Multi-armed bandit for species discovery: a Bayesian nonparametric approach

Multi-armed bandit for species discovery: a Bayesian nonparametric approach

Marco Battiston, Stefano Favaro and Yee Whye Teh*

ABSTRACT. Let $(P_1, \ldots, P_J)$ denote J populations of animals from distinct regions. A priori, it is unknown which species are present in each region and what are their corresponding frequencies. Species are shared among populations and each species can be present in more than one region with its frequency varying across populations. In this paper we consider the problem of sequentially sampling these populations in order to observe the greatest number of different species. We adopt a Bayesian nonparametric approach and endow $(P_1, \ldots, P_J)$ with a Hierarchical Pitman-Yor process prior. As a consequence of the hierarchical structure, the J unknown discrete probability measures share the same support, that of their common random base measure. Given this prior choice, we propose a sequential rule that, at every time step, given the information available up to that point, selects the population from which to collect the next observation. Rather than picking the population with the highest posterior estimate of producing a new value, the proposed rule includes a Thompson sampling step to

*Marco Battiston is Ph.D student of Statistics, Department of Decision Sciences, Bocconi University, Italy. (E-mail address: marco.battiston@phd.unibocconi.it). Stefano Favaro is Associate Professor of Statistics, Department of Economics and Statistics, University of Torino and Collegio Carlo Alberto, Italy. (E-mail address: stefano.favaro@unito.it). Yee Whye Teh is Professor of Statistical Machine Learning, Department of Statistics, University of Oxford, United Kingdom. (E-mail address: y.w.teh@stats.ox.ac.uk)

better balance the exploration-exploitation trade-off. We also propose an extension of the algorithm to deal with incidence data, where multiple observations are collected in a time period. The performance of the proposed algorithms is assessed through a simulation study and compared to three other strategies. Finally, we compare these algorithms using a dataset of species of trees, collected from different plots in South America.

KEYWORDS. Bayesian nonparametric statistic; discovery probability; hierarchical Pitman-Yor process; multi-armed bandit; species sampling models; Thompson Sampling.

1. INTRODUCTION

Species sampling problems have a long history in ecological and biological studies and interest in them has recently grown in many other fields, like machine learning, linguistics and genetics. The setting of these problems is very general. Specifically, a sample of size n is collected from a discrete population, denoted by P, and interest lies either in predicting the realization of future observations or in estimating some particular feature of P. Using a species metaphor, we can think of P as a population of animals of different species. Each observation is an animal and its realized value is its specific species. Statistical issues common in species sampling problems are, for instance, making inference on the number of unseen species or estimating the probability that in a further sample of m units k new distinct species will be observed. Starting from the seminal works of Good (1953), Good and Toulmin (1956) and Efron and Thisted (1976), a full range of statistical approaches, parametric and nonparametric as well as frequentist and Bayesian, have been proposed for making inference in species sampling problems. See, e.g., Chao (1981), Chao and Lee (1992), Mao and Lindsay (2002), Mao (2004), Lijoi et al. (2007), Ionita-Laza et al. (2009), Barger and Bunge (2010) and Favaro et al. (2012a).

In this work we consider a different setting for species sampling problems. Let $(P_1, \ldots, P_J)$ denote J populations of animals of distinct regions. Each population P_j is assumed to contain a large number of species of animals, but both the represented species and their frequencies are unknown a priori. Also, the J populations are allowed to share the same species of animals and each species can have different frequencies in distinct regions. In this paper we consider the problem of sequentially sampling these populations with the goal of maximizing the number of distinct species observed. In ecology and biology this problem arises naturally when different environments are explored in search of new species. In order for the exploration to remain cost-effective, redirecting it to a different environment is necessary whenever the probability of discovering a new species at the next draw becomes unacceptably low in the current environment. Similarly, in genetics the goal is to increase the number of genetic variants one expect to discover. See, e.g., Ionita-Laza et al. (2009) and Ionita-Laza and Laird (2010) where it has been shown that combining data from multiple populations in a discovery study increase the number of genetic variants identified relative to studies on single populations. Other applications aries in electrical engineering, in the context of security analysis of electrical power systems, and in software engineering, in the context of bug discovery. See, e.g, Fonteneau-Belmudes et al. (2010), Bubeck et al. (2013b) and references therein.

The framework of our species sampling problem resembles that of stochastic multi-armed bandit (MAB) problems. These are problems in reinforcement learning which can be described using a gambling metaphor. We imagine a gambler facing J slot machines ("one armed bandit" is the colloquial term for a slot machine in American slang) with different unknown reward distribution functions. At every step, given the history of plays and realized rewards, the player can choose on which machine to play next and he will receive a random reward from the distribution of that arm. In the bandit lit-

erature the most common formalization of the problem is that of independent rewards from J unknown distributions. In this setting the two most popular sequential strategies are the Upper Confidence Bound (UCB) algorithm, introduced in Lai and Robbins (1985) and further developed by Auer et al. (2002), and the Thompson sampling (TS), proposed by Thompson (1933). The former solves the classical exploration-exploitation trade-off inherent in any bandit problem by constructing deterministic upper bounds for the expected reward in each arm, and then playing the arm with the highest bound value. The latter is a Bayesian algorithm which assigns priors to the unknown parameters and plays an arm according to its posterior probability of being the best one.

The problem analyzed in this paper can be traced back to a similar bandit formalization. In our setting, at every time step, a reward of one unit is received if a new species is observed and zero otherwise. Hence, rewards are Bernoulli distributed, but with dependent and time varying success probabilities. Indeed every time a new species is observed the probability of observing another new species in the next steps decreases. We propose a sequential rule that, given the information available up to a point, select the population from which to collect the next observation, with the goal of discovering as many distinct species as possible. Our sequential rule contains two elements: a Bayesian nonparametric procedure for the estimation of the $(P_1, \ldots, P_J)$, seen as random discrete probability measures on a suitable space of species of animals, together with a TS strategy for the sequential choice of the best arm. We choose a Hierarchical Pitman-Yor (HPY) process as nonparametric prior for the unknown populations $(P_1, \ldots, P_J)$. This prior choice induces a prior also for the J mean parameters of the Bernoulli reward distributions. Given the induced prior and given a set of data from $(P_1, \ldots, P_J)$, we derive the corresponding posterior, which is then used to implement the TS strategy. We refer to this strategy as HPY-TS. In addition, we also

propose an extension of it, to deal with incidence data, in which animals are collected in groups.

There is a connection between our strategy and adaptive sampling techniques. These are modifications of stratified random sampling, see e.g. Cochran (1977) chapter 5, in which the choice of the sampling units are not fixed prior to making observations, but units are sequentially chosen depending on previously observed values of some variable of interest. Theoretical advantages of adaptive selection designs were already pointed out in Basu (1969) and Zacks (1969) and can be remarkable, particularly when dealing with rare or elusive species. The first successful adaptive sampling procedure was by Thompson (1990), who proposed adaptive cluster sampling. With this technique, biologists search for rare species of interest close to locations where the species was previously observed. Extensions and refinements of Thompson's work can be found in subsequent papers, e.g. Thompson (1991a), Thompson (1992) and Thompson (1991b). Two good references for adaptive sampling techniques are Thompson (2002) and Thompson and Seber (1996). Our algorithms have a similar flavor, but rather than focusing on re-observing a particular rare species, the goal is now to detect new ones.

We assess the performances of the proposed HPY-TS algorithms through simulations and using a dataset from biology. We compare the HPY-TS algorithms to three other strategies: an Oracle strategy in which the composition of the $(P_1, \ldots, P_J)$ are known; a Uniform strategy that selects at every step a population uniformly at random; a rule recently proposed in Bubeck et al. (2013b) based on the Good-Turing missing mass estimator introduced in Good (1953). Our simulation study considers different scenarios, by varying the level of heterogeneity in species variety among populations. Simulated results show that the HPY-TS performs better than the Uniform and the strategy of Bubeck et al. (2013b) in all scenarios, discovering more new species both in

the abundance and in the incidence case. These results suggests also that the HPY-TS algorithms are robust to changes in the level of heterogeneity in species variety across the J populations, without the need of tuning parameters to regulate the exploration rate. We also compare the algorithms using a dataset of species of trees, collected in different plots in South America, analyzed in Pyke et al. (2001) and Condit et al. (2002).

This work is organized as follows. Section 2 reviews the Pitman-Yor process, its hierarchical counterpart, and the MAB problem. Section 3 introduces HPY-TS algorithms for abundance and incidence data. Section 4 describes the setting and reports the results of the simulation study and of the real data illustration. Concluding remarks close the work. Proofs, the MCMC sampler used to estimate the parameters of the HPY model, and additional numerical results are available as online supplementary material.

2. PRELIMINARIES

2.1 The Pitman-Yor process

The two parameter Poisson-Dirichlet process, now known as the Pitman-Yor (PY) process, was introduced in Pitman and Yor (1997) as a generalization of the Dirichlet process by Ferguson (1973). Like the Dirichlet process, the PY process is a probability measure that assigns probability one to the set of discrete distributions. It is parametrized by (σ, θ, P_0), where P_0, called base distribution, is a distribution on the sample space, and σ and θ are two scalars satisfying $0 \leq \sigma < 1$ and $\theta > -\sigma$, respectively called the concentration and the mass parameter. The Dirichlet process corresponds to the special case $\sigma = 0$. The PY process admits a stick breaking representation. Specifically, if P is a random probability measure distributed according to $\text{PY}(\sigma, \theta, P_0)$, then $P \overset{d}{=} \sum_{i \geq 1} p_i \delta_{Y_i^*}$, where $(Y_i^*)_{i \geq 1}$ are i.i.d. random variables with distribution P_0, and

$(p_i)_{i\geq 1}$ are such that $p_i = V_i \prod_{1\leq k\leq i-1}(1-V_k)$, with $V_i \sim \text{beta}(1-\sigma, \theta+i\sigma)$ for all $i \geq 1$.

A description of the posterior distribution of the PY process was derived in Pitman (1996). Given a sample $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$, such that $Y_i|P \overset{iid}{\sim} P$ for all $1 \leq i \leq n$ and $P \sim \text{PY}(\sigma, \theta, H)$, the posterior of P given $\mathbf{Y}_n$ satisfies the following distributional equation

$$P|\mathbf{Y}_n \overset{d}{=} \sum_{i=1}^{K_n} w_i \delta_{Y_i^*} + w_0 \tilde{P}, \tag{1}$$

where K_n is the number of distinct values in the sample $\mathbf{Y}_n$, denoted by $(Y_1^*, \ldots, Y_{K_n}^*)$ and having multiplicities $(n_1, \ldots, n_{K_n})$, $(w_0, w_1, \ldots, w_{K_n})$ is a random vector distributed according to $\text{Dir}(\theta + K_n\sigma, n_1 - \sigma, \ldots, n_{K_n} - \sigma)$ and $\tilde{P} \sim \text{PY}(\sigma, \theta + K_n\sigma, H)$. Using this posterior representation, it is also possible to derive the so called Chinese Restaurant representation of the PY process, which describes the conditional distribution of the next observation, when the underlying random distribution P has been integrated out, that is

$$Y_{n+1}|\mathbf{Y}_n, \sigma, \theta, P_0 \sim \sum_{i=1}^{K_n} \frac{n_i - \sigma}{\theta + n} \delta_{Y_i^*} + \frac{\theta + K_n\sigma}{\theta + n} P_0.$$

Following this predictive distribution, the observation Y_{n+1} is assigned to an old cluster with value Y_i^* with probability proportional to $n_i - \sigma$, or it is sampled from P_0 and forms its own cluster with probability proportional to $\theta + K_n\sigma$. See Pitman (1996) for details.

