Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits

Yasin Abbasi-Yadkori **Dávid Pál** Csaba Szepesvári

AISTATS 2012, La Palma, Canary Islands
April 22, 2012

Outline

- Confidence sets for linear model:

$$Y = \theta_*^T X + \text{noise}$$

- Linear bandit problem with side information
- Sparse θ_*

Linear Model

$$Y_t = \theta_*^T X_t + \eta_t \qquad\qquad t = 1, 2, \ldots$$

- η_t is zero-mean, R-sub-Gaussian
- We observe $(X_1, Y_1), (X_2, Y_2), \ldots$
- $X_t \in \mathbb{R}^d$ and can depend on past observations

Goal: Estimate θ_* and construct a confidence set for it.

Confidence Set

Given $\delta \in (0, 1)$, construct

$$C_n := C_n(X_1, Y_1, \ldots, X_n, Y_n, \delta) \subseteq \mathbb{R}^d$$

such that

$$\Pr[\theta_* \in C_n] \geq \delta$$

Previous Construction: Least Squares

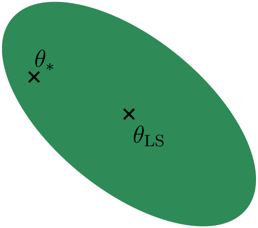
- Least squares solution

$$\mathbf{X} = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad \theta_{\mathsf{LS}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{Y}$$

- Confidence set is an ellipsoid centered at θ_{LS}

$$C_n = \left\{ \theta \in \mathbb{R}^d \ : \ (\theta - \theta_{\mathsf{LS}})^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})(\theta - \theta_{\mathsf{LS}}) \le \text{“radius”} \right\}$$

- “Radius” depends on $n, d, \delta, \mathbf{X}, \lambda, R$ etc.



Previous Construction: Theorem

[Dani et al., 2008], [Rusmevichientong and Tsitsiklis, 2010]

Theorem ([Abbasi-Yadkori et al., 2011])

Assume $\|\theta_*\|_2 \leq S$ *and* $\|X_t\|_2 \leq L$. *With probability* $\geq 1 - \delta$, θ_* *lies in the set*

$$C_n = \left\{ \theta \in \mathbb{R}^d \ : \ \sqrt{(\theta - \theta_{LS})^T (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})(\theta - \theta_{LS})} \leq R \sqrt{2d \log \left(\frac{1 + nL^2/\lambda}{\delta} \right)} + S\sqrt{\lambda} \right\}$$

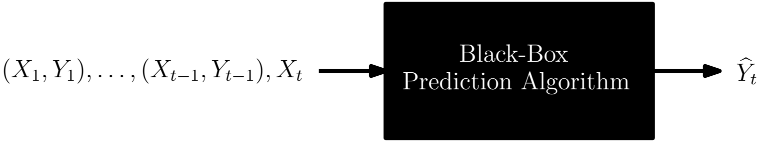
Note: More refined version exists.

Why a different confidence set?

- There are algorithms that are good at **estimating** sparse θ_*
- Can "radius" of the ellipsoid be smaller if θ_* is sparse? (Yes!)

Our construction: Reduction

Assume that we have a black-box prediction algorithm



with regret at most B_n

$$\text{Regret} = \sum_{t=1}^{n} (\widehat{Y}_t - Y_t)^2 - \sum_{t=1}^{n} (\widehat{Y}_t - \theta_*^T X_t)^2 \le B_n$$

Such black-boxes do exist!

Our construction, continued

- Collect black-box predictions $\widehat{Y}_1, \ldots, \widehat{Y}_n$
- Confidence set

$$C_n = \left\{ \theta \in \mathbb{R}^d \ : \ \sum_{t=1}^{n} (\widehat{Y}_t - \theta^T X_t)^2 \le \mathrm{poly}(B_n, R, \log(1/\delta)) \right\}$$

- Note 1: It's an ellipsoid centered at unregularized least squares solution

$$\theta'_{\mathrm{LS}} = (\mathbf{X}^T \mathbf{X})^\dagger \mathbf{X}^T \widehat{\mathbf{Y}}$$

  where we **replaced $\mathbf{Y}$** by $\widehat{\mathbf{Y}}$!
- Note 2: The smaller B_n, the tighter the confidence set.

Aside: Low-regret Prediction Algorithms

Assume $\|X_t\|_2 \leq 1$ and $|Y_t| \leq 1$

Theorem ([Vovk, 2001] & [Azoury and Warmuth, 2001])

If $\|\theta_*\|_2 \leq 1$, online regularized least squares has regret $O(d \log n)$

Theorem ([Gerchinovitz, 2011])

If $\|\theta_*\|_\infty \leq 1$ and $\|\theta\|_0 \leq p$, SeqSEW has regret $O(p \log(nd))$

Note: Confidence set via Vovk-Azoury is roughly the same as best known confidence set for least squares.

Application: Linear Bandits

- Online game. In round t
  1. receive set of actions $D_t \subseteq \mathbb{R}^d$
  2. choose an action $X_t \in D_t$
  3. receive reward $Y_t = \theta_*^T X_t + \eta_t$
- Minimize regret

$$\rho = \sum_{t=1}^{n} \left(\max_{X_t^* \in D_t} \theta_*^T X_t^* \right) - \sum_{t=1}^{n} \theta_*^T X_t$$

- Note: Classical d-armed bandit problem is $D_t = \{e_1, \ldots, e_d\}$

Optimistic Algorithm

- Maintain confidence set C_t
- In round t choose

$$(\widehat{\theta}_t, X_t) = \operatorname*{argmax}_{(\theta, X) \in C_{t-1} \times D_t} \theta^T X$$

- Note: This reduces to UCB for $D_t = \{e_1, \ldots, e_d\}$

Regret of Optimistic Algorithm

Theorem
If $|\theta_*^T X| \leq 1$ for all $X \in D_t$ and t, then with probability $\geq 1 - \delta$, for all n, regret is

$$O\left(\sqrt{dnB_n} \cdot \mathrm{polylog}(n, d, 1/\delta, B_n)\right)$$

For $\|\theta\|_0 \leq p$ using $\textsc{SeqSEW}$ we get

$$O\left(\sqrt{pdn} \cdot \mathrm{polylog}(n, d, 1/\delta)\right)$$

Improvement over $O(d\sqrt{n} \cdot \mathrm{polylog}(n, d, 1/\delta))$ in [Dani et al., 2008]

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011).
Improved algorithms for linear stochastic bandits.
In Shawe-Taylor, J., Zemel, R. S., Bartlett, P., Pereira, F., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 24, (NIPS 2011).*

Azoury, K. S. and Warmuth, M. K. (2001).
Relative loss bounds for on-line density estimation with the exponential family of distributions.
Machine Learning, 43:211–246.

Dani, V., Hayes, T. P., and Kakade, S. M. (2008).
Stochastic linear optimization under bandit feedback.
In Servedio, R. and Zhang, T., editors, *Proceedings of the 21st Annual Conference on Learning Theory (COLT 2008)*, pages 355–366.

Gerchinovitz, S. (2011).
Sparsity regret bounds for individual sequences in online linear regression.
In *Proceedings of the 24st Annual Conference on Learning Theory (COLT 2011).*

Rusmevichientong, P. and Tsitsiklis, J. N. (2010).
Linearly parameterized bandits.
Mathematics of Operations Research, 35(2):395–411.

Vovk, V. (2001).
Competitive on-line statistics.
International Statistical Review, 69:213–248.