Target Tracking for Contextual Bandits: Application to Demand Side Management

Margaux Brégère

Joint work with Gilles Stoltz (CNRS Université Paris-Sud), Yannig Goude (EDF R&D) and Pierre Gaillard (Inria)

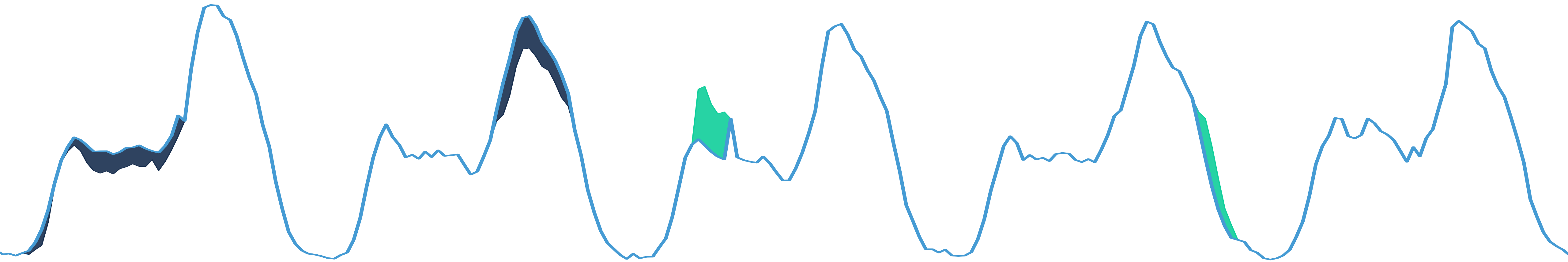
International Conference on Machine Learning

June, 12th 2019

Introduction

Motivation: maintain balance between production and demand

▸ **Apply contextual-bandit theory to demand side management by offering price incentives**



K=3 tariffs: **Low (L)**, **Normal (N)**, **High (H)**

Modeling of the Electricity Consumption (general additive model)

$$Y_t = f_1(\text{temperature}) + f_2(\text{position in the year}) + f_3(\text{tariff}) + f_4(\text{hour}) + \ldots + \text{noise}$$

$$\rightarrow \exists \; \boldsymbol{\phi} \text{ (known transfer function)}, \boldsymbol{\theta} \text{ (unknown parameter)} \mid \mathbb{E}[\mathbf{Y}] = \boldsymbol{\phi}(\mathbf{X})^{\mathbf{T}}\boldsymbol{\theta}$$

At each round $t = 1, \ldots$

- Observe a context $x_t \in \mathcal{X}$
- Pick an allocation of price levels $p_t \in \mathcal{P} \subset \Delta_K = \{(p_1, \ldots, p_K) \in [0,1]^K \,|\, \sum_k p_k = 1\}$
- Observe the consumption $\mathbf{Y}_{\mathbf{t},\mathbf{p_t}} = \boldsymbol{\phi}(\mathbf{x_t}, \mathbf{p_t})^{\mathbf{T}}\boldsymbol{\theta} + \mathbf{p_t^T}\boldsymbol{\varepsilon}_\mathbf{t}$ with $\mathbb{V}[\boldsymbol{\varepsilon}_\mathbf{t}] = \boldsymbol{\Gamma} \in \mathcal{M}_\mathbf{K}(\mathbb{R})$ in $[0, C]$

Target tracking for contextual bandits

At each round $t = 1, \ldots$

- Observe a context $x_t \in \mathcal{X}$ and a **target c_t** in $[0, C]$
- Pick $p_t \in \mathcal{P}$ and observe Y_{t,p_t}
- Suffer a **loss** $\left(\mathbf{Y}_{t,p_t} - \mathbf{c}_t\right)^2$

Aim. Minimize the pseudo-regret $$R_T = \sum_{t=1}^{T} \ell_{t,p_t} - \sum_{t=1}^{T} \min_{p \in \mathcal{P}} \ell_{t,p}$$

$$\text{with } \ell_{t,p} = \mathbb{E}\left[\left(Y_{t,p} - c_t\right)^2 \mid \text{past}\right] = \left(\phi(x_t, p)^T \theta - c_t\right)^2 + p^T \Gamma p$$

- Classical setting $R_T = \sum_{t=1}^{T} \max_{p \in \mathcal{P}} \phi(x_t, p)^T \theta - \sum_{t=1}^{T} \phi(x_t, p_t)^T \theta$
- Reach a bias-variance trade-off by estimating $\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}$

