory of **GAN**s, with a focus on Wasserstein GANs. An important component was reasoning about what was the right theorem to prove. We again used Hoeffding's inequality.
- The next section was **online learning**. Here we moved beyond the i.i.d. assumption (non-statistical), and saw the connection to **online convex optimization**.
- The last topic was on **bandit problems**. Here we were back to statistical flavor. Finally we covered some **Information Theory** concepts in the context of analyzing **Thompson Sampling** in the Bayesian setting.

We did *not* cover the following topics, given time constraints:

- Kernel methods - these are approaches for non-linear models. With kernels we learn linear models in a high dimensional feature space (implicitly defined by kernels), but are non-linear in the original feature space.
- Spectral methods - these are very theoretically elegant methods for unsupervised learning (particularly in clustering). Both spectral methods and kernel methods, though theoretically elegant, have recently fallen out of favor since other methods have demonstrated much stronger empirical performance.
- Theories of Deep Learning - this field is also not mature enough to dedicate a full section on it. The core questions in the field are broadly around:
  - Why do deep learning models even generalize (given that they have such a large number of parameters)?
  - Questions around optimization of their highly non-convex loss surfaces.

Looking forward, there are several aspects of statistical learning that justify further attention and research, such as:

- Saftety.
- Fairness.
- Interpretability and Explainability.
- Quantification of Uncertainty.

These are not only interesting, but also important aspects to study and improve, especially as machine learning methods make their way into our everyday lives in the form of self driving cars, automation, applications in law, healthcare etc.

With this, we conclude the course, and thank you for participating!

References

[1] Daniel Russo and Benjamin Van Roy. "An Information-theoretic Analysis of Thompson Sampling". In: *J. Mach. Learn. Res.* 17.1 (Jan. 2016), pp. 2442–2471. ISSN: 1532-4435. URL: http://www.jmlr.org/papers/volume17/14-087/14-087.pdf.