



POLITECNICO
MILANO 1863



e l l i s

THOMPSON SAMPLING-LIKE ALGORITHMS FOR STOCHASTIC RISING RESTED BANDITS

MARCO FIANDRI, ALBERTO MARIA METELLI AND FRANCESCO TROVÒ

marco.fiandri@mail.polimi.it, {albertomaria.metelli, francesco1.trovo}@polimi.it

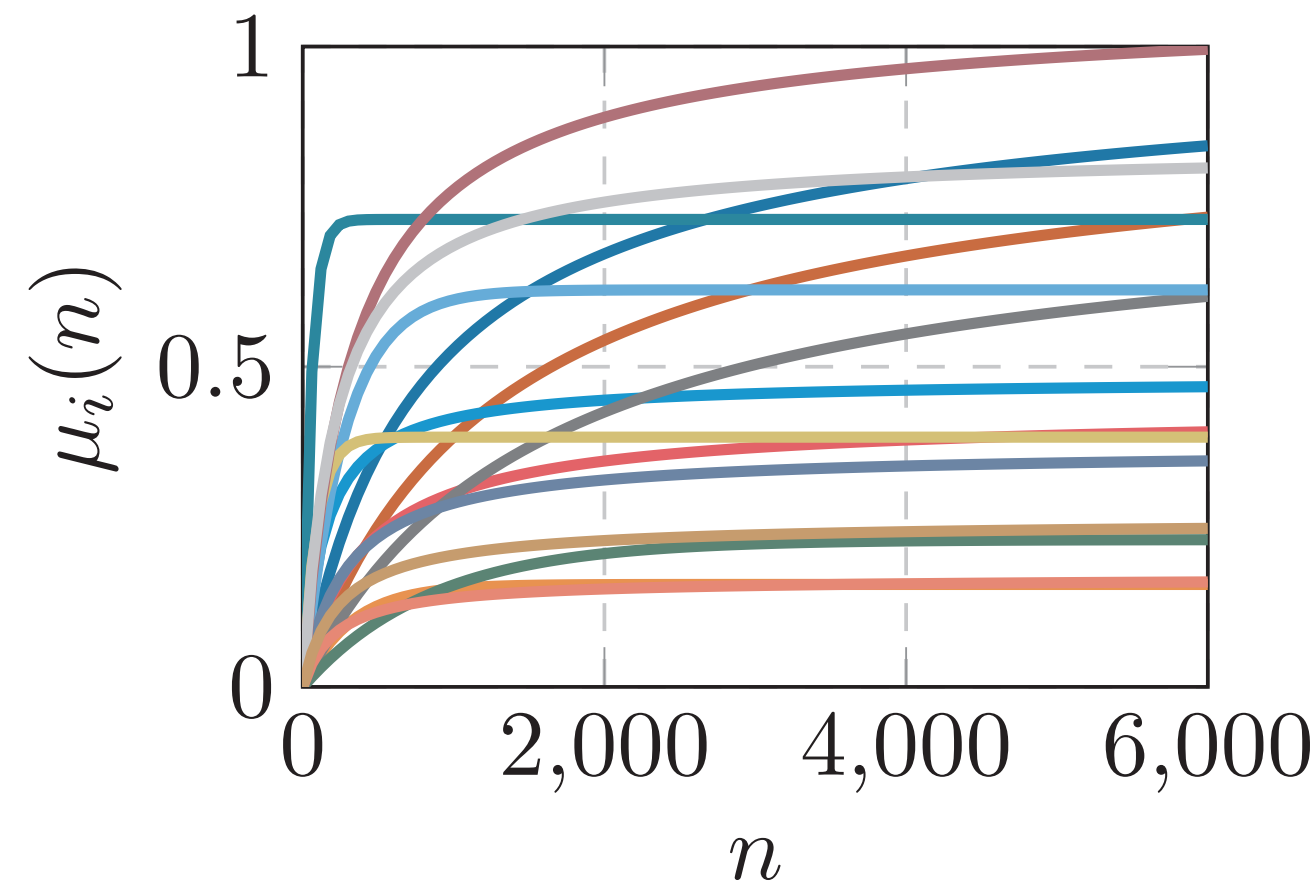


SETTING - OVERVIEW

STOCHASTIC RISING RESTED BANDITS

- Samples of arm i taken from either a Bernoulli or a σ^2 -subgaussian distribution with expectation $\mu_i(n)$, being n the number of pulls of the arm.
- Goal: minimize the *expected cumulative regret*:

$$R(\mathcal{A}, T) := T \bar{\mu}_{i^*(T)}(T) - \mathbb{E} \left[\sum_{t=1}^T \mu_{I_t}(N_{I_t, t}) \right]$$



MOTIVATION

- Real-world applications present scenarios where the rewards associated with each arm evolve.
- We study the domain of *stochastic rising rested bandits* (SRRB, Heidari et al., 2016; Metelli et al., 2022).
- The SRRB models situations where the arms' expected rewards *increase* with the pulls, describing one of the most fundamental settings in Automatic Machine Learning (AutoML).
- Bayesian methods, which were notoriously hard to study, lacked an analysis in this setting, providing excellent experimental results (Metelli et al. (2022)).

