

A BNP Approach to the Multi-Armed Bandit Problem for Traits Allocation Models

Mario Beraha Federico Camerlenghi **Claudio Del Sole**
University of Milano – Bicocca, Milan, Italy,
claudio.delsol@unimib.it



Traits allocation models

Consider a **countable set of traits or features** (e.g. species, topics, genes):

- each observation displays a **finite subset** of traits, with a **level of belonging** (here, a count) for each trait;
- observations are collected from **multiple populations**, e.g. forests, books;
- the same trait may appear in different observations both **within** the same population and **across** populations → the set of traits is **shared** among populations!

$X_{11} =$	$X_{12} =$	$X_{13} =$
$X_{21} =$	$X_{22} =$	$X_{23} =$ empty
$X_{31} =$	$X_{32} = \dots$	$X_{33} =$
$X_{41} = \dots$		$X_{43} = \dots$

Consider d **populations**, with n_j **observations** collected from population $j = 1, \dots, d$. The observation X_{ij} from population j is represented as

$$X_{ij} = \sum_{k \geq 1} A_{ijk} \delta_{\psi_k}, \quad i = 1, \dots, n_j,$$

where $A_{ijk} \geq 0$ is the **count** for trait ψ_k in population j .

Nonparametric model: each observation X_{ij} is an i.i.d. realization of a **Poisson process**, driven by the finite discrete measure μ_j :

$$A_{ijk} \stackrel{\text{ind}}{\sim} \text{Poisson}(\lambda_{jk}), \quad \mu_j = \sum_{k \geq 1} \lambda_{jk} \delta_{\psi_k}.$$

Some useful notation:

- the **total count** for trait X_k in population j is $M_{jk} = \sum_{i=1}^{n_j} A_{ijk}$;
- the set of traits appearing **at least once** in population j is $\mathcal{K}_j = \{k \geq 1 : M_{jk} > 0\}$.

Multi-armed bandit for traits discovery

Problem: choose from which population to sample the next observation, with the goal of **discovering new traits**.

$X_{11} =$	$X_{12} =$	$X_{13} =$
$X_{21} =$	$X_{22} =$	$X_{23} =$ empty
$X_{31} =$	$X_{32} =$	$X_{33} =$
$X_{41} =$		$X_{43} =$

Multi-armed bandit formulation: sequentially select one among the possible **actions** (sampling from a certain population) to maximize the cumulative **reward** (the number of discovered traits).

The **random reward** of population j is the number of **new traits** discovered in a new observation:

$$R_j = \sum_{k \geq 1} \mathbb{1} \left(A_{(n_j+1)jk} > 0, \sum_{j=1}^d M_{jk} = 0 \right) = \sum_{k \notin \mathcal{K}} \mathbb{1} \left(A_{(n_j+1)jk} > 0 \right),$$

where $\mathcal{K} = \cup_{j=1}^d \mathcal{K}_j$ is the set of traits appearing in **at least one** population.

Optimal strategy: choose the population **maximizing** the conditional **expected reward**, given the past observations,

$$\mathbb{E}[R_j | \mathbf{X}] = \sum_{k \notin \mathcal{K}} (1 - e^{-\lambda_{jk}}), \quad j = 1, \dots, d.$$

Statistical problem: how to **estimate** the population **parameters** μ_1, \dots, μ_d ?

Bayesian nonparametric approach

Hierarchical prior: the distribution of measures μ_1, \dots, μ_d is a hierarchical **process**:

$$\mu_j = \sum_{k \geq 1} \lambda_{jk} \delta_{\psi_k}, \quad \lambda_{jk} | \lambda_{0k} \stackrel{\text{ind}}{\sim} \text{Gamma}(\theta_j \lambda_{0k}, 1), \quad j = 1, \dots, d,$$

$$\mu_0 = \sum_{k \geq 1} \lambda_{0k} \delta_{\psi_k} \sim \text{CRM}(\theta_0 \rho),$$

- where $\theta_0, \theta_1, \dots, \theta_d > 0$ are **concentration parameters**,
- and ρ is the Lévy measure characterizing the **distribution of jumps** $(\lambda_{0k})_{k \geq 1}$ in μ_0 .