The base distribution P_0 corresponds to the mean of the process P, i.e, $\mathbb{E}[P(\cdot)] = P_0(\cdot)$. Furthermore, from the stick breaking representation it is clear that it is also the distribution of the locations of support points of the random distribution P. The mass parameter θ regulates the variance of the prior around the prior mean P_0. Specifically, with high θ, the prior guess is strong, i.e. the variance of the prior is small, whereas with low θ there is more uncertain about possible values of P. A posteriori, a high θ implies more weight on prior information, while a low θ more weight on the informa-

tion given from the sample. This fact can also be read from the Chinese Restaurant representation. In fact, the weight of the prior mean is higher when θ is high. The concentration parameter σ affects how the total mass of P is spread across its support points. When σ is low, the prior samples with high probability distributions with few points of very high mass. With a value of σ close to 1, distributions with the total mass evenly spread across many support points are more likely to be sampled.

The PY process offers more flexibility than the Dirichlet process and it has desirable properties when dealing with species sampling problems. In particular, it offers a flexible predictive structure in which the probability of observing a new species depends not only on the sample size, like for example in Dirichlet process, but also on the number of distinct values observed in the sample. At the same time, it maintains mathematical tractability. Moreover, differently from the Dirichlet process, with a PY process the number of distinct observations K_n grows following a power law behavior as the sample size increases, a feature common to many real world datasets, as observed in Mitzenmacher (2004) and Goldwater et al. (2006). It is a useful prior when the population size is large but unknown and the number of species in the population is also unknown. However, it is not an adequate prior when the goal is estimating the total number of species in the population. Indeed, from the stick-breaking representation, a sample from a PY process has an infinite number of atoms with probability one. Hence, a point estimate for the total number of species in the population is always infinite.

2.2 The Hierarchical Pitman-Yor process

The HPY process was introduced in Teh (2006). See also the review by Teh and Jordan (2010). This is a useful model in problems in which multiple groups of data are available and where we wish to introduce probabilistic dependence across populations. In particular, it is an appropriate model when data incorporates a discrete variable

of unknown cardinality. The discrete variable can either be at the observations level, as in our context, or at the latent level, like when it parametrizes the distribution of continuous observations or when it works as a classification variable in mixture models settings.

To describe the HPY process we use a similar notation as in Teh and Jordan (2010). Specifically, $\mathbf{Y}_{n_{j\cdot}} = \left(Y_{j,1}, \ldots, Y_{j,n_{j\cdot}}\right)^T$ denotes the column vector of observations from the j-th population, K stands for total number of distinct values $(Y_1^{**}, \ldots, Y_K^{**})$ observed in the joined sample containing observations from all populations, n_{jtk} denotes the number of observations in population j, belonging to cluster t and having value Y_k^{**}, while m_{jk} is the number of clusters in population j with value Y_k^{**}, $(Y_{j,1}^*, \ldots, Y_{j,m_{j\cdot}}^*)$ are the values of the $m_{j\cdot}$ clusters in population j. As in Teh and Jordan (2010), dots in the indexes denote that we are summing over that index, e.g. $n_{j\cdot}$ is the number of observations from the j-th population.

The HPY process is described by the hierarchical representation

$$\begin{aligned} Y_{j,i}|P_j &\overset{iid}{\sim} P_j \qquad j = 1, \ldots, J \; i = 1, \ldots, n_{j\cdot} \\ P_j|\sigma_j, \theta_j, P_0 &\overset{ind}{\sim} \mathrm{PY}\left(\sigma_j, \theta_j, P_0\right) \qquad j = 1, \ldots, J \\ P_0|\alpha, \gamma, H &\sim \mathrm{PY}\left(\alpha, \gamma, H\right), \end{aligned}$$

where H is a fixed and diffuse probability measure and the $J+1$ couples of hyperparameters (σ_j, θ_j) and (α, γ) are chosen to satisfy the conditions $\sigma_j, \alpha \in [0, 1)$, $\theta_j > -\sigma_j$ and $\gamma > -\alpha$, for all $j \in \{1, \ldots, J\}$. Also, the hyperparameters σ_j, θ_j, α and γ are usually assumed to be unknown and endowed with priors. In the HPY process, observations from the j-th group, when conditioned to the realization of the unknown P_j, are independent and identically distributed with distribution P_j. Moreover, they are conditionally independent of observations from other populations. The P_j's are treated as random objects and endowed with PY processes with the same base measure P_0. This latter hyperparameter is not fixed by the modeler, but is considered as a random

element to be inferred from data. Another PY process is used as nonparametric distribution for P_0. Due to the almost sure discreteness of P_0, this recursive construction has the effect that the support of the P_j's is contained in that of P_0. As a consequence, all populations share the same random support of P_0.

The HPY process admits a useful representation in terms of the so-called Chinese Restaurant Franchise process. See Teh and Jordan (2010) for details on this culinary metaphor. The Chinese Restaurant Franchise representation of the HPY process is summarized by the following predictive distributions for the observables and for the cluster values in population j

$$Y_{j,i+1}|Y_{j,1},\ldots,Y_{j,i},\sigma_j,\theta_j,P_0 \sim \sum_{t=1}^{m_{j\cdot}} \frac{n_{jt\cdot}-\sigma_j}{\theta_j+n_{j\cdot\cdot}}\delta_{Y^*_{j,t}} + \frac{\theta_j+m_{j\cdot}\sigma_j}{\theta_j+n_{j\cdot\cdot}}P_0 \tag{2}$$

and

$$Y^*_{j,m_{j\cdot}+1}|Y^*_{1,1},\ldots,Y^*_{J,m_{J\cdot}},\alpha,\gamma,H \sim \sum_{k=1}^{K} \frac{m_{\cdot k}-\alpha}{\gamma+m_{\cdot\cdot}}\delta_{Y^{**}_k} + \frac{\gamma+K\alpha}{\gamma+m_{\cdot\cdot}}H. \tag{3}$$

Equation (2) is the Chinese Restaurant representation of P_j. The new observation $Y_{j,i+1}$ belongs to an old cluster $Y^*_{j,t}$ with probability proportional to $n_{jt\cdot}-\sigma_j$ or it forms a new cluster and it is sampled from P_0 with probability proportional to $\theta_j+m_{j\cdot}\sigma_j$. The sequence of cluster values is sampled from P_0, which, being distributed as a PY process, also admits a Chinese Restaurant representation, summarized by Equation (3). The new cluster in population j has the same value as one already observed in the joined sample with probability proportional to $m_{\cdot\cdot}-K\alpha$ or it has a new value, sampled from H, with probability proportional to $\gamma+K\alpha$.

In the HPY process, the hyperparameters (σ_j,θ_j) have the same interpretation as in the PY case. By contrast, the hyperparameters (α,γ) regulate the total number and the sharing of cluster values among populations. If γ is low, the total number of cluster values K will be very low on average and, when a new sample from P_0 is needed, it will coincide with high probability with an already observed one. If α is high, the sharing

of cluster values among populations is low, while, with α low, there is a small set of popular cluster values which are seen many times among all populations.

2.3 The multi-armed bandit problem and Thompson sampling

A MAB problem is a sequential allocation problem under limited information. We imagine J slot machines with unknown reward distributions. The goal is to maximize the expected pay-off by exploiting machines that give high profits while exploring machines for which we have limited information. See Bubeck and Cesa-Bianchi (2012) for a review. In the stochastic formalization, the J slot machines have unknown reward distributions functions and, at every time step, a draw from the distribution of the chosen machine is collected. A strategy is a sequential rule that, given the history up to that point, chooses the next arm to play. To evaluate its performances, its expected total reward is usually compared with that of an "Oracle" strategy, the strategy that chooses the arm with the highest expected payoff, when uncertainness about the reward distributions is removed. The difference from their total rewards is termed regret. The goal is to find strategies that minimize the expected regret.

Two popular strategies have been shown to effectively address the stochastic bandit problem: the UCB algorithm and TS. The UCB algorithm was initially suggested by Lai and Robbins (1985) and further developed by Auer et al. (2002). This algorithm constructs a deterministic upper confidence bound for the expected reward of each arm and then plays the arm with highest bound. This algorithm has good theoretical guarantees for the i.i.d. case: Auer et al. (2002) proved that its expected regret matches, up to a constant factor, the lower bound of Lai and Robbins (1985). This is a lower bound for the expected regret of any strategy satisfying mild conditions, in the i.i.d. context. TS was initially proposed by Thompson (1933) as a randomized Bayesian algorithm to minimize regret in a clinical trial setting. The idea is to assume a prior for the unknown parameters in the distributions of each arm and, at every time step,

play an arm according to its posterior probability of being the best one. Its most canonical examples are for Bernoulli bandits. In this setting, a Bernoulli distribution with unknown parameter is set as reward distribution for each arm, and the unknown reward means are endowed with a Beta prior distributions. TS thus consists in sampling a draw from each of these J Beta distributed posteriors and then play the arm with the highest realization.

Even though it was proposed eighty years ago, TS has attracted attention only recently. Several recent studies have empirically demonstrated the efficacy of TS. Chapelle and Li (2011) have empirically demonstrated that TS achieves regret comparable to the lower bound of Lai and Robbins (1985). In addition, the algorithm is more robust to delayed or batched feedback than other methods. Chapelle and Li (2011) also show that TS performs equally or better of popular methods, such as UCB algorithms, in applications like display advertising and news article recommendation. Other empirical works on TS or randomized probability matching algorithm (to which TS belongs) are Granmo (2010), Scott (2010) and May and Leslie (2011). Theoretical investigations of the TS are in Agrawal and Goyal (2012), Kaufmann et al. (2012), Russo and Van Roy (2014) and Szabo and Tran-Thanh (2015). A recent promising theoretical result for TS is in Russo and Van Roy (2016), where the authors provide Baysian regret bounds for a broad range of on-line optimization algorithms, with TS being a particular case.

3. HPY-TS ALGORITHM

The problem of sequential species discovery in presence of many populations can be cast as a stochastic bandit problem by imagining each population produces a reward when sampled: 1 if a new species is discovered and 0 otherwise. It is thus a Bernoulli bandit problem with reward probabilities that depend on past observations.

The reward probabilities are non-increasing functions of the number of draws, because every time a new species is observed the remaining missing mass will be lower in following time steps. Subsection 3.1 derives the joint posterior of these J Bernoulli means and introduces the HPY-TS algorithm in the case of abundance data. Subsection 3.2 proposes an extension of this algorithm to deal with incidence data.

3.1 Abundance data

In the abundance data scenario, a single animal is observed at each sampling time. Given a model choice, the TS draws a value for each population from its posterior probability of being the best arm. In the species discovery problem with many populations, this posterior distribution is given by the joint distribution of the J random probabilities of observing a new value in each arm, given all observations. Proposition 1 derives the relevant probability in the case of a HPY process for $(P_1, \ldots, P_J)$. In particular, denoting with $\mathbf{Y}_{n_{j\cdot}} = \left(Y_{j,1}, \ldots, Y_{j,n_{j\cdot}}\right)^T$ the vector of observations from the population j taking values on a measurable space $\mathcal{Y}$, with $\mathbf{Y_n} = (\mathbf{Y}_{n_{1\cdot}}, \ldots, \mathbf{Y}_{n_{J\cdot}})$ the joint sample (the array containing observations from all populations) and with $A = \{y \in \mathcal{Y} : y \notin \mathbf{Y_n}\}$ the set of possible new species, what is needed is the distribution of

$$(P_1(A), \ldots, P_J(A)) | \mathbf{Y_n}, \sigma_1, \ldots, \sigma_J, \theta_1, \ldots, \theta_J, \alpha, \gamma, H.$$

For ease of notation, from now on we omit the reference to the hyperparameters of the HPY process, $\sigma_j, \theta_j, \alpha, \gamma, H$ when conditioning on them. The density function of this joint distribution is provided in the following proposition, whose proof is available in the on-line supplementary material. In its statement, we adopt the notation for table counts and distinct values previously introduced for the Chinese Franchise Representation of the HPY process. Also, $\text{beta}(p|a, b)$ stands for a beta density function with parameters a and b, evaluated at p.

