Thompson Sampling for Linear Bandits

In MAB we played arm I_t: $\mathbb{P}(I_t = i) = \mathbb{P}(i = i_* \mid (x_s, y_s)_{s=1}^{t-1})$

Equivalently, maintain a posterior over the parameter $\theta_{*,i}$

say $\cong \mathcal{N}(\hat{\theta}_{t,i}, \frac{1}{T_i})$ and draw $\theta'_{t,i} \sim \mathcal{N}(\hat{\theta}_{t,i}, \frac{1}{T_i}) \ \forall i$

$I_t = \arg\max_i \theta'_{t,i}$.

Input: arm set $\mathcal{X} \subset \mathbb{R}^d$
Init. prior $P_0 = \mathcal{N}(0, I)$
for $t = 1, 2, \ldots$

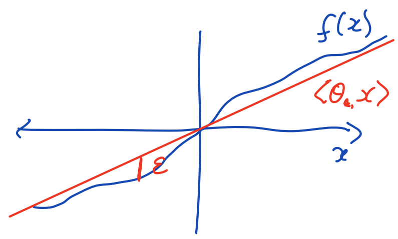
$\mathcal{N}(0,1)$

Draw $\theta'_t \sim P_{t-1}$

Pull arm $x_t = \arg\max_{x \in \mathcal{X}} \langle x, \theta'_t \rangle$, observe $y_t = \langle x_t, \theta_* \rangle + \eta_t$

Update posterior $P_{t-1} \to P_t$. (More or less: $P_t = \mathcal{N}(\hat{\theta}_t, V_t^{-1})$ where

$V_t = \sum_{s=1}^{t} x_s x_s^T + \lambda I \quad \hat{\theta}_t = V_t^{-1} \sum_{s=1}^{t} x_s y_s$)



Suppose $\mathbb{E}[y_t \mid x_t] = f(x_t)$ where f is arbitrary.

but $\exists\, \theta_* \in \mathbb{R}^d$: $\max_x |\langle x, \theta_* \rangle - f(x)| \leq \varepsilon$.

Then the elimination algorithm satisfies

$$\max_x \mathbb{E}\left[\sum_{t=1}^{T} f(x) - f(x_t)\right] \leq d\sqrt{T \log |\mathcal{X}|} + T d \sqrt{\varepsilon}.$$

Suppose we have some algorithm that pulls arms $x_1, x_2 \ldots$

and it has low regret: $\sum_{t=1}^{T} \langle x_* - x_t, \theta_* \rangle \leq: R_T$

Then if I choose $\zeta \sim \text{uniform}(\{x_t\}_{t=1}^{T})$ then

$$\mathbb{E}[\langle x_* - x_\zeta, \theta_* \rangle] \leq \frac{R_T}{T}$$

"online to batch" conversion.

Proof

$$\mathbb{E}[\langle x_* - x_\zeta, \theta_* \rangle] = \sum_t \mathbb{P}(\zeta = t) \langle x_* - x_t, \theta_* \rangle$$
$$= \frac{1}{T} \sum_{t=1}^{T} \langle x_* - x_t, \theta_* \rangle$$
$$= \frac{R_T}{T}.$$

We say $x \in \mathcal{X}$ is ε-good if $\langle x_* - x, \theta_* \rangle \leq \varepsilon$.

$\Longrightarrow$ Elim/UCB can find ε-good arm (in expectation)

after just $T = \frac{d \log|\mathcal{X}| \wedge d^2}{\Delta^2 \vee \varepsilon^2}$ $\quad \Delta = \min_{x \neq x_*} \langle x_* - x, \theta_* \rangle$

min $= \wedge \equiv \cap \equiv XY$

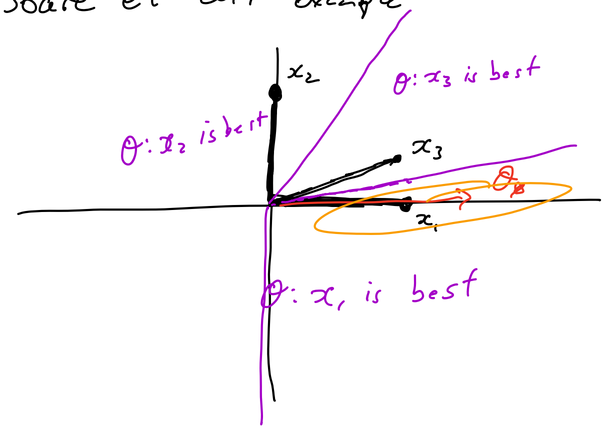
max $= \vee \equiv \cup \equiv X+Y$

If G-optimal design then constructed LS estimator $\hat{\theta}$, then once

$T \geq \frac{d \log|\mathcal{X}| \wedge d^2}{\Delta^2 \vee \varepsilon^2}$, $\quad \operatorname{argmax}_{x \in \mathcal{X}} \langle x, \hat{\theta} \rangle$ is ε-good w.h.p.

Best arm identification for Linear Bandits.

Soare et 2014 example:



$$x_3 = \begin{pmatrix} \cos\gamma \\ \sin\gamma \end{pmatrix}$$