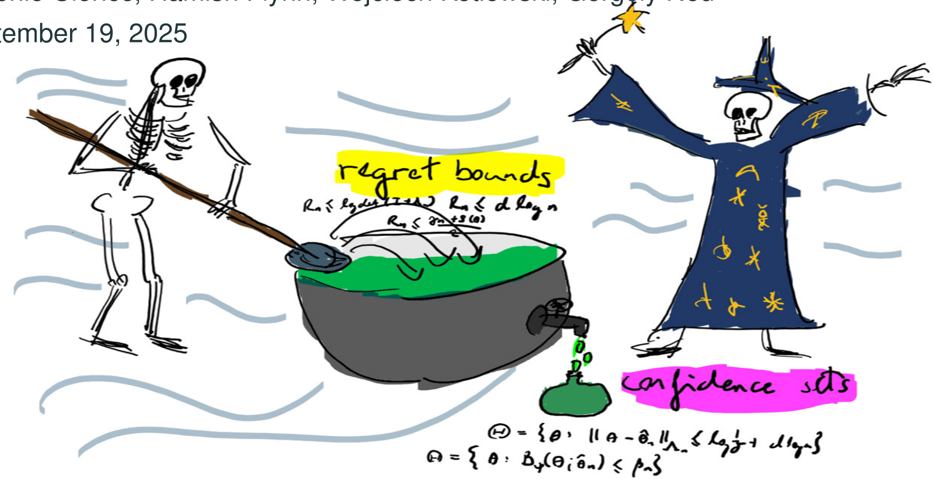
Confidence sequences for generalised linear models via regret analysis

Eugenio Clerico, Hamish Flynn, Wojciech Kotłowski, Gergely Neu

September 19, 2025

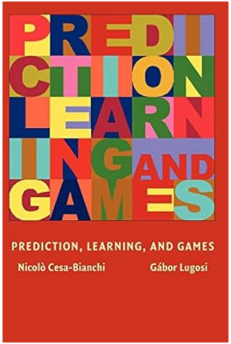


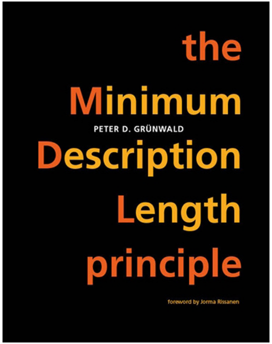
Why are confidence bounds/sets/sequences interesting (for RL)?

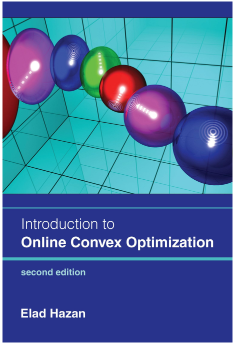
- Exploration-exploitation trade-offs (OFU, Thompson Sampling, etc.)
- Stopping rules for pure exploration
- Safe exploration
- Asymptotic optimality/instance-optimality

This work

Assume we know a bit about online learning.

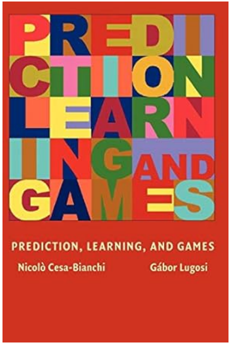


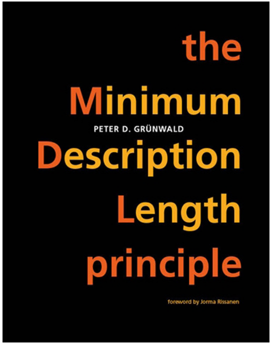


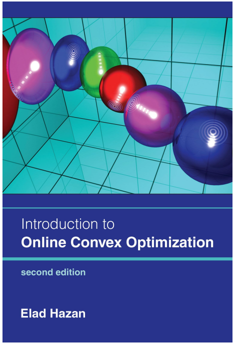


This work

Assume we know a bit about online learning.







We want to construct confidence sequences for GLMs without doing any actual work.

Setting

Linear model.