Proposition 1. *Let* $\mathbf{Y_n}$ *denote the joined sample from a HPY proces and let* $A =$

$\{y \in \mathcal{Y} : y \notin \mathbf{Y_n}\}$. *Then,* $(P_1(A), \ldots, P_J(A))|\mathbf{Y_n}$ *admits the following multivariate density*

$$f_{(P_1(A),\ldots,P_J(A))|\mathbf{Y_n}}(p_1, \ldots p_J) = \int_0^1 \prod_{j=1}^{J} f_j(p_j|\beta_0, m_{j\cdot}, n_{j\cdot\cdot}) f_0(\beta_0|K, m_{\cdot\cdot}) d\beta_0,$$

where

$$f_j(p_j|\beta_0, m_{j\cdot}, n_{j\cdot\cdot}) = beta(p_j|(\theta_j + m_{j\cdot}\sigma_j)\beta_0, (\theta_j + m_{j\cdot}\sigma_j)(1-\beta_0) + n_{j\cdot\cdot} - \sigma_j m_{j\cdot})$$

and

$$f_0(\beta_0|K, m_{\cdot\cdot}) = beta(\beta_0|\gamma + K\alpha, m_{\cdot\cdot} - \alpha K).$$

The following corollary provides a Bayesian nonparametric point estimate of the missing mass for each populations. This result follows by a direct application of Proposition 1.

Corollary 1. *Under squared loss function, the Bayesian nonparametric point estimate for the probability of discovering a new value in population* j*,given the joined sample* $\mathbf{Y_n}$*, is*

$$\mathbb{E}[P_j(A)|\mathbf{Y_n}] = \left(\frac{\theta_j + m_{j\cdot}\sigma_j}{\theta_j + n_{j\cdot\cdot}}\right)\left(\frac{\gamma + K\alpha}{\gamma + m_{\cdot\cdot}}\right).$$

With the posterior distribution obtained in Proposition 1, HPY-TS strategy prescribes to sample a draw from it and to select the population with the highest realized value. This strategy usually outperforms the greedy strategy that selects the arm with the highest posterior point estimate, $j^{\text{greedy}} = \text{argmax}\{\mathbb{E}[P_j(A)|\mathbf{Y_n}] : j \in \{1, \ldots, J\}\}$, since it better balances the exploration step. Intuitively, suppose to have only a few observations, with an unlucky sample, from a "winning" arm (a population with a very high species variety), resulting in a low point estimate for its missing mass. This

population will not be chosen by the greedy strategy, which only exploits arms with good past behavior. Whereas, with HPY-TS strategy, the posterior distribution of the missing mass of this population will have high variance, due to the small sample size. This implies a positive probability for that arm to be chosen, if its Thompson draw results in a high value. The HPY-TS strategy for abundance data is summarized in Algorithm 1.

Algorithm 1: HPY-TS - Abundance Data

for *i in 1:additional sample* **do**

draw $\beta_0 \sim \text{beta}\left(\gamma + K\alpha, m_{\cdot\cdot} - \alpha K\right)$;

for *j in 1:J* **do**

draw $p_j \sim \text{beta}\left(\left(\theta_j + m_{j\cdot}\sigma_j\right)\beta_0, \left(\theta_j + m_{j\cdot}\sigma_j\right)\left(1-\beta_0\right) + n_{j\cdot\cdot} - \sigma_j m_{j\cdot}\right)$;

end

Compute $j^* = \text{argmax}\{p_j : j \in \{1, \ldots J\}\}$;

Sample the next observation from population j^*;

Update table counts and estimates of the HPY hyperparameters;

end

Note that in Algorithm 1 the parameters of the beta distributions depend on the counts in the Chinese Franchise Representation of the HPY process. In particular, they depend on the number of observations for each population $(n_{j\cdot\cdot} : j \in \{1, \ldots, J\})$, the number of clusters in each population $(m_{j\cdot} : j \in \{1, \ldots, J\})$ and the total number of distinct species observed in the joint sample, K. We must remark that the collection clusters counts, $(m_{j\cdot} : j \in \{1, \ldots, J\})$, are latent variables, namely they are not directly observed. Hence, if an initial sample is available, we must estimate these components before running the algorithm. In the on-line supplementary material, we describe a MCMC procedure to handle this problem, together with the problem of inferring the hyperparameters $((\sigma_j, \theta_j) : j \in \{1, \ldots, J\})$ and (α, γ), in case they are treated as

unknown components.

3.2 Incidence data

There are applications where we cannot sample an animal at a time. Instead, multiple individuals are jointly collected in the sample. In these situations an extension of Algorithm 1 is needed. Suppose that, as in the case of abundance data, at every time step we can choose the next population to sample from, but now, instead of one animal, a collection z animals are observed from that population. In such a context, what must be maximized is the expected number of new distinct values observed in an additional sample of size z. In particular, given the array of data $\mathbf{Y_n}$, let us denote by $K_j^{(z)}|\mathbf{Y_n}$ the random number of new distinct species observed in a new sample of size z, collected from population j. By new distinct species, we mean species that are observed in the additional sample, but which were not previously observed in any of the J populations. In such a context, the reward distribution for arm j is the distribution of the random variable $\mathbb{E}[K_j^{(z)}|\mathbf{Y_n}]$. Note that $\mathbb{E}[K_j^{(z)}|\mathbf{Y_n}]$ is a random variable (since P_j is random), but if conditioned to $P_j|\mathbf{Y_n}$, it becomes a number. Another remark is that, when $z = 1$, we are back to the abundance case. In fact, $\mathbb{E}[K_j^{(1)}|\mathbf{Y_n}] = \mathbb{E}[I\left(Y_{n_{j\cdot\cdot}+1} \in A\right)|\mathbf{Y_n}] = P_j\left(A\right)|\mathbf{Y_n}$, where I is the indicator function and $A = \{y \in \mathcal{Y} : y \notin \mathbf{Y_n}\}$. In Proposition 2, we derive the distribution of $(\mathbb{E}[K_1^{(z)}|\mathbf{Y_n}], \ldots, \mathbb{E}[K_J^{(z)}|\mathbf{Y_n}])$. Its proof is available in the on-line supplementary material.

Proposition 2. *Conditionally to* $\beta_0|\mathbf{Y_n} \sim beta\left(\beta_0|\gamma + K\alpha, m_{\cdot\cdot} - \alpha K\right)$ *and conditionally to* $P_j\left(A\right)|\mathbf{Y_n}, \beta_0 = p_j$*, where*

$$P_j\left(A\right)|\mathbf{Y_n}, \beta_0 \sim beta\left(p_j|\left(\theta_j + m_{j\cdot}\sigma_j\right)\beta_0, \left(\theta_j + m_{j\cdot}\sigma_j\right)\left(1-\beta_0\right) + n_{j\cdot\cdot} - \sigma_j m_{j\cdot}\right),$$

$\mathbb{E}[K_j^{(l)}|\mathbf{Y_n}]$ *is a constant, independent of the other arms and* $\mathbb{E}[K_j^{(l)}|\mathbf{Y_n}, \beta_0, p_j]$ *can be computed as*

$$\sum_{k=0}^{l} k \sum_{i=k}^{l} \binom{l}{i} p_j^i \left(1-p_j\right)^{l-i} \sum_{\tilde{m}=k}^{i} F\left(\tilde{m}, k, \alpha, \gamma + K\alpha\right) F\left(i, \tilde{m}, \sigma, \left(\theta + m_{j\cdot}\right)\beta_0\right),$$

where the function $F(n,k,\sigma,\theta)$ (recalled in the supplementary material) is the probability of having k distinct values in a sample of size n sampled from a Pitman-Yor process with hyperparameters (σ,θ).

HPY-TS for incidence data, prescribes to sample a draw from the joint distribution of $(\mathbb{E}[K_1^{(z)}|\mathbf{Y_n}],\dots,\mathbb{E}[K_J^{(z)}|\mathbf{Y_n}])$ and to choose the population corresponding to the highest value. A schematic description of the algorithm for incidence data is summarized in Algorithm 2.

Algorithm 2: HPY-TS - Incidence Data

for *i in 1: number of new samples* **do**

fix z equal to its posterior estimate ;

draw $\beta_0 \sim \text{beta}\left(\gamma + K\alpha, m_{\cdot\cdot} - \alpha K\right)$;

for *j in 1:J* **do**

draw $p_j \sim \text{beta}\left((\theta_j + m_{j\cdot}\sigma_j)\,\beta_0, (\theta_j + m_{j\cdot}\sigma_j)\,(1-\beta_0) + n_{j\cdot\cdot} - \sigma_j m_{j\cdot}\right)$;

compute $\mathbb{E}[K_j^{(z)}|\mathbf{Y_n},\beta_0,p_j]$ as in Proposition 2 ;

end

Compute $j^* = \text{argmax}\{\mathbb{E}[K_j^{(z)}|\mathbf{Y_n},\beta_0,p_j] : j \in \{1,\dots,J\}\}$;

Sample the next group of observations from population j^*;

Update table counts and estimates of the HPY hyperparameters;

end

A remark is that, up to now, we considered the additional sample size z as fixed. However, in some applications it could not be the case. For instance, if we capture animals using traps, we could not know in advance how many animals will be captured in the next step. In these circumstances, the approach can be extended to handle this case by incorporating the distribution of z into to posterior sampler. For example, we can assume independence of the rest of the model, adopt a simple parametric model for z and, at every time step, use its posterior point estimate to compute $\mathbb{E}[K_j^{(z)}|\mathbf{Y_n}]$.

4. APPLICATIONS

4.1 Simulated results

In the following simulations, the true distribution of each arm is supported on a subset of size 2500, randomly chosen from a total number of 3000 possible species, hence allowing for a partial sharing of the supports. These J distributions are assumed to follow Zipf laws. The mass assigned to the k-th most common species in population j, is

$$p_j\left(k; s_j\right) = \frac{1/k^{s_j}}{\sum_{n=1}^{2500}\left(1/n^{s_j}\right)}$$

where $s_j > 1$ is a real parameter controlling how the total mass is spread along the support points. When s_j is high, the total mass is concentrated on a few points and the ordered masses steeply decrease toward zero. As s_j approaches 1, the total mass is more spread, with many points of high mass.

In the bandit context, an arm with low parameter s_j can be viewed as a "winning arm", an arm with high species variety. Whereas, a high value for s_j implies that, after the few very common species have been discovered, the discovery probability for that arm will be very close to zero.

The three competing strategies, used as a term of comparison, are the following:

- an *Oracle strategy*: this strategy knows the $(P_1, \ldots, P_J)$ that generates the data. Hence, uncertainty on the underlying data generating process is removed and, at every time step, this strategy selects the arm with the highest missing mass, so maximizing the probability of observing a new value in the next observation.