Posterior distribution: the posterior of the parameters for the **unseen traits** is **tractable**:

$$\mu_j^* = \sum_{k \notin \mathcal{K}} \lambda_{jk} \delta_{\psi_k}, \quad \lambda_{jk} | \lambda_{0k}, \mathbf{X}_j \stackrel{\text{ind}}{\sim} \text{Gamma}(\theta_j \lambda_{0k}, 1 + n_j), \quad j = 1, \dots, d,$$

$$\mu_0^* = \sum_{k \notin \mathcal{K}} \lambda_{0k} \delta_{\psi_k} | \mathbf{X} \sim \text{CRM}(\theta_0 \rho^*), \quad \rho^*(ds) = \prod_{j=1}^d (1 + n_j)^{-\theta_j s} \rho(ds).$$

Important remarks:

- the posterior parameters of unseen traits depend exclusively on the **number of observations** in each population!
- to model **biodiversity** across populations, consider exponential **hyperpriors** on concentration parameters $\theta_0, \theta_1, \dots, \theta_d$.

Thompson sampling for traits discovery

Since population parameters are **unknown**, common **heuristic** strategies balance:

- exploration**, selecting less sampled populations to learn their parameters;
- exploitation**, selecting populations with highest chances of seeing new traits.

Thompson sampling: choose the population **randomly**, according to the **posterior probability** of maximizing the conditional expected reward:

$$p_j = \int \mathbb{1} \left(j = \arg \max_{\ell} \mathbb{E}[R_{\ell} | \mathbf{X}, \boldsymbol{\mu}] \right) p(\boldsymbol{\mu} | \mathbf{X}).$$

In practice, consider two possible **implementations**:

- standard Thompson sampling:**
 - sample posterior measures μ_1, \dots, μ_d ;
 - select the population $j^* = \arg \max_j \mathbb{E}[R_j | \mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\theta}] = \arg \max_j \sum_{k \notin \mathcal{K}} (1 - e^{-\lambda_{jk}})$
- partially marginalized Thompson sampling:**
 - sample posterior parameters $\theta_1, \dots, \theta_d$;
 - select the population $j^* = \arg \max_j \mathbb{E}[R_j | \mathbf{X}, \boldsymbol{\theta}] \approx \arg \max_j \theta_j \log \left(1 + \frac{1}{1+n_j} \right)$.

Idea: the more you **marginalize**, the more you **exploit**!