- Covariates $X_1, \ldots, X_n \in \mathbb{R}^d$
- Responses $Y_1, \ldots, Y_n \in \mathbb{R}$
- Likelihood $p(Y_t|X_t, \theta^\star) = \frac{1}{\sqrt{2\pi}} \exp\big(- \frac{1}{2}(Y_t - \langle \theta^\star, X_t \rangle)^2 \big)$

Setting

Linear model.

- Covariates $X_1, \ldots, X_n \in \mathbb{R}^d$
- Responses $Y_1, \ldots, Y_n \in \mathbb{R}$
- Likelihood $p(Y_t | X_t, \theta^\star) = \frac{1}{\sqrt{2\pi}} \exp\big(- \frac{1}{2}(Y_t - \langle \theta^\star, X_t \rangle)^2 \big)$

Log-likelihood loss. Define $\ell_t(\theta) = -\log(p(Y_t | X_t, \theta)) = \frac{1}{2}(Y_t - \langle \theta^\star, X_t \rangle)^2 + \frac{1}{2}\log(2\pi)$.

Setting

Linear model.

- Covariates $X_1, \ldots, X_n \in \mathbb{R}^d$
- Responses $Y_1, \ldots, Y_n \in \mathbb{R}$
- Likelihood $p(Y_t|X_t, \theta^\star) = \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{1}{2}(Y_t - \langle\theta^\star, X_t\rangle)^2\big)$

Log-likelihood loss. Define $\ell_t(\theta) = -\log(p(Y_t|X_t, \theta)) = \frac{1}{2}(Y_t - \langle\theta^\star, X_t\rangle)^2 + \cancel{\frac{1}{2}\log(2\pi)}$.

Setting

Linear model.

- Covariates $X_1, \dots, X_n \in \mathbb{R}^d$
- Responses $Y_1, \dots, Y_n \in \mathbb{R}$
- Likelihood $p(Y_t | X_t, \theta^\star) = \frac{1}{\sqrt{2\pi}} \exp\big(- \frac{1}{2}(Y_t - \langle \theta^\star, X_t \rangle)^2 \big)$

Log-likelihood loss. Define $\ell_t(\theta) = -\log(p(Y_t | X_t, \theta)) = \frac{1}{2}(Y_t - \langle \theta^\star, X_t \rangle)^2 + \cancel{\frac{1}{2}\log(2\pi)}$.

Adaptive design. X_t depends on $X_1, Y_1, \dots, X_{t-1}, Y_{t-1}$.

Objective and claim

For $\delta \in (0, 1]$, a δ-*confidence sequence* for $\theta^\star$ is a sequence of sets $\Theta_1, \Theta_2, \ldots$, such that

$$\mathbb{P}(\forall n \geq 1 : \theta^\star \in \Theta_n) \geq 1 - \delta\,.$$

Objective and claim

For $\delta \in (0,1]$, a δ-*confidence sequence* for $\theta^\star$ is a sequence of sets $\Theta_1, \Theta_2, \dots$, such that

$$\mathbb{P}(\forall n \geq 1 : \theta^\star \in \Theta_n) \geq 1 - \delta\,.$$

Gold standard (e.g. OFUL). Θ_n is the ellipsoid

$$\Theta_n := \left\{ \theta \in \mathbb{R}^d : \|\theta - \widehat{\theta}_n\|^2_{\Lambda_n + \frac{1}{\gamma^2}\mathrm{Id}} \leq \beta_n \right\}\,,$$

where $\widehat{\theta}_n := \operatorname{argmin}_{\theta \in \mathbb{R}^d} \{\sum_{t=1}^n \ell_t(\theta) + \frac{1}{2\gamma^2}\|\theta\|_2^2\}$, $\Lambda_n = \sum_{t=1}^n X_t X_t^\top$, $\beta_n = \mathcal{O}(d \log n)$.