- a *Uniform strategy*: this strategy, at every time step, picks an arm uniformly at random, i.e. every arm has probability $1/J$ of being played next. Another similar strategy can be a deterministic strategy that cycles through the experts, i.e., at time t, it draws from population $(t \mod [J])$.

- a UCB algorithm, recently proposed by Bubeck et al. (2013b), based on the Good and Turing missing mass estimator, derived in Good (1953). We refer to this algorithm as *Good-Turing strategy* for simplicity. It is an UCB-like algorithm introduced to solve the issue of security analysis of a power system. This algorithm uses an adaptation of Good and Turing missing mass estimator of Good (1953) to produce a point estimate of the probability of observing a new item, in each arm. Then, it constructs a deterministic upper bound for this estimate, inversely proportional to the number of times that that arm has been played. The chosen arm is the one with the highest upper bound. More precisely, the adapted Good and Turing estimator counts the number of items with frequency one in joined sample that has been observed in arm j and divides it by the number of plays of that arm. This ratio is the point estimator of the missing mass in arm j. The upper bound is constructed by summing $C(\log(4n)/n_{j\cdot})^{1/2}$ to the point estimate, where $n = \sum_{j=1}^{J} n_{j\cdot}$ is the total number of plays and C is a tuning parameter to be fixed.

We consider three different scenarios, corresponding to different levels of heterogeneity or homogeneity in species variety across arms. Heterogeneity in species variety depends on how different the parameters of the Zipf laws are across arms. When heterogeneity is high, "winning" and "losing" arms emerge. Winning arms are those with high species variety (with a low Zipf parameter), while the losing ones (those with high Zipf parameters) are those in which the mass will be concentrated on just a few dominating species. In presence of heterogeneity, a good strategy must be able to detect winning arms soon and play them only. Whereas, in presence of homogeneity, there will not be "winning" arms and all arms will have similar probabilities of producing new species. In this case, a strategy must be able not to get stuck exploiting only a few arms, but to carefully explore all of them. In our simulations, we fix $J = 8$ and

consider the following three scenarios:

1. *Pure Exploitation*, Zipf parameters=(1.3,1.3,2,2,2,2,2,2): in this scenario, there are two "winning" arms. A good strategy should be able to intensively exploit these two arms, without exploring much the other six suboptimal arms.

2. *Pure Exploration*, Zipf parameters=(1.3,1.3,1.3,1.3,1.3,1.3,2,2): in this scenario, the majority of arms are equally profitable. A good strategy should not get stuck exploiting just a few of them, but continue to explore all the six good arms.

3. *Exploration plus Exploitation*, Zipf parameters=(1.3,1.3,1.3,1.3,2,2,2,2): in this scenario, there are four good arms and four bad ones. A good strategy should adequately balance exploitation and exploration, by stopping to play the four suboptimal arms soon, but continuing to play all the other four.

Figure 1, 2 and 3 report the results of simulations in the three scenarios just described for both abundance and incidence data. Each figure displays the average number of species discovered by the four algorithms as a function of the additional samples observed. In particular, results are averages of 60 runs. For each run, we assume an initial sample of 30 observations per arm to be available and collect further 300 observations, following the four possible strategies. In the abundance case, new observations arrive one at a time. In the incidence one, they arrive as 30 bunches of size 10. The hyperparameters of the HPY are endowed with priors, $\alpha, \sigma_1, \ldots, \sigma_J \overset{iid}{\sim} \text{beta}\,(1,2)$ and $\gamma, \theta_1, \ldots, \theta_J \overset{iid}{\sim} \exp(1)$, and then estimated using the MCMC described in the on-line supplementary material. In this on-line section, we also provide Tables containing the weights given to each arm by the four algorithms in the first 10 simulations of each scenario, i.e. the number of times each arm has been chosen by the four algorithms in each simulation.

(Figure 1, Figure 2 and Figure 3 about here)

In the simulations, the HPY-TS algorithm performs well in all scenarios, discovering fewer new species than the Oracle strategy, but more than the Uniform and the Good-Turing strategy. Figure 1, 2 and 3 show how these latter strategies seem to balance the exploration-exploration trade-off worse than HPY-TS. They perform relatively well only in the two extreme cases of pure exploration or pure exploitation, Figure 1 and 2. This guess is strongly confirmed by looking at the Tables in the on-line supplementary material providing the weights of each arm. On the one hand, the Good-Turing strategy does too much exploitation. It selects the arm that seems the most profitable at initial time point and exploits it only, without exploring the others. This behaviour is evident by looking at the Tables with the weights. The algorithm performs well only in the pure exploitation scenario, Figure 1, in which exploiting just one arm is a profitable strategy. However, this strategy becomes suboptimal in presence of more "winning" arms, as displayed in Figure 2 and 3. On the other hand, as expected, the Uniform strategy does too much exploration. It continues to play all arms, irrespectively of their past behaviors. Its performances are very poor, except in the extreme scenario of pure exploration, Figure 2. Instead, the HPY-TS algorithm seems to be robust to changes in species variety across arms. In all scenarios, it performs well, standing behind only to the Oracle strategy. In particular, in the intermediate scenario, Figure 3, its results are very close to the Oracle's ones, while in the extreme cases, Figure 1 and 2, it is still as good as or better than both the Uniform and the Good-Turing strategies.

4.2 Illustration using species of trees in South America

In this section, we compare the HPY-TS algorithm with the three competing alternatives previously described, using a dataset of species of trees, collected in South America. This dataset was studied in Pyke et al. (2001) and Condit et al. (2002) and contains species of trees observed in 100 plots near the Panama Canal, in Ecuador's

Yasunì National Park and in Perù's Manu Biosphere Reserve. All plots were in terra firme forests and, in each of these plots, trees of $\geq$ 10-cm stem diameter were tagged, measured and sorted to morphospecies. A total of 41688 trees have been observed, displaying totally 802 distinct species. The dataset is freely downloadable in the supplementary on-line section of Condit et al. (2002).

In Condit et al. (2002), this dataset has been used to study β-diversity in tropical forest trees. This is a measure of how species composition changes with distance. This notion was firstly introduced by Whittaker (1960) together with the terms α and γ-diversities to describe species variety in a landscape. In particular, the total species variety (γ-diversity) can be viewed as the product of the mean species diversity in the habitat level (α-diversity) times the differentiation among habitats (β-diversity). α-diversity has been studied in many locations, in particular a high α-diversity has been amply documented for tropical forests under consideration. Using this dataset, Condit et al. (2002) study their β-diversity by comparing the actual data with those predicted by a neutral model in which habitat is uniform and only dispersal and speciation influence species turnover. The result of their study is that the data is inconsistent with the neutral model and a high level of β-diversity is observed. In particular, they show that β-diversity is higher in Panama than in western Amazonia.

We use the same dataset of counts of species to test the performances of the two HPY-TS algorithms against the three alternative strategies. In order to test HPY-TS algorithm, we aggregated the 100 individual plots into 4 bigger groups, according to spatial location. In particular, we joined columns in the dataset with code starting with BCI, P, S and C. In Table 1, we computed the Sorensen similarity index. This similarity index measures the fraction of species shared in two plots and is computed as $2A/(2A+B+C)$, where A is the number of species shared between plots and B and C are the number of species unique to each plot. As a measure of similarity, the Sorensen

index has the feature that it weights all species equally. Also, we computed in Table 2, the Shannon and Simpson indexes for the four aggregated plots. These two indexes measure species variety in a location and, given a (finite or infinite) discrete probability vector $\mathbf{p} = (p_1, p_2, \ldots)$, are computed as $S_{Shan}(\mathbf{p}) = -\sum_j p_j log(p_j)$ and $S_{Simp}(\mathbf{p}) = 1 - \sum_j p_j^2$. Both indexes suggest a high species variety in the four plots. Given the four aggregated plots, we used their empirical distributions as data generating process. As for simulated results, we consider a small sample of 30 observations for each arm as initial sample and then we let the four competing algorithms choose where to collect further observations. These 300 additional observations are sampled one at a time in the abundance data case and as 30 bunches of size 10 when dealing with incidence data. The average results over 60 runs are displayed in Figure 4. Tables 3 and 4 report the weights given to each region by the four algorithms in the first 10 runs.

(Table 1 and Table 2 about here)

(Figure 4 about here)

Figure 4 shows how the Oracle strategy outperforms the three other algorithms. The HPY-TS algorithm performs slightly better than the Good-Turing and better than the Uniform strategy, discovering on average more new species both in the abundance and in the incidence case. As shown in Condit et al. (2002), the level of species variety in the four regions is very different, with the plots of Panama having a species variety much higher compared to the other three locations. Therefore this example is a quite extreme scenario and it is similar to scenario 1 of the simulated results, what we called pure exploitation scenario, with the arm P being the "winning" one. As in the simulated case in this scenario results of the HPY-TS algorithms and Good-Turing are not too different, with the HPY-TS discovering on average just a few species

more than the Good-Turing strategy. However, from the tables with the weights we note that the behaviours of the two algorithms are completely different. On the one side, the HPY-TS does exploration until it starts sampling only from region P. On the other side, the Good-Turing picks one arm at the beginning and exploits it only. Hence, in some simulations the Good-Turing algorithm selects the suboptimal arm BCI and remains stuck exploiting it, without doing any exploration and without realizing that indeed the best arm is P. In presence of more "winning" arms, this greedy behaviour would turn out to be much less profitable than in this example. Moreover, we remark that a big advantage of the HPY-TS algorithm is that it does not require any tuning parameter to regulate the exploration rate, but it balances exploration and exploitation automatically. This feature is very important because the right exploration rate depends on the level of heterogeneity in species variety in the populations, which is usually not known in advance. The HPY-TS seems robust to changes in the level of heterogeneity among populations and it is able to correctly balance the level of exploration and exploitation. Finally, as expected, the Uniform strategy performs worse than all the other strategies. However, the differences in performances among algorithms are now less remarked than in the simulated results, since, having only four arms rather than eight, the probability of picking suboptimal arms is lower in this context.

(Table 3 and Table 4 about here)

CONCLUDING REMARKS

In this work we have introduced a new methodology to choose where to allocate resources when J distinct locations are available. This procedure works sequentially, suggesting at every time step from which location to collect the next sample. This sample can be composed of one observation only or by a group of them and the group sizes

can be either fixed or random. For both cases we have provided HPY-TS algorithms, based on a joint use of tools from the Bayesian nonparametric and the multi-armed bandit literature. In particular a HPY is used to estimate the unknown probability measures $(P_1, \ldots, P_J)$ and TS for the sequential allocation problem. Up to our knowledge, this is the first instance that such tools have been used together, particularly with the aim of discovering items from multiple populations. Results from simulated and real data are good, showing that the HPY-TS algorithms are competitive or better than other strategies already proposed, both when dealing with abundance and with incidence data.

These good empirical results encourage to continue the research in this direction, also in consideration of the wide applicability of the algorithm in a variety of fields, like ecology, biology or genetics. Moreover, the proposed HPY-TS algorithm can be easily adapted to deal with other problems in species sampling. For instance it can be used to solve the problem of detecting rare or elusive species, as studied in the adaptive sampling literature started with Thompson (1990). In this problem we are interested in re-observing a particular rare species, labeled by Y_k^{**}, we already observed in an initial sample. We can modify Algorithm 1 by sampling from the distribution of $(P_1(\{Y_k^{**}\}), \ldots, P_J(\{Y_k^{**}\}))|\mathbf{Y_n}$. This joint posterior distribution can be derived in the same manner as in the proof of Proposition 1 and is still a mixture of a product of j beta distributions, with mixing measure another beta distribution, but with different parameters, depending also on table counts relative to the label Y_k^{**}.