CONTRIBUTIONS

- Analysis of Thompson Sampling variants with forced exploration: ET-Beta-TS and its Gaussian extension γ -ET-GTS, with novel analytical techniques.
- A lower bound dependent on a new complexity measure.
- Sliding-window extensions (ET-Beta-SWTS, γ -ET-SWGTS) for limited horizon T .

BOUNDING THE ONLINE REGRET

REGRET FOR ET-BETA-TS:

$$\mathbb{E}[N_{i,T}] \leq O \left(\Gamma + (1+\epsilon) \frac{\log(T)}{d(\bar{\mu}_i(T), \bar{\mu}_1(\sigma))} + \frac{1}{\epsilon^2} + \sum_{j=\Gamma}^{\sigma} \frac{1}{(1-\bar{\mu}_1(\sigma))^{j+1}} \delta_{TV}(\text{Bin}(j, \bar{\mu}_1(j)), \text{Bin}(j, \bar{\mu}_1(\sigma))) \right)$$

For $\Gamma \geq \sigma$:

$$\mathbb{E}[N_{i,T}] \leq O \left(\Gamma + \frac{\log(T)}{\Delta_i^2} \right)$$

THE SAMPLE COMPLEXITY OF THE RISING RESTED PROBLEM:

$\mathcal{M}_{\bar{\sigma}} := \{\mu \text{ SRRB} : \sigma \leq \bar{\sigma}\}$. Let $T \in \mathbb{N}$ and $\bar{\sigma} \in [2, \frac{T-1}{2}]$. For every algorithm $\mathcal{A}$, it holds that:

$$\sup_{\mu \in \mathcal{M}_{\bar{\sigma}}^{\det}} R_{\mu}(\mathcal{A}, T) \geq \frac{K}{64} (\bar{\sigma} - 2).$$

Thus, *any algorithm* that wishes to be no-regret, cannot avoid playing every arm a number of times proportional to $\bar{\sigma}$, suggesting that for $\Gamma \propto \bar{\sigma}$ we are tight.

FACING THE LIMITED TIME HORIZON PROBLEM-FORGETTING THE PAST:

$$\mathbb{E}[N_{i,T}] \leq O \left(\Gamma + \frac{T}{\tau} \frac{\log(T)}{(\Delta'_i(T; \tau))^2} + ((\sigma'(T; \tau) - \Gamma)\tau) \right)$$

For $\tau \sqrt{\sigma'(T; \tau)} = \Theta(\sqrt{T \log(T)})$, even when $\Gamma = 0$:

$$R(\text{Beta-SWTS}) \leq O \left(\frac{1}{\Delta^2} \sqrt{\sigma'(T; \tau) T \log(T)} \right)$$

NOVELTY: CONCENTRATION PROPERTIES FOR BETA-TS IN NON-STATIONARY SETTINGS

Beta distribution does not enjoy nice concentration properties like the Gaussian distribution.

BOUNDING A UBIQUITOUS TERM IN TS ANALYSIS FOR NON-STATIONARY SETTINGS:

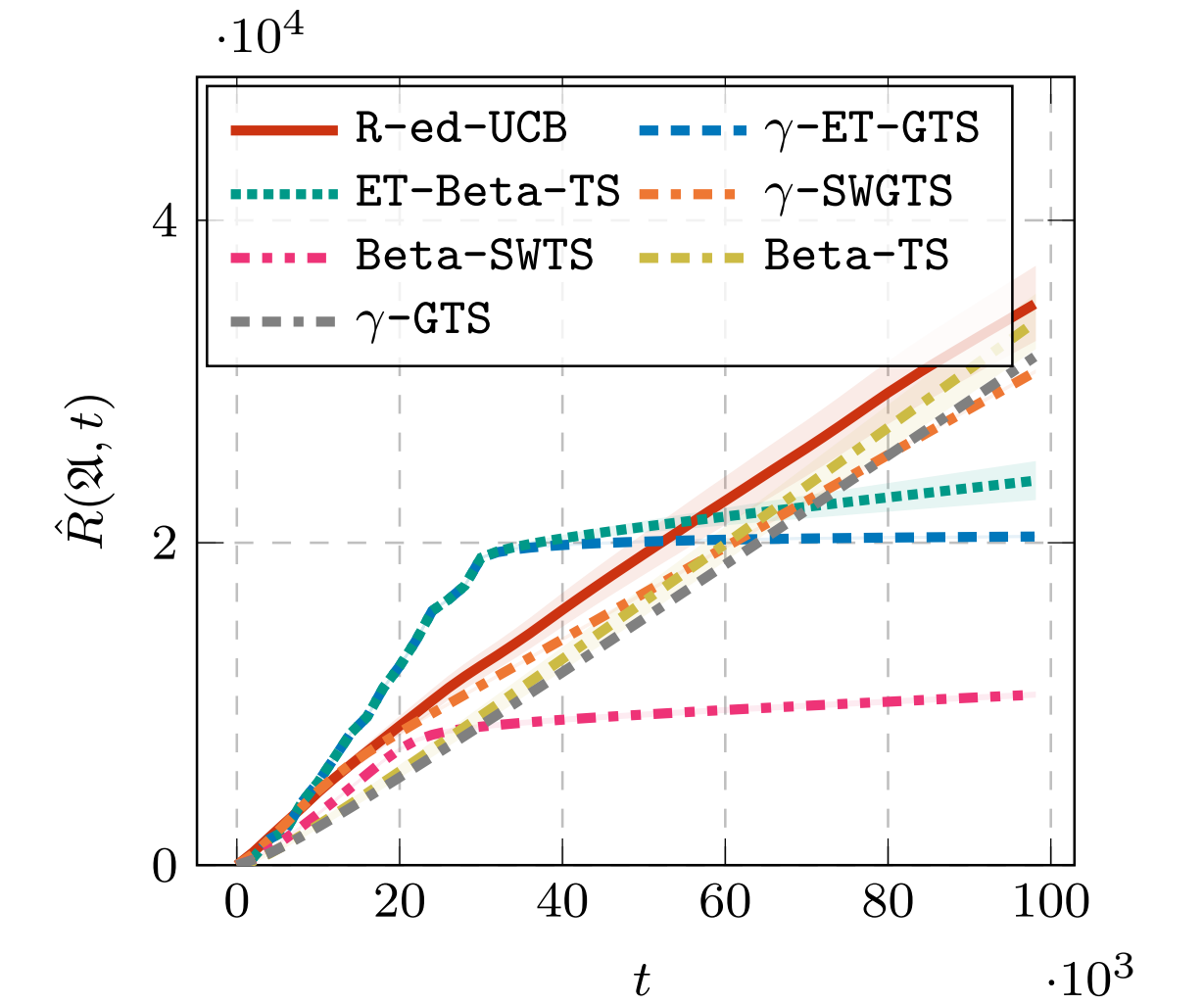
Being $p_{i,t} := \Pr(\text{Beta}(S_{1,t} + 1, F_{1,t} + 1) > y_i | \mathcal{F}_{t-1})$.

$$\mathbb{E}_{S_{1,t} \sim \text{PB}(\bar{\mu}_1(j))} \left[\frac{1}{p_{i,t}} \mid N_{1,t} = j \right] \leq \mathbb{E}_{S_{1,t} \sim \text{Bin}(j, \bar{\mu}_1(j))} \left[\frac{1}{p_{i,t}} \mid N_{1,t} = j \right] \leq \mathbb{E}_{S_{1,t} \sim \text{Bin}(j, x)} \left[\frac{1}{p_{i,t}} \mid N_{1,t} = j \right],$$