Objective and claim

For $\delta \in (0,1]$, a δ-*confidence sequence* for $\theta^\star$ is a sequence of sets $\Theta_1, \Theta_2, \dots$, such that

$$\mathbb{P}(\forall n \geq 1 : \theta^\star \in \Theta_n) \geq 1 - \delta\,.$$

Gold standard (e.g. OFUL). Θ_n is the ellipsoid

$$\Theta_n := \left\{ \theta \in \mathbb{R}^d : \|\theta - \widehat{\theta}_n\|^2_{\Lambda_n + \frac{1}{\gamma^2}\mathrm{Id}} \leq \beta_n \right\}\,,$$

where $\widehat{\theta}_n := \mathrm{argmin}_{\theta \in \mathbb{R}^d}\{\sum_{t=1}^n \ell_t(\theta) + \frac{1}{2\gamma^2}\|\theta\|_2^2\},\ \Lambda_n = \sum_{t=1}^n X_t X_t^\top,\ \beta_n = \mathcal{O}(d \log n)$.

Online-to-confidence-set conversion. Use the regret bound of an online learning algorithm to determine β_n.

Objective and claim

For $\delta \in (0, 1]$, a δ-*confidence sequence* for $\theta^\star$ is a sequence of sets $\Theta_1, \Theta_2, \ldots,$ such that

$$\mathbb{P}(\forall n \geq 1 : \theta^\star \in \Theta_n) \geq 1 - \delta\,.$$

Gold standard (e.g. OFUL). Θ_n is the ellipsoid

$$\Theta_n := \left\{ \theta \in \mathbb{R}^d : \|\theta - \widehat{\theta}_n\|^2_{\Lambda_n + \frac{1}{\gamma^2}\mathrm{Id}} \leq \beta_n \right\}\,,$$

where $\widehat{\theta}_n := \mathrm{argmin}_{\theta \in \mathbb{R}^d} \{\sum_{t=1}^n \ell_t(\theta) + \frac{1}{2\gamma^2}\|\theta\|_2^2\},\ \Lambda_n = \sum_{t=1}^n X_t X_t^\top,\ \beta_n = \mathcal{O}(d \log n)$.

Online-to-confidence-set conversion. Use the regret bound of an online learning algorithm to determine β_n.

Claim. We can recover or improve upon all confidence sequences for GLMs via OTCS.

Online-to-confidence-set conversion attempt 1

(don’t do this)

Online linear regression

Protocol. For $t = 1, 2, \ldots, n$:

1. Environment reveals X_t to the learner
2. Learner picks $\theta_t \in \Theta$
3. Environment reveals Y_t to the learner,
4. Learner incurs the loss $\ell_t(\theta_t)$

Online linear regression

Protocol. For $t = 1, 2, \dots, n$:

1. Environment reveals X_t to the learner
2. Learner picks $\theta_t \in \Theta$
3. Environment reveals Y_t to the learner,
4. Learner incurs the loss $\ell_t(\theta_t)$

Regret. The regret of $\theta^n := (\theta_1, \dots, \theta_n)$ w.r.t. a comparator $\bar{\theta} \in \Theta$ is

$$\mathrm{regret}_{\theta^n}(\bar{\theta}) = \sum_{t=1}^{n} \left(\ell_t(\theta_t) - \ell_t(\bar{\theta})\right) .$$

Online linear regression

Protocol. For $t = 1, 2, \ldots, n$:

1. Environment reveals X_t to the learner
2. Learner picks $\theta_t \in \Theta$
3. Environment reveals Y_t to the learner,
4. Learner incurs the loss $\ell_t(\theta_t)$

Regret. The regret of $\theta^n := (\theta_1, \ldots, \theta_n)$ w.r.t. a comparator $\bar{\theta} \in \Theta$ is

$$\mathrm{regret}_{\theta^n}(\bar{\theta}) = \sum_{t=1}^{n} \left(\ell_t(\theta_t) - \ell_t(\bar{\theta})\right).$$