Another possible adaptation of the algorithm is to the case in which we want to maximize the sum of the distinct values observed in each location. In this problem, the target function is affected when we observe a species that we have not observed in that location before, irrespectively of the fact that that species has already been observed it in other location. Proposition 1 and Algorithm 1 can be easily adapted

to this problem. Denoting by $A_j = \left\{y \in \mathcal{Y} : y \notin \mathbf{Y}_{n_{j\cdot\cdot}}\right\}$ the set of species not observed in population j and following the same steps of the proof of Proposition 1, we derive the posterior density of $\left(P_1\left(A_1\right), \ldots, P_J\left(A_j\right)\right)|\mathbf{Y_n}$ as a mixture of the product of J beta densities with parameters of the j-th factor being $(\beta_0(\theta_j + m_{j\cdot}\sigma_j) + \sum_{k:Y_k^{**}\notin\mathbf{Y}_{n_{j\cdot\cdot}}}((\theta_j + m_{j\cdot}\sigma_j)\beta_k + n_{j\cdot k} - \sigma_j m_{jk}), \sum_{k:Y_k^{**}\in\mathbf{Y}_{n_{j\cdot\cdot}}}((\theta_j + m_{j\cdot}\sigma_j)\beta_k + n_{j\cdot k} - \sigma_j m_{jk}))$, and with mixing measure for $(\beta_0, \ldots, \beta_K)$ being a Dirichlet distribution of parameters $(\gamma + K\alpha, m_{\cdot 1} - \alpha, \ldots, m_{\cdot K} - \alpha)$. A HPY-TS algorithm can easily be implemented by substituting the joint posterior of Proposition 1 with this posterior density in Algorithm 1.

Another possible direction of research is to improve the proposed models by including spatial and covariate dependence. Indeed, the HPY model assumes exchangeability of the J distributions $(P_1, \ldots, P_J)$. In applied settings, this assumption may not be adequate, because the marginal distribution of each P_j could be different from place to place, depending on spatial or environmental factors. A possible extension of the proposed strategy can be to use dependent spatial models. In particular, in Bayesian nonparametric literature, a set of spatial models have been proposed as particular cases of the general Dependent Dirichlet Processes (DDPs) of MacEachern (1999). The DDPs are extensions of the Dirichlet Process to account for spatial or temporal dependence, but the same kind of generalization can be applied to any nonparametric prior admitting a stick breaking representation. The dependence on time or location is introduced by indexing the weights, the locations or both by a temporal or spatial variable. The most popular bayesian nonparametric spatial models are the spatial dependent Dirichlet process in Gelfand et al. (2005), and its generalizations in Duan et al. (2007) and Gelfand et al. (2007), the hybrid Dirichlet mixture model of Petrone et al. (2009), the order-based dependent Dirichlet drocess of Griffin and Steel (2006), and the spatial kernel stick-breaking prior of Reich and Fuentes (2007). For a concise

description of all of these models, the reader is referred to Steel and Fuentes (2010), section 11.2.

What is surely missing in this work is a theoretical analysis of the algorithm, which assures, not just empirically, its good properties. We think the next step is to analytically study the behavior of the HPY-TS algorithm, trying to provide a finite time bound for its regret. However, our context seems to be more challenging than that of the classical multi-armed bandit problem, due to the dependence of rewards both across time and populations. We think it could be helpful to substitute TS with an UCB strategy, in which the upper bound for the missing mass of each arm is constructed as a credible interval around the point estimate of Corollary 1, using the joint posterior distribution derived in Proposition 1. A concentration inequality providing an upper bound for the probability that the true missing mass of that arm is outside the credible interval, would then be very helpful to prove a finite time regret bound for the new HPY-UCB strategy.

ACKNOWLEDGEMENTS

The authors thank the Associate Editor and two anonymous referees, whose comments and suggestions helped to improve the paper substantially. Special thanks to Sergio Bacallado, Levi Boyles and Lorenzo Trippa for useful discussions, and to Alberto Galvan and Boyu Ren for support in the computational parts of this work. Stefano Favaro is supported by the European Research Council (ERC) through StG N-BNP 306406. Yee Whye Teh research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617071.

References

Agrawal, S. and Goyal, N. (2012). Analysis of Thompson sampling for the multi-armed bandit problem. In *Conference On Learning Theory*.

Auer, P., Cesa-Bianchi, N. and Fischer, P. (2002). Finite-time analysis of the multi-armed bandit problem. *Machine Learning*, **47**, 235–256.

Barger, K. and Bunge, J. (2010). Objective Bayesian estimation of the number of species. *Bayesian Analysis*, **5**, 619–639.

Basu, D.(1969). Role of the Sufficiency and Likelihood Principles in Sample Survey Theory. *Sankhya, Series A*, **31**, 441–454.

Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, **5**.

Bubeck, S., Ernst, D. and Garivier, A. (2013). Optimal discovery with probabilistic expert advice: finite time analysis and macroscopic optimality. *Journal of Machine Learning Research*, **14**, 601–623.

Chapelle, O. and Li, L. (2011). An empirical evaluation of Thompson sampling. In *Neural Information Processing Systems, 2011*.

Chao, A. (1981). On estimating the probability of discovering a new species. *The Annals of Statistics*, **9**, 1339–1342.

Chao, A. and Lee, S.M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210–217.

Cochran, W.G. (1977). *Sampling Techniques*. Wiley, New York.

Condit, R. et al. (2002). Beta-Diversity in Tropical Forest Trees. *Science*, **295**, 666–669.

Duan, J.A., Guindani, M. and Gelfand, A.E. (2007). Generalized spatial Dirichlet process models. *Biometrika*, **77**, 1–11.

Efron, B. and Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika*, **63**, 435–447.

Favaro, S., Lijoi, A. and Prünster, I. (2012a). A new estimator of the discovery probability. *Biometrics*, **68**, 1188–1196.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209–230.

Fonteneau-Belmudes, F., Ernst, D., Druet, P., Panciatici, P. and Wehenkel, L. (2010). Consequence driven decomposition of large-scale power system security analysis. In *Proceedings of the 2010 IREP Symposium*

Gelfand, A.E., Kottas, A. and MacEachern S.N. (2005). Bayesian nonparametric spatial modelling with Dirichlet processes mixing. *Journal of the American Statistical Association*, **100**, 1021–1035.

Gelfand, A.E., Guindani, M. and Petrone, S. (2007). Bayesian nonparametric modeling for spatial data analysis using Dirichlet processes. In *Bayesian Statistics* **8**, (eds. J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West). Oxford University Press, Oxford.

Goldwater, S., Griffiths, T., and Johnson, M. (2006). Interpolating between types and tokens by estimating power-law generators. In *Advances in Neural Information Processing Systems*, **18**. MIT Press, Cambridge, MA.

Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237–264.

Good, I.J. and Toulmin, G.H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika*, **43**, 45–63.

Granmo, O.C. (2010). Solving two-armed bernoulli bandit problems using a Bayesian learning automaton. *International Journal of Intelligent Computing and Cybernetics*, **3**, 207–234.

Griffin, J.E. and Steel, M.F.J. (2006). Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, **101**, 179–194.

Kaufmann, E., Korda, N. and Munos, R. (2012). Thompson Sampling: An Optimal Finite-Time Analysis. *The International Conference on Algorithmic Learning Theory*.

Ionita-Laza, I., Lange, C. and Laird, N.M. (2009). Estimating the number of unseen variants in the human genome. *Proceedings of the National Academy of Sciences*, **106**, 5008–5013.

Ionita-Laza, I. and Laird, N.M. (2010). On the optimal design of genetic variant discovery studies. *Statistical Applications in Genetics and Molecular Biology*, **9**, 33

Lai, T.L. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, **6**, 4–22.

Lijoi, A., Mena, R.H. and Prünster, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika*, **94**, 715-740.

MacEachern, S.N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association, Alexandria.

May, B.C. and Leslie, D.S. (2011). Simulation studies in optimistic Bayesian sampling in contextual-bandit problems. Technical Report 11:02, Statistics Group, Department of Mathematics, University of Bristol.

Mao, C.X. and Lindsay, B.G. (2002). A Poisson model for the coverage problem with a genomic application. *Biometrika*, **89**, 669–681.

Mao, C. X. (2004). Predicting the conditional probability of discovering a new class. *Journal of the American Statistical Association*, **99**, 1108–1118.

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, **1**, 226–251.

Petrone, S., Guindani, M. and Gelfand, A.E. (2009). Hybrid Dirichlet mixture models for functional data. *Journal of the Royal Statistical Society, series B*, **71**, 755–782.

Pyke, C.R., Condit, R., Aguilar, S. an Lao, S. (2001). Floristic composition across a climatic gradient in a neotropical lowland forest. *Journal of Vegetation Science*, **12**, 553-566.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability andGame Theory* (eds. T. S. Ferguson, L. S. Shapley and J. B. MacQueen), pp. 245–267. Institute of Mathematical Statistics, Hayward.

Pitman, J. and Yor, M. (1997). The two parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability*, **25**, 855–900.

Reich, B. and Fuentes, M. (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics*, **1**, 249–264.

Russo, D. and Van Roy, B. (2014). Learning to Optimize Via Posterior Sampling. *Mathematics of Operations Research*, **39**, 1221–1243

Russo, D. and Van Roy, B. (2016). An Information-Theoretic Analysis of Thompson Sampling. *Journal of Machine Learning Research*, **17**, 1–30

Scott, S. (2010). A modern Bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, **26**, 639–658

Steel, M.F.J. and Fuentes, M. (2010). Non-Gaussian and Nonparametric Models for Continuous Spatial Data. In *Handbook of Spatial Statistics* (eds. A.E. Gelfand, P.J. Diggle, M. Fuentes and P.Guttorp), 149–167. CRC Press.

Szabo, B. and Tran-Thanh, L. (2015). Finite-Time Concentration Inequalities for the Posteriori and Regret Analysis of Bayesian Online Learning Algorithms. Working Paper.

Teh, Y.W. and Jordan, M. (2010). Hierarchical Bayesian Nonparametric Models with Applications. In *Bayesian Nonparametrics* (eds. N.L. Hjort, C. Holmes, P. Müller and S.G. Walker), pp. 158-207. Cambridge University Press, Cambridge.

Teh, Y. W. (2006). A hierarchical Bayesian language model based on Pitman–Yor processes. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 985–92. Association for Computational Linguistics, Morristown.

Thompson, S.K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, **85**, 1050–1059.

Thompson, S.K. (1991). Stratified Adaptive Cluster Sampling. *Biometrika*, **78**, 389–397.

Thompson, S.K. (1991). Adaptive Cluster Sampling: Designs with Primary and Secondary Units. *Biometrics*, **47**, 1103–1115.

Thompson, S.K. (1992). An Adaptive Procedure for Sampling Animal Populations. *Biometrics*, **48**, 1195–1199.

Thompson, S.K. (2002). *Sampling*. Wiley, New York.

Thompson, S.K. and Seber, G.A.F. (1996). *Adaptive Sampling*. Wliey, New York.

Thompson, W.R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, **25**, 285–294.

Whittaker, R.H. (1960). Vegetation of the Siskiyou Mountains. *Ecological Monographs*, **30**, 279–338.

Zacks, S. (1969). Bayes Sequential Designs of Fixed Size Samples From Finite Populations. *Journal of the American Statistical Association*, **64**, 1342–1349.