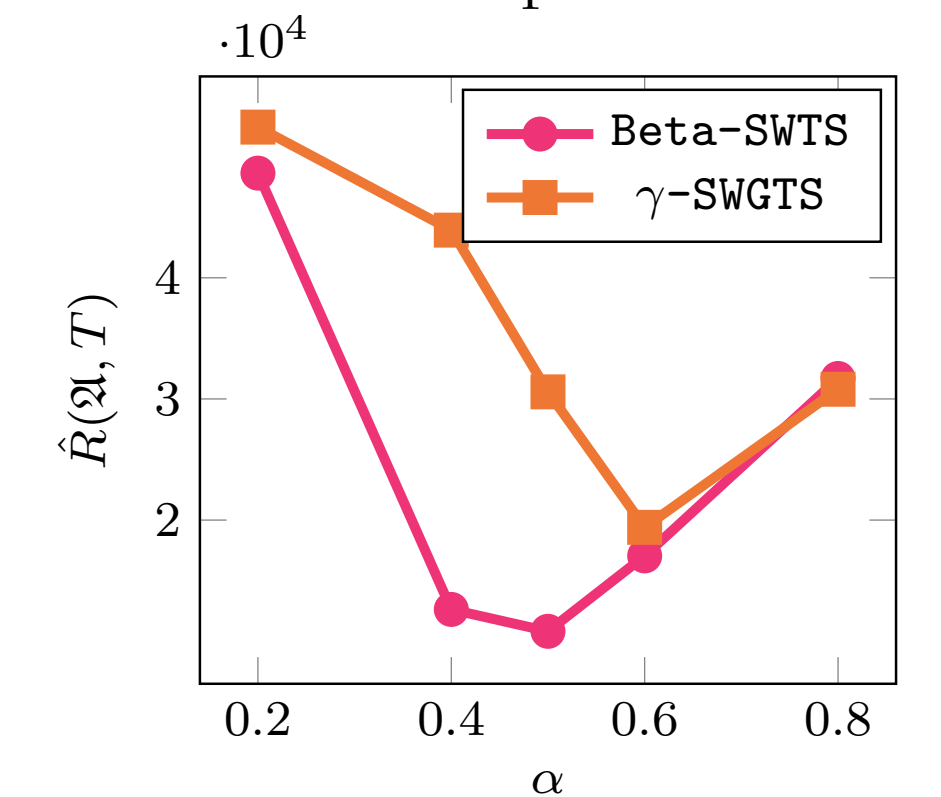
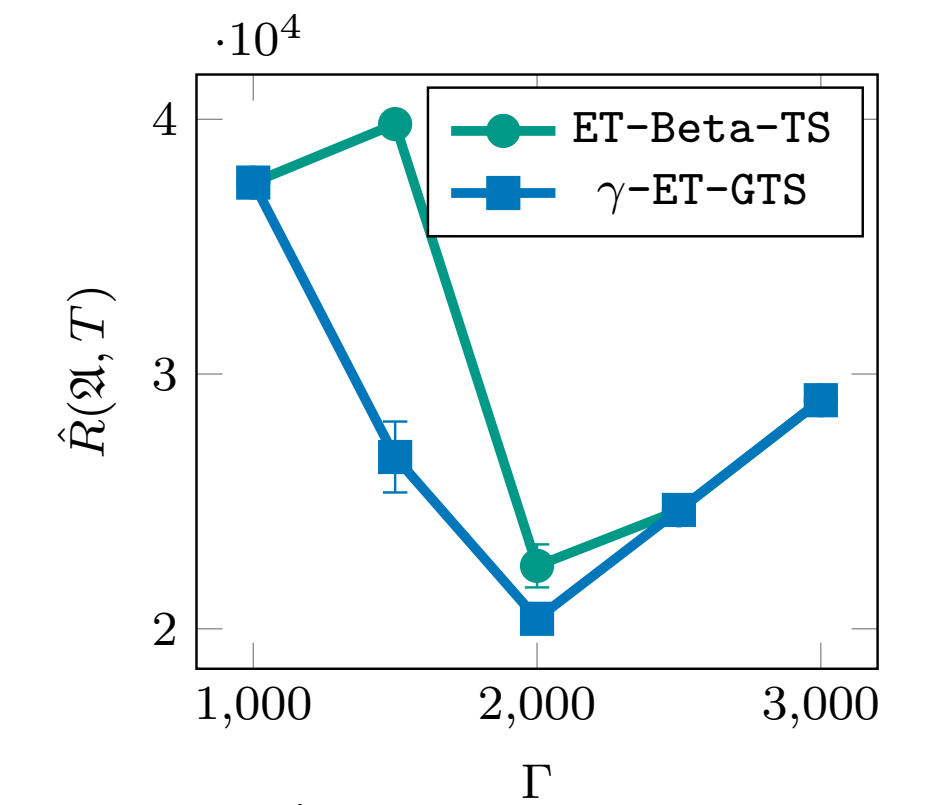
These results state that under the right circumstances those concentration properties are the same of those already retrieved in the non-stationary literature that can now be leveraged to expand the analysis of variants of Thompson Sampling to a large variety of evolving environments, making possible to widen results for rested and restless scenarios.

EXPERIMENTAL VALIDATION

ONLINE REGRET



SENSITIVITY ANALYSIS



REFERENCES

- Hoda Heidari, Michael Kearns, and Aaron Roth. Tight policy regret bounds for improving and decaying bandits. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, page 1562–1570, 2016.
- Alberto Maria Metelli, Francesco Trovò, Matteo Pirola, and Marcello Restelli. Stochastic rising bandits. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 15421–15457, 2022.