If the Vovk-Azoury-Warmuth forecaster (with parameter γ) is used to generate θ^n, then

$$\mathrm{regret}_{\theta^n}(\bar{\theta}) \leq \frac{1}{2\gamma^2}\|\bar{\theta}\|_2^2 + \frac{\max_{t\in[n]} Y_t^2}{2} \log\det\left(\gamma^2 \Lambda_n + \mathrm{Id}\right) = \mathcal{O}(d(\log n)^2).$$

Online-to-confidence-set conversion (attempt 1)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \ldots$ and any strategy θ^n, the sets $\Theta_1, \Theta_2, \ldots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n)\big) \leq \mathrm{regret}_{\theta^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\} .$$

Online-to-confidence-set conversion (attempt 1)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \ldots$ and any strategy θ^n, the sets $\Theta_1, \Theta_2, \ldots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n)\big) \leq \operatorname{regret}_{\theta^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\} .$$

Proof. First,

$$\sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\bar{\theta}_n)\big) = \sum_{t=1}^{n} \big(\ell_t(\theta_t) - \ell_t(\bar{\theta}_n)\big) + \sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\theta_t)\big) .$$

Online-to-confidence-set conversion (attempt 1)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \dots$ and any strategy θ^n, the sets $\Theta_1, \Theta_2, \dots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n) \big) \leq \mathrm{regret}_{\theta^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\}.$$

Proof. First,

$$\sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\bar{\theta}_n) \big) = \underbrace{\sum_{t=1}^{n} \big(\ell_t(\theta_t) - \ell_t(\bar{\theta}_n) \big)}_{\mathrm{regret}_{\theta^n}(\bar{\theta}_n)} + \sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\theta_t) \big).$$

Online-to-confidence-set conversion (attempt 1)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \ldots$ and any strategy θ^n, the sets $\Theta_1, \Theta_2, \ldots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n) \big) \leq \mathrm{regret}_{\theta^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\} .$$

Proof. First,

$$\sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\bar{\theta}_n) \big) = \underbrace{\sum_{t=1}^{n} \big(\ell_t(\theta_t) - \ell_t(\bar{\theta}_n) \big)}_{\mathrm{regret}_{\theta^n}(\bar{\theta}_n)} + \underbrace{\sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\theta_t) \big)}_{\leq \ \log \frac{1}{\delta} \ \ \text{w.p.} \geq 1-\delta} .$$

Online-to-confidence-set conversion (attempt 1)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \ldots$ and any strategy θ^n, the sets $\Theta_1, \Theta_2, \ldots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n)\big) \leq \mathrm{regret}_{\theta^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\} .$$

Proof. First,

$$\sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\bar{\theta}_n)\big) = \underbrace{\sum_{t=1}^{n} \big(\ell_t(\theta_t) - \ell_t(\bar{\theta}_n)\big)}_{\mathrm{regret}_{\theta^n}(\bar{\theta}_n)} + \underbrace{\sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\theta_t)\big)}_{\leq \ \log \frac{1}{\delta} \ \text{ w.p.} \geq 1-\delta} .$$

Therefore,

$$\mathbb{P}\left(\forall n \geq 1, \sum_{t=1}^{n} \big(\ell_t(\theta^\star) - \ell_t(\bar{\theta}_n)\big) \leq \mathrm{regret}_{\theta^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right) \geq 1 - \delta \, .$$

Online-to-confidence-set conversion (attempt 1)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \dots$ and any strategy θ^n, the sets $\Theta_1, \Theta_2, \dots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \left(\ell_t(\theta) - \ell_t(\bar{\theta}_n) \right) \leq \mathrm{regret}_{\theta^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\}.$$

If it is known that $\|\theta^\star\|_2 \leq B$, then by plugging in the VAW regret bound and then completing some squares, we obtain

$$\Theta_n := \left\{ \theta \in \mathbb{R}^d : \|\theta - \widehat{\theta}_n\|^2_{\Lambda_n + \frac{1}{\gamma^2}\mathrm{Id}} \leq \max_{t \in [n]} Y_t^2 \log\det(\gamma^2 \Lambda_n + \mathrm{Id}) + \frac{B^2}{\gamma^2} + 2 \log \tfrac{1}{\delta} \right\}.$$