Table 1: Sorensen Index

Aggr.Plots	BCI	P	S	C
BCI	1	0.97	0.358	0.44
P	0.97	1	0.196	0.267
S	0.358	0.196	1	0.134
C	0.44	0.267	0.134	1

Table 2: Shannon and Simpson Indeces

Aggr.Plots	Shannon	Simpson
BCI	4.27	0.974
P	5.25	0.988
S	3.412	0.936
C	3.953	0.97

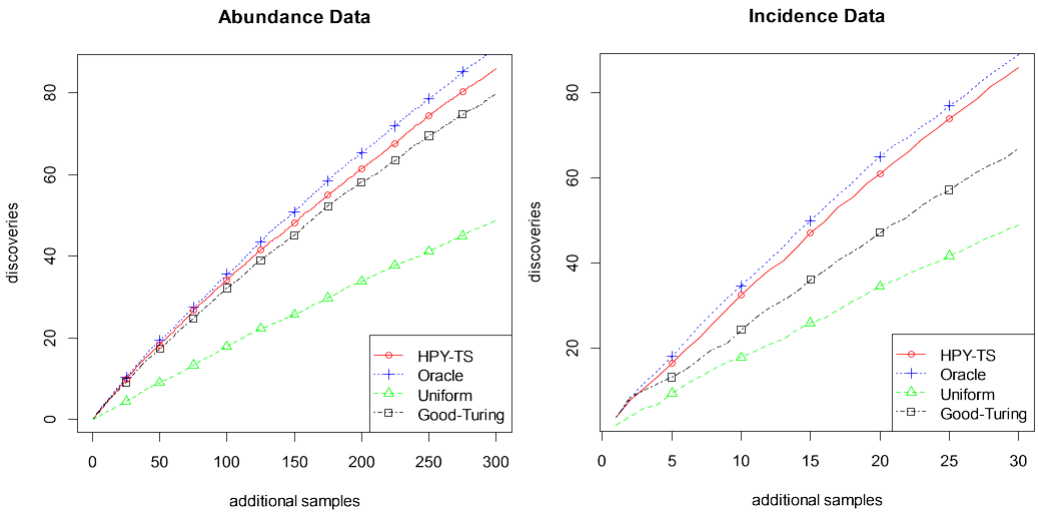

Figure 1: Simulated results. Pure Exploitation Scenario. Zipf parameters=(1.3,1.3,2,2,2,2,2,2).

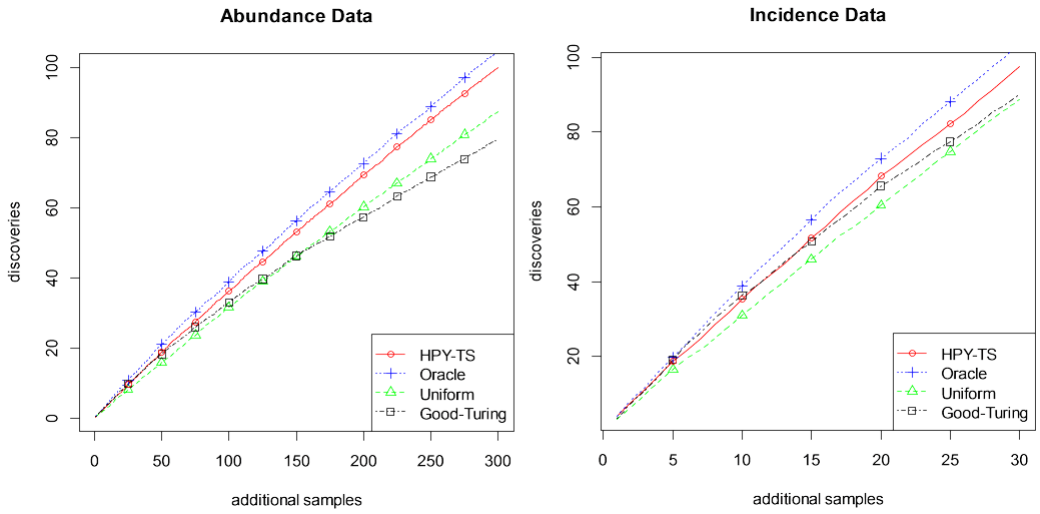

Figure 2: Simulated results. Pure Exploration Scenario. Zipf parameters=(1.3,1.3,1.3,1.3,1.3,1.3,2,2).

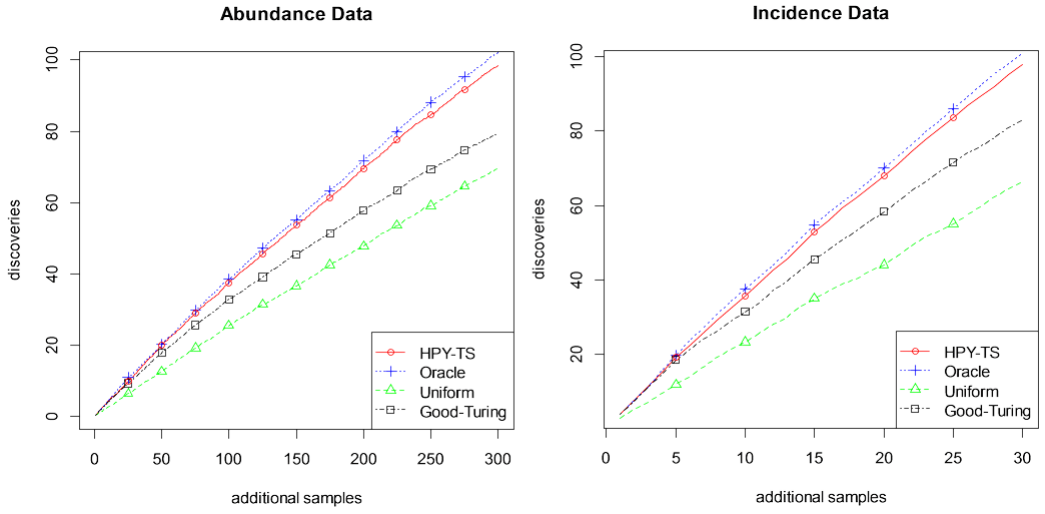

Figure 3: Simulated results. Exploitation plus Exploration Scenario. Zipf parameters=(1.3,1.3,1.3,1.3,2,2,2,2).

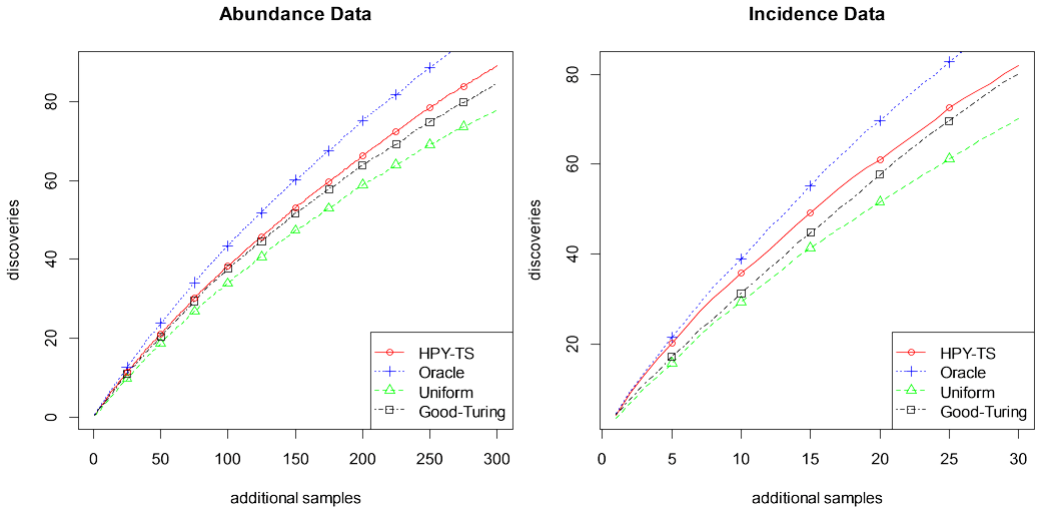

Figure 4: Real data example. Species of trees in South America

Table 3: Real Data Example. Abundance data.

	HPY-TS				Good-Turing			
Runs/Aggr.Plots	BCI	P	S	C	BCI	P	S	C
1	58	129	41	72	294	3	2	1
2	45	182	13	60	0	299	0	1
3	41	211	23	25	20	273	3	4
4	67	148	20	65	3	292	1	4
5	76	104	51	69	1	296	2	1
6	47	170	14	69	2	296	1	1
7	39	210	8	43	1	297	1	1
8	44	202	4	50	3	289	3	5
9	74	148	1	77	300	0	0	0
10	72	131	26	71	291	5	2	2
	Uniform				Oracle			
Runs/Aggr.Plots	BCI	P	S	C	BCI	P	S	C
1	75	73	64	88	0	300	0	0
2	72	85	69	74	0	300	0	0
3	93	65	69	73	0	284	0	16
4	86	73	68	73	0	300	0	0
5	73	79	73	75	0	298	0	2
6	78	92	70	60	5	285	0	10
7	81	60	83	76	0	278	0	22
8	84	64	88	64	0	300	0	0
9	60	91	58	91	0	284	0	16
10	77	81	54	88	0	293	0	7

Table 4: Real Data Example. Incidence data.

	HPY-TS				Good-Turing			
Runs/Aggr.Plots	BCI	P	S	C	BCI	P	S	C
1	8	13	2	7	5	16	4	5
2	7	18	1	4	3	23	2	2
3	7	17	0	6	5	18	3	4
4	7	23	0	0	4	20	3	3
5	4	20	0	6	11	9	4	6
6	5	17	0	8	7	16	4	3
7	7	17	0	6	5	14	4	7
8	8	16	0	6	7	11	5	7
9	6	17	2	5	23	3	2	2
10	5	9	3	13	1	27	1	1
	Uniform				Oracle			
Runs/Aggr.Plots	BCI	P	S	C	BCI	P	S	C
1	5	12	9	4	0	28	0	2
2	9	8	7	6	0	30	0	0
3	7	7	9	7	1	29	0	0
4	7	5	9	9	0	30	0	0
5	8	8	10	4	0	30	0	0
6	8	10	5	7	0	30	0	0
7	5	9	7	9	0	30	0	0
8	3	13	6	8	1	29	0	0
9	7	6	10	7	0	30	0	0
10	9	8	7	6	0	30	0	0

Supplementary material to the paper "Multi-armed bandit for species discovery: a Bayesian nonparametric approach"

Marco Battiston, Stefano Favaro and Yee Whye Teh

1. PROOFS.

In the following proofs, we assume all random variables to be defined on a common probability space, and we denote by $\mathbb{P}$ its probability measure and by $\mathbb{E}$ the corresponding expectation operator. Furthermore, before proving Proposition 1 and Proposition 2, we recall the following result by Gnedin and Pitman (2006) that will be used in our proofs.