Trees species in Japanese forests

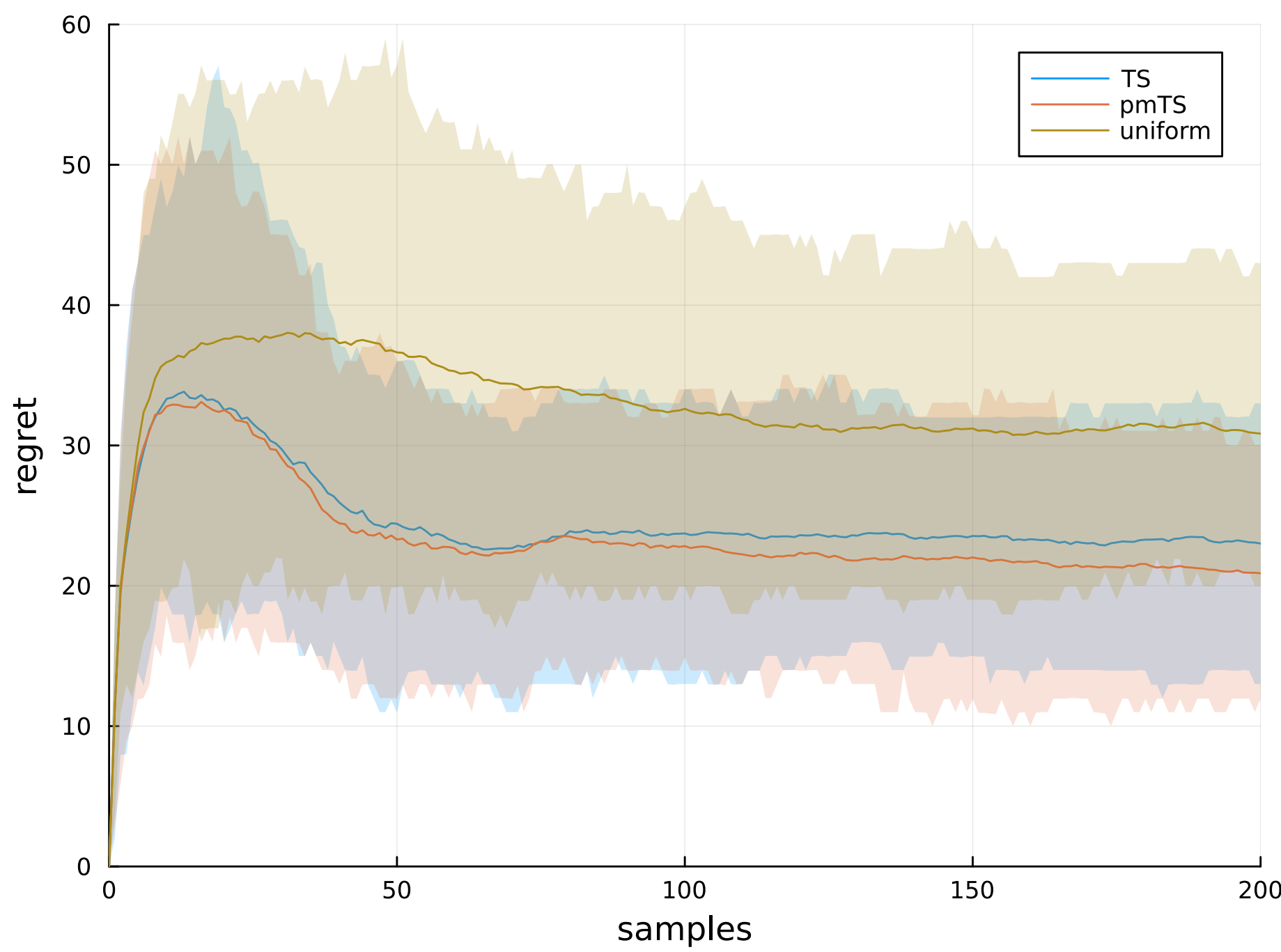
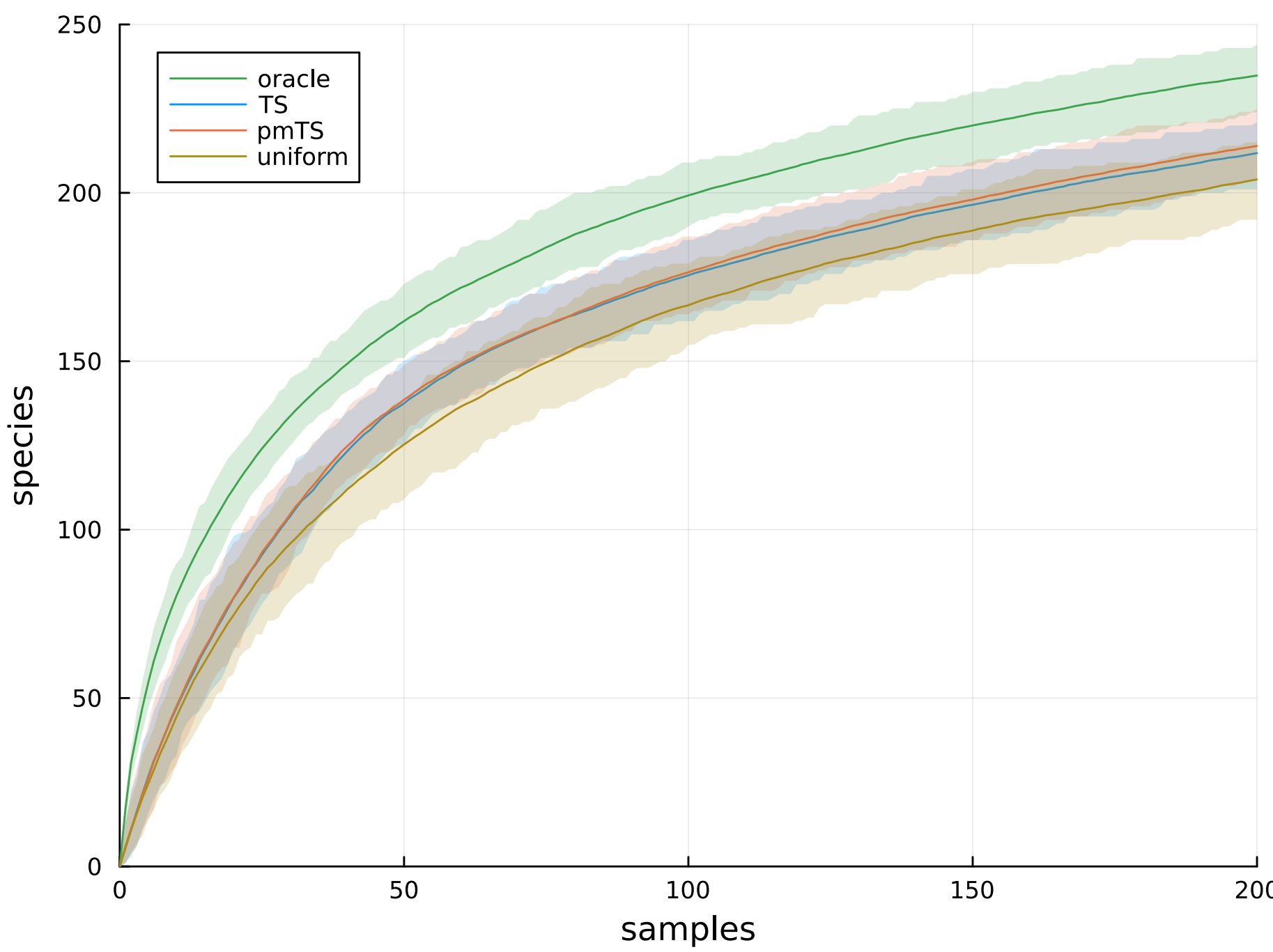
Dataset containing trees species counts from ecological **survey plots** collected in various forest sites in Japan:

- consider $d = 41$ **populations** (forest plots) classified in 4 macro-types;
- around 100 **observations** per forest → trees counts in 10×10 m areas;
- records for 365 **species** across forest sites.

Preliminary analysis:

- number of **discovered** species;
- **regret** from optimal strategy.

Results are averaged over 100 experiments.



[1] Battiston, Favaro & Teh (2018). *Multi-Armed Bandit for Species Discovery: A Bayesian Nonparametric Approach*. Journal of the American Statistical Association.

[2] Camerlenghi, Dumitrascu, Ferrari, Engelhardt & Favaro (2020). *Nonparametric Bayesian Multiarmed Bandits for Single-Cell Experiment Design*. The Annals of Applied Statistics.

[3] Auer, Cesa-Bianchi, Fisher (2002). *Finite-time Analysis of the Multiarmed Bandit Problem*. Machine Learning.

[4] Masoero, Camerlenghi, Favaro, Broderick (2023). *Posterior representations of hierarchical completely random measures in trait allocation models*. NeurIPS Workshop.