Online-to-confidence-set conversion (attempt 1)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \ldots$ and any strategy θ^n, the sets $\Theta_1, \Theta_2, \ldots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \left(\ell_t(\theta) - \ell_t(\bar{\theta}_n) \right) \leq \mathrm{regret}_{\theta^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\}.$$

If it is known that $\|\theta^\star\|_2 \leq B$, then by plugging in the VAW regret bound and then completing some squares, we obtain

$$\Theta_n := \left\{ \theta \in \mathbb{R}^d : \|\theta - \widehat{\theta}_n\|^2_{\Lambda_n + \frac{1}{\gamma^2}\mathrm{Id}} \leq \max_{t \in [n]} Y_t^2 \log\det(\gamma^2 \Lambda_n + \mathrm{Id}) + \frac{B^2}{\gamma^2} + 2 \log \tfrac{1}{\delta} \right\}.$$

Problem. For this confidence set, $\beta_n = \mathcal{O}(d(\log n)^2)$, whereas we should have $\beta_n = \mathcal{O}(d \log n)$.

Can we remove the factor of $\max_{t \in [n]} Y_t^2$?

Yes

Online-to-confidence-set conversion attempt 2

(do this)

Sequential probability assignment

Protocol. For $t = 1, 2, \ldots, n$:

1. Environment reveals X_t to the learner
2. Learner picks $Q_t \in \Delta_\Theta$ with density q_t
3. Environment reveals Y_t to the learner,
4. Learner incurs the log loss $\mathcal{L}_t(q_t) = -\log \int \exp(-\ell_t(\theta)) q_t(\theta)\, \mathrm{d}\theta$

Sequential probability assignment

Protocol. For $t = 1, 2, \ldots, n$:

1. Environment reveals X_t to the learner
2. Learner picks $Q_t \in \Delta_\Theta$ with density q_t
3. Environment reveals Y_t to the learner,
4. Learner incurs the log loss $\mathcal{L}_t(q_t) = -\log \int \exp(-\ell_t(\theta)) q_t(\theta)\, \mathrm{d}\theta$

Regret. The regret of $q^n = (q_1, \ldots, q_n)$ w.r.t. a comparator $\bar{\theta} \in \Theta$ is

$$\mathrm{regret}_{q^n}(\bar{\theta}) = \sum_{t=1}^{n} \left(\mathcal{L}_t(q_t) - \ell_t(\bar{\theta}) \right) .$$

Sequential probability assignment

Protocol. For $t = 1, 2, \ldots, n$:

1. Environment reveals X_t to the learner
2. Learner picks $Q_t \in \Delta_\Theta$ with density q_t
3. Environment reveals Y_t to the learner,
4. Learner incurs the log loss $\mathcal{L}_t(q_t) = -\log \int \exp(-\ell_t(\theta)) q_t(\theta) \, \mathrm{d}\theta$

Regret. The regret of $q^n = (q_1, \ldots, q_n)$ w.r.t. a comparator $\bar{\theta} \in \Theta$ is

$$\mathrm{regret}_{q^n}(\bar{\theta}) = \sum_{t=1}^{n} \left(\mathcal{L}_t(q_t) - \ell_t(\bar{\theta}) \right) .$$

If Vovk's Aggregating Algorithm (a.k.a. the Exponentially Weighted Average forecaster) is used with the prior $Q_1 = \mathcal{N}(0, \gamma^2 \mathrm{Id})$, then

$$\mathrm{regret}_{q^n}(\bar{\theta}) \leq \frac{1}{2\gamma^2} \|\bar{\theta}\|_2^2 + \frac{1}{2} \log \det \left(\gamma^2 \Lambda_n + \mathrm{Id} \right) = \mathcal{O}(d \log n)$$

Online-to-confidence-set conversion (attempt 2)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \ldots$ and any strategy q^n, the sets $\Theta_1, \Theta_2, \ldots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n)\big) \leq \mathrm{regret}_{q^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\}.$$

Online-to-confidence-set conversion (attempt 2)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \dots$ and any strategy q^n, the sets $\Theta_1, \Theta_2, \dots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n)\big) \leq \operatorname{regret}_{q^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\}.$$

Proof. Exactly the same as before. $\sum_{t=1}^{n}(\ell_t(\theta^\star) - \mathcal{L}_t(q_t))$ is still the logarithm of a non-negative martingale.