Proposition *Let $P \sim PY(\sigma, \theta, H)$ and let $(Y_1, \dots, Y_n)$ be a sample from it. Then, the probability of observing k distinct values in $(Y_1, \dots, Y_n)$ is denoted by $F(n, k, \sigma, \theta)$ and*

$$F(n, k, \sigma, \theta) = \frac{\prod_{r=1}^{k-1}(\theta + r\sigma)}{\sigma^k (\theta + 1)_{n-1}} \mathcal{C}(n, k, \sigma)$$

where $\mathcal{C}$ is the generalized factorial coefficient, defined for all $n \in \mathbb{N}, k \leq n, 0 \leq \sigma \leq 1$ as $\mathcal{C}(n, k; \sigma) = (1/k!) \cdot \sum_{0 \leq j \leq k} (-1)^j \binom{k}{j} (-j\sigma)_n$, with the proviso $\mathcal{C}(0, 0; \sigma) = 1$ and $\mathcal{C}(n, 0; \sigma) = 0$ $\forall n$ and where $(\theta + 1)_{n-1} = (\theta + 1)(\theta + 2) \cdots (\theta + n - 1)$ is the rising factorial coefficient.

We also recall the characterization of the posterior distribution of the PY process derived in Pitman (1996), and the Chinese Restaurant Representation of the HPY.

Given a sample $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$, such that $Y_i|P \overset{iid}{\sim} P$ for all $1 \leq i \leq n$ and $P \sim \text{PY}(\sigma, \theta, H)$, the posterior of P given $\mathbf{Y}_n$ satisfies the following distributional equation

$$P|\mathbf{Y}_n \overset{d}{=} \sum_{i=1}^{K_n} w_i \delta_{Y_i^*} + w_0 \tilde{P} \tag{1}$$

where K_n is the number of distinct values in the sample $\mathbf{Y}_n$, denoted by $(Y_1^*, \ldots, Y_{K_n}^*)$ and having multiplicities $(n_1, \ldots, n_{K_n})$, $(w_0, w_1, \ldots, w_{K_n})$ is a random vector distributed according to $\text{Dir}(\theta + K_n\sigma, n_1 - \sigma, \ldots, n_{K_n} - \sigma)$ and $\tilde{P} \sim \text{PY}(\sigma, \theta + K_n\sigma, H)$.

The Chinese Restaurant Franchise representation of the HPY process is described by the following two predictive distributions for the observables and for the cluster values in population j

$$Y_{j,i+1}|Y_{j,1}, \ldots, Y_{j,i}, \sigma_j, \theta_j, P_0 \sim \sum_{t=1}^{m_{j\cdot}} \frac{n_{jt\cdot} - \sigma_j}{\theta_j + n_{j\cdot\cdot}} \delta_{Y_{j,t}^*} + \frac{\theta_j + m_{j\cdot}\sigma_j}{\theta_j + n_{j\cdot\cdot}} P_0 \tag{2}$$

and

$$Y_{j,m_{j\cdot}+1}^*|Y_{1,1}^*, \ldots, Y_{J,m_{J\cdot}}^*, \alpha, \gamma, H \sim \sum_{k=1}^{K} \frac{m_{\cdot k} - \alpha}{\gamma + m_{\cdot\cdot}} \delta_{Y_k^{**}} + \frac{\gamma + K\alpha}{\gamma + m_{\cdot\cdot}} H. \tag{3}$$

Proof of Proposition 1. From Equation (3), the franchise-wide distinct values $(Y_1^{**}, \ldots, Y_K^{**})$ are governed by P_0 and $P_0 \sim \text{PY}(\alpha, \gamma, H)$. Using formula (1), the posterior distribution of P_0, given the observations, satisfies the distributional equation

$$P_0|\mathbf{Y_n} \overset{d}{=} \sum_{k=1}^{K} \beta_k \delta_{Y_k^{**}} + \beta_0 P_0^{'}$$

where

$$\begin{aligned} P_0^{'}|\mathbf{Y_n} &\sim \text{PY}(\alpha, \gamma + K\alpha, H) \\ \beta|\mathbf{Y_n} = (\beta_0, \ldots, \beta_K)\,|\mathbf{Y_n} &\sim \text{Dir}(\gamma + K\alpha, m_{\cdot 1} - \alpha, \ldots, m_{\cdot K} - \alpha) \end{aligned}$$

Similarly, from formula (2), we can apply formula (1) to P_j to find a distributional equation for P_j, conditionally on P_0 and the data. Also, using the distributional equation for the posterior of P_0, we find the following distributional equation for P_j

$$P_j|\beta, P_0', \mathbf{Y_n} \stackrel{d}{=} \sum_{k=1}^{K} \pi_{j,k}\delta_{Y_k^{**}} + \pi_{j,0}P_j' \tag{4}$$

where

$$\begin{aligned} P_j'|P_0', \mathbf{Y_n} &\sim \text{PY}(\sigma_j, (\theta_j + m_{j\cdot}\sigma_j)\beta_0, P_0') \\ (\pi_{j,0}, \ldots, \pi_{j,K})|\beta, \mathbf{Y_n} &\sim \text{Dir}((\theta_j + m_{j\cdot}\sigma_j)\beta_0, (\theta_j + m_{j\cdot}\sigma_j)\beta_1 + n_{j\cdot 1} - \sigma_j m_{j1}, \ldots \\ &\quad \ldots, (\theta_j + m_{j\cdot}\sigma_j)\beta_K + n_{j\cdot K} - \sigma_j m_{jK}) \end{aligned}$$

So, the distribution of $P_j(A)|\mathbf{Y_n}, P_0$ satisfies

$$P_j(A)|\beta, P_0', \mathbf{Y_n} \stackrel{d}{=} \sum_{k=1}^{K} \pi_{j,k}\delta_{Y_k^{**}}(A) + \pi_{j,0}P_j'(A)$$

for all $j \in \{1, \ldots, J\}$, which implies

$$P_j(A)|\beta_0, \mathbf{Y_n} \sim \text{beta}((\theta_j + m_{j\cdot}\sigma_j)\beta_0, (\theta_j + m_{j\cdot}\sigma_j)(1-\beta_0) + n_{j\cdot\cdot} - \sigma_j m_{j\cdot})$$

where we made use of the following facts:

1. $\delta_{Y_k^{**}}(A) = 0 \ \forall k = 1, \ldots, K$: since $\{Y_1^{**}, \ldots, Y_K^{**}\} = A^c$.

2. $P_j'(A) \stackrel{as}{=} 1$: P_j' can be rewritten as $P_j' = \sum_{i\geq 1}\gamma_i\delta_{X_i}$ for some weights $\{\gamma_i\}_{i\geq 1}$ and atoms $\{X_i\}_{i\geq 1} \stackrel{iid}{\sim} H$. Then, $\mathbb{P}(\cap_{i\geq 1}\{X_i \in A\}) = \prod_{i\geq 1}\mathbb{P}(X_i \in A) = \prod_{i\geq 1} 1 = 1$, since H is diffuse and A^c is a finite set of points. Finally, $\mathbb{P}(\cap_{i\geq 1}\{X_i \in A\}) = 1 \Rightarrow P_j'(A) \stackrel{as}{=} 1$.

3. $\pi_{j,0}|\beta_0, \mathbf{Y_n} \sim \text{beta}((\theta_j + m_{j\cdot}\sigma_j)\beta_0, (\theta_j + m_{j\cdot}\sigma_j)(1-\beta_0) + n_{j\cdot\cdot} - \sigma_j m_{j\cdot})$: by the aggregation property of Dirichlet distribution.

Also, since we are conditioning on P_0 (through β, P_0'), $P_j(A)|\beta_0, \mathbf{Y_n}$ is independent of $P_i(A)|\beta_0, \mathbf{Y_n}$ for all $i, j \in \{1, \ldots J\}$, $i \neq j$. Hence, their joint distribution is simply the product of the marginals. The last step is to integrate β_0 out

$$(P_1(A), \ldots, P_J(A))|\mathbf{Y_n} = \int_0^1 \prod_{j=1}^{J} P_j(A)|\beta_0, \mathbf{Y_n} \cdot dF_{\beta_0}(\beta_0)$$

where the distribution of β_0 is another beta (again by aggregation of Dirichlet distribution). So, $(P_1(A), \ldots, P_J(A))|\mathbf{Y_n}$ admits a density as stated.

Proof of Proposition 2. Using the distributional equation (4) for the posterior of P_j and working conditionally on $\beta_0|\mathbf{Y_n} \sim \text{beta}(\gamma + K\alpha, m_{\cdot\cdot} - \alpha K)$, we compute $\mathbb{P}(K_j^{(z)} = k|\mathbf{Y_n}, \beta_0)$.

From the distributional equation, we know that, given

$$\pi_{j,0}|\beta_0, \mathbf{Y_n} \sim \text{beta}\left((\theta_j + m_{j\cdot}\sigma_j)\beta_0, (\theta_j + m_{j\cdot}\sigma_j)(1-\beta_0) + n_{j\cdot\cdot} - \sigma_j m_{j\cdot}\right)$$

an observation $Y_{n_{j\cdot\cdot}+i}$ with $i = 1, \ldots, z$ does not coincide with any of the K distinct species (in the joint sample) with probability $\pi_{j,0}$. To have $K_j^{(z)} = k$, at least k of the z data $Y_{n_{j\cdot\cdot}+1}, \ldots, Y_{n_{j\cdot\cdot}+z}$ must be allocated to the k new distinct species that have not previously observed. Hence,

$$\mathbb{P}(K_j^{(z)} = k|\mathbf{Y_n}, \beta_0, \pi_{j,0}) = \sum_{i=k}^{z} \binom{z}{i} {\pi_{j,0}}^i (1-\pi_{j,0})^{z-i} \mathbb{P}(K_i = k|\beta_0)$$

where K_i is now the number of distinct species in a sample of size i generated by a $\text{PY}(\sigma_j, (\theta_j + m_{j\cdot}\sigma_j)\beta_0, P_0')$, where $P_0' \sim \text{PY}(\alpha, \gamma + K\alpha, H)$.

We need to find $\mathbb{P}(K_i = k|\beta_0)$. Using the Chinese Franchise Representation and the result by Gnedin and Pitman (2006), denoting by M_i the number of tables, we have that, for $\tilde{m} = 1, \ldots, i$, $\mathbb{P}(M_i = \tilde{m}) = F(i, \tilde{m}, \sigma_j, (\theta_j + m_{j\cdot}\sigma_j)\beta_0)$. Moreover, conditionally on $M_i = \tilde{m}$, for $k = 1, \ldots, \tilde{m}$, $\mathbb{P}(K_i = k|M_i = \tilde{m}) = F(\tilde{m}, k, \alpha, \gamma + K\alpha)$. Finally, $\mathbb{P}(K_j^{(z)} = k|\mathbf{Y_n}, \beta_0, \pi_{j,0})$ can be computed as

$$\sum_{i=k}^{z} \binom{z}{i} {\pi_{j,0}}^i (1-\pi_{j,0})^{z-i} \sum_{\tilde{m}=k}^{i} F(\tilde{m}, k, \alpha, \gamma + K\alpha) F(i, \tilde{m}, \sigma_j, (\theta_j + m_{j\cdot}\sigma_j)\beta_0)$$

The conditional mean $\mathbb{E}(K_j^{(z)}|\mathbf{Y_n}, \beta_0, \pi_{j,0})$ is found by averaging over $\{0, \ldots, z\}$ and, being constant, they are trivially independent among arms. Hence, the joint distribution of $(\mathbb{E}(K_1^{(z)}|\mathbf{Y_n}), \ldots, \mathbb{E}(K_J^{(z)}|\mathbf{Y_n}))$ is found by integrating β_0, $(\pi_{j,0} : j \in \{1, \ldots J\})$ out from the product of these J conditional (constant) distributions.