Optimistic algorithm for tracking target with context

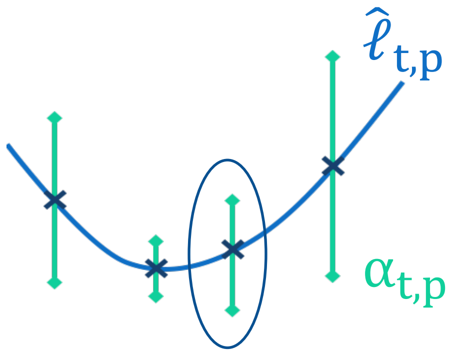
- For $t = 1, \dots, n$,
  - Estimate Γ and get a confidence bound for $p_1, \dots p_n$ well chosen

$$\hat{\Gamma}_n = \arg\min_{\hat{\Gamma}} \sum_{s=1}^{n} \left(\left(Y_{t,p_t} - \left[\phi(x_t, p_t)^T \hat{\theta}_n \right]_C \right)^2 - p_t^T \hat{\Gamma} p_t \right)^2$$

$$\text{with } \hat{\theta}_n = \arg\min_{\hat{\theta}} \sum_{s=1}^{n} \left(Y_{s,p_s} - \phi(x_s, p_s)^T \hat{\theta} \right)^2 + \lambda \left\| \hat{\theta} \right\|^2$$

- For $t \geq n + 1$,
  - Re-estimate θ and get a confidence set – as in Lin-UCB (Li et al. 2010)
  - For each p, estimate loss and get a confidence bound

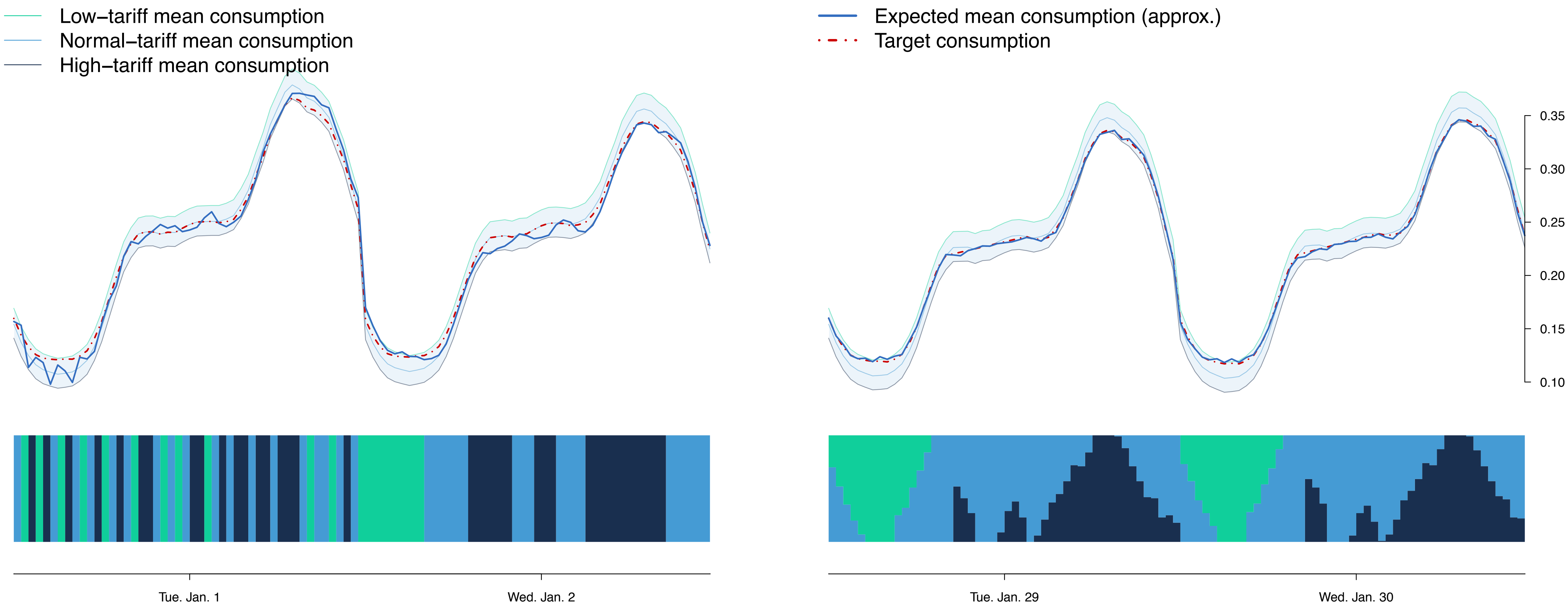
$$\hat{\ell}_{t,p} = \left(\left[\phi(x_t, p)^T \hat{\boldsymbol{\theta}}_{t-1} \right]_C - c_t \right)^2 + p^T \hat{\boldsymbol{\Gamma}}_{\mathbf{n}} p \text{ and } \left\| \hat{\ell}_{t,p} - \ell_{t,p} \right\| \leq \alpha_{t,p}$$



  - Select price levels optimistically $\mathbf{p_t} \in \underset{\mathbf{p} \in \mathcal{P}}{\mathbf{arg\,min}} \left\{ \hat{\ell}_{\mathbf{t,p}} - \boldsymbol{\alpha}_{\mathbf{t,p}} \right\}$

> **Theorem.** For proper choice of confidence level $\alpha_{t,p}$ and $n = \mathcal{O}(T^{2/3})$, with probability at least $1 - \delta$, $R_T \lesssim \mathbf{T}^{2/3} \ln^2 (T/\delta) \sqrt{\ln(1/\delta)}$.

Experiments



Thank you and see you at the poster session! #122