Online-to-confidence-set conversion (attempt 2)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \ldots$ and any strategy q^n, the sets $\Theta_1, \Theta_2, \ldots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n)\big) \leq \mathrm{regret}_{q^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\} .$$

This time, if $\|\theta^\star\|_2 \leq B$, then by plugging in the VAA/EWA regret bound and completing some squares, we obtain

$$\Theta_n := \left\{ \theta \in \mathbb{R}^d : \|\theta - \widehat{\theta}_n\|^2_{\Lambda_n + \frac{1}{\gamma^2}\mathrm{Id}} \leq \log\det(\gamma^2 \Lambda_n + \mathrm{Id}) + \frac{B^2}{\gamma^2} + 2\log \tfrac{1}{\delta} \right\} .$$

Online-to-confidence-set conversion (attempt 2)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \dots$ and any strategy q^n, the sets $\Theta_1, \Theta_2, \dots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n) \big) \le \mathrm{regret}_{q^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\} .$$

This time, if $\|\theta^\star\|_2 \le B$, then by plugging in the VAA/EWA regret bound and completing some squares, we obtain

$$\Theta_n := \left\{ \theta \in \mathbb{R}^d : \|\theta - \widehat{\theta}_n\|^2_{\Lambda_n + \frac{1}{\gamma^2}\mathrm{Id}} \le \log\det(\gamma^2 \Lambda_n + \mathrm{Id}) + \frac{B^2}{\gamma^2} + 2\log\tfrac{1}{\delta} \right\} .$$

We now have $\beta_n = \mathcal{O}(d \log n)$.

Online-to-confidence-set conversion (attempt 2)

Claim. For any comparators $\bar{\theta}_1, \bar{\theta}_2, \ldots$ and any strategy q^n, the sets $\Theta_1, \Theta_2, \ldots$ form a δ-CS, where

$$\Theta_n = \left\{ \theta \in \mathbb{R}^d : \sum_{t=1}^{n} \big(\ell_t(\theta) - \ell_t(\bar{\theta}_n)\big) \leq \mathrm{regret}_{q^n}(\bar{\theta}_n) + \log \tfrac{1}{\delta} \right\} .$$

This time, if $\|\theta^\star\|_2 \leq B$, then by plugging in the VAA/EWA regret bound and completing some squares, we obtain

$$\Theta_n := \left\{ \theta \in \mathbb{R}^d : \|\theta - \widehat{\theta}_n\|^2_{\Lambda_n + \frac{1}{\gamma^2}\mathrm{Id}} \leq \log\det(\gamma^2 \Lambda_n + \mathrm{Id}) + \frac{B^2}{\gamma^2} + 2\log \tfrac{1}{\delta} \right\} .$$

We now have $\beta_n = \mathcal{O}(d \log n)$.

Conclusion. Use regret bounds for sequential probability assignment.

What about reinforcement learning?

RL stuff

Bandits.

- Obvious applications to UCB algorithms (for generalised linear bandits)
- Other applications to batched bandit algorithms

RL stuff

Bandits.

- Obvious applications to UCB algorithms (for generalised linear bandits)
- Other applications to batched bandit algorithms

Model-based RL.

- Applications to online (generalised) linear control

RL stuff

Bandits.

- Obvious applications to UCB algorithms (for generalised linear bandits)
- Other applications to batched bandit algorithms

Model-based RL.

- Applications to online (generalised) linear control

Model-free RL.

- A bit tricky. Perhaps we need a reduction to a game in which previous predictions influence future responses
- Confidence sets for temporal difference estimators?
- New confidence sets for value functions of linear MDPs? (replace covering numbers by data-dependent regret bounds)

The end. Thank you!