2. IMPLEMENTATION ISSUES - MCMC ALGORITHM FOR THE HPY PARAMETERS.

The number of clusters in each population $\mathbf{m}_J = (m_{j\cdot} : j \in \{1, \ldots, J\})$ appearing in the parametrization of the beta distributions in both Algorithms 1 and 2 of Section 3.1 and 3.2 of the paper, are latent variables. In subsection 2.1 we describe a simple MCMC scheme to estimate them in case an initial sample is available. The MCMC algorithm directly follows from paragraph 5.1 of Teh et al. (2006). Moreover if the hyperparameters of the HPY model are unknown, they must added to the MCMC sampler too, as outlined in subsection 2.2.

2.1. MCMC for $\mathbf{m}_J$*:* In principle a Gibbs sampler to estimate $\mathbf{m}_J$ should sequentially draw samples from the full conditionals $\pi(m_{j\cdot}|m_{1\cdot}, \ldots, m_{j-1\cdot}, m_{j+1\cdot}, \ldots, m_{J\cdot}, \mathbf{Y_n})$. However both the joint $\pi(m_{1\cdot}, \ldots, m_{J\cdot}|\mathbf{Y_n})$ and the full conditional posterior distributions are difficult combinatorial objects and cannot be derived in closed form. A possible solution is a Gibbs sampler that, rather than directly updating $m_{j\cdot}|m_{-j\cdot}, \mathbf{Y_n}$, updates the cluster allocations $(t_{ji} : i \in \{1, \ldots, n_{j\cdot\cdot}\})$ and then computes $m_{j\cdot}|m_{-j\cdot}, \mathbf{Y_n}$. As in Teh et al. (2006), the cluster allocation variable t_{ji} specifies the cluster to which the i-th observation of population j belongs. Let $\mathbf{t}_{-jp}^{(i-1)}$ denote the array of cluster allocations after iteration $i-1$ of the sampler and with the p-th observation of the j-th population removed. Then $t_{jp}^{(i)}|(\mathbf{Y_n}, \mathbf{t}_{-jp}^{(i-1)})$ is proportional to

$$\sum_{t:\psi_{jt}=\psi_{jt_{ji}}} \frac{n_{jt\cdot} - I\left(t = t_{jp}^{(i-1)}\right) - \sigma_j}{\theta_j + n_{j\cdot\cdot} - 1}\delta_t + \frac{\theta_j + m_{j\cdot}^{(i-1,p-1)}\sigma_j}{\theta_j + n_{j\cdot\cdot} - 1}\frac{m_{\cdot k_{jp}}^{(i-1,p-1)} - \alpha}{\gamma + m_{\cdot\cdot}^{(i-1,p-1)}}\delta\left(m_{j\cdot}^{(i-1,p-1)} + 1\right)$$

where $m_{j\cdot}^{(i-1,p-1)}$ denotes the number of clusters in population j at the i-th iteration after having updated the first $p-1$ cluster allocations of that population, ψ_{jt} is a classification variable that tells us the species of the observations in the t-th cluster in population j and k_{jp} is the species of the observations in the p-th cluster in population

j. If $n_{jt_{jp}\cdot} = 1$ (i.e. the observation is forming its own cluster), before updating $t_{jp}^{(i)}$ we must remove its cluster and subtract one to all the m's. The updated value for $m_{j\cdot}^{(i)}$ can also be taken as the highest $t_{jp}^{(i)}$ for $p \in \{1, \ldots, n_{j\cdot\cdot}\}$, rather than the number of distinct values in the $t_{jp}^{(i)}$.

The algorithm is time expensive because at every iteration it re-samples the cluster allocations of all populations and of all observations. However, we experienced that a good choice of the starting value makes the chain converge to its stationary distribution in just a few iterations. We suggest to run a Chinese Franchise given the data to find the initial point for cluster allocations to start the Gibbs sampler.

When the HPY-TS algorithm is run, the vector $\mathbf{m}_J$ can be updated by allocating new observations to either old or new clusters using the Chinese Restaurant Franchise. If the observation is new, it forms a new cluster. If it is old, say of type Y_k^{**}, then the corresponding observation either will form a new cluster with probability proportional to $((m_{\cdot k} - \alpha)/(\gamma + m_{\cdot\cdot}))(\theta_j + m_{j\cdot}\sigma_j)/((\theta_j + n_{j\cdot\cdot}))$ or it will join an existing cluster (with dish Y_k^{**}) with probability proportional to $(n_{j\cdot k} - m_{jk}\sigma_j)/(\theta_j + n_{j\cdot\cdot})$.

2.2. HPY Hyperparameters: If the hyperparameters are considered as unknown, they must be included in the Gibbs sampler for the cluster sizes. Assuming independent priors for hyperparameters of different Pitman-Yor processes, the full conditional distributions can be derived from

$$\pi(\alpha, \gamma | (m_{jk} : j \in \{1, \ldots, J\}, k \in \{1, \ldots, K\}), (\sigma_j, \theta_j : j \in \{1, \ldots, J\}), \mathbf{Y_n}) =$$
$$= \pi(\alpha, \gamma | m_{\cdot\cdot}, K) \propto \frac{\Gamma(\frac{\gamma}{\alpha} + K)\Gamma(\gamma)\mathcal{C}(m_{\cdot\cdot}, K, \alpha)}{\Gamma(\frac{\gamma}{\alpha})\Gamma(\gamma + m_{\cdot\cdot})} \pi^{prior}(\alpha, \gamma)$$

and, for each couple $((\sigma_j, \theta_j) : j \in \{1, \ldots J\})$, from

$$\pi(\sigma_j, \theta_j | (m_{jk} : j \in \{1, \ldots, J\}, k \in \{1, \ldots, K\}), \sigma_{-j}, \theta_{-j}, \alpha, \gamma, \mathbf{Y_n}) =$$
$$= \pi(\sigma_j, \theta_j | n_{j\cdot\cdot}, m_{j\cdot}) \propto \frac{\Gamma(\frac{\theta_j}{\sigma_j} + m_{j\cdot})\Gamma(\theta_j)\mathcal{C}(n_{j\cdot\cdot}, m_{j\cdot}, \sigma_j)}{\Gamma(\frac{\theta_j}{\sigma_j})\Gamma(\theta_j + n_{j\cdot\cdot})} \pi^{prior}(\sigma_j, \theta_j)$$

3. SIMULATIONS RESULTS - TABLES OF WEIGHTS.

The following tables report the weights given to each arm by the four algorithms in the simulation study. We consider the behaviour of the HPY-TS algorithm and of the three other competing strategies in the 3 scenarios described in Section 4.1 of the paper. Specifically, we consider the following three scenarios:

1. *Pure Exploitation Scenario*, corresponding to the Zipf parameter vector for the true distributions equal to $(1.3, 1.3, 2, 2, 2, 2, 2, 2)$;

2. *Pure Exploration Scenario*, corresponding to the Zipf parameter vector $(1.3, 1.3, 1.3, 1.3, 1.3, 1.3, 2, 2)$;

3. *Exploration-Exploitation Scenario*, corresponding to the Zipf parameter vector $(1.3, 1.3, 1.3, 1.3, 2, 2, 2, 2)$.

For each scenario we run the four algorithms for 60 times. We report here only the results of the first 10 simulations. Each row in the tables is the result of one simulation. The columns correspond to the possible arms. Finally, we repeat the same simulations both for abundance and for incidence data.

References

Gnedin, A. and Pitman, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *Journal of Mathematical Science*, **138**, 5674–5685.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability andGame Theory* (eds. T. S. Ferguson, L. S. Shapley and J. B. MacQueen), 245–267. Institute of Mathematical Statistics, Hayward.

Teh, Y.W., Jordan, M., Beal, M. and Blei, D. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, **101**, 1566–1581.

Table 1: Simulations: Pure Exploitation. Abundance data.

	HPY-TS								Good-Turing							
Runs/Zipf	1.3	1.3	2	2	2	2	2	2	1.3	1.3	2	2	2	2	2	2
1	157	137	4	1	1	0	0	0	3	292	1	1	1	1	1	0
2	147	121	3	5	1	4	13	6	282	11	1	1	1	1	2	1
3	168	125	0	0	0	4	2	1	2	298	0	0	0	0	0	0
4	10	279	2	0	6	0	2	1	5	269	4	4	4	5	4	5
5	216	66	0	0	5	3	0	10	293	1	1	1	1	1	1	1
6	134	150	0	1	0	12	0	3	1	298	0	0	0	1	0	0
7	147	100	1	9	1	35	0	7	291	2	1	1	1	2	1	1
8	137	161	0	2	0	0	0	0	290	5	1	1	0	1	1	1
9	146	149	1	0	0	4	0	0	299	1	0	0	0	0	0	0
10	134	111	8	3	4	6	13	21	288	5	1	1	1	1	1	2
	Uniform								Oracle							
Runs/Zipf	1.3	1.3	2	2	2	2	2	2	1.3	1.3	2	2	2	2	2	2
1	31	48	34	39	40	32	38	38	139	161	0	0	0	0	0	0
2	45	39	33	43	38	26	39	37	178	122	0	0	0	0	0	0
3	40	39	41	33	41	31	34	41	147	153	0	0	0	0	0	0
4	51	35	40	28	32	27	48	39	157	143	0	0	0	0	0	0
5	40	40	41	36	29	42	30	42	112	188	0	0	0	0	0	0
6	26	45	37	42	39	37	36	38	176	124	0	0	0	0	0	0
7	38	38	56	37	31	29	30	41	126	174	0	0	0	0	0	0
8	29	38	38	40	43	31	47	34	197	103	0	0	0	0	0	0
9	39	30	37	37	47	43	35	32	179	121	0	0	0	0	0	0
10	36	31	34	38	31	39	46	45	120	180	0	0	0	0	0	0

Table 2: Simulations: Pure Exploitation. Incidence data.

	HPY-TS								Good-Turing							
Runs/Zipf	1.3	1.3	2	2	2	2	2	2	1.3	1.3	2	2	2	2	2	2
1	14	16	0	0	0	0	0	0	5	9	2	2	3	3	3	3
2	15	14	0	0	0	1	0	0	3	15	1	2	2	2	2	3
3	22	6	0	0	0	0	2	0	7	6	3	3	2	3	3	3
4	9	19	0	2	0	0	0	0	24	1	1	1	1	0	1	1
5	11	19	0	0	0	0	0	0	14	3	2	2	3	2	2	2
6	10	20	0	0	0	0	0	0	14	4	1	2	3	2	2	2
7	16	13	0	0	0	0	0	1	4	13	2	2	2	3	2	2
8	17	12	0	1	0	0	0	0	6	11	3	2	2	2	2	2
9	16	14	0	0	0	0	0	0	18	2	2	2	1	2	1	2
10	13	16	0	0	0	0	1	0	11	5	2	3	2	2	2	3
	Uniform								Oracle							
Runs/Zipf	1.3	1.3	2	2	2	2	2	2	1.3	1.3	2	2	2	2	2	2
1	4	5	6	3	3	0	7	2	15	15	0	0	0	0	0	0
2	4	4	3	3	2	6	4	4	15	15	0	0	0	0	0	0
3	2	4	6	4	3	5	2	4	13	17	0	0	0	0	0	0
4	4	8	3	2	3	2	4	4	20	10	0	0	0	0	0	0
5	5	2	1	3	5	3	6	5	12	18	0	0	0	0	0	0
6	3	10	1	2	4	4	3	3	12	17	1	0	0	0	0	0
7	3	2	2	4	3	4	6	6	15	15	0	0	0	0	0	0
8	7	5	4	0	2	2	4	6	12	18	0	0	0	0	0	0
9	2	2	9	1	2	5	3	6	13	17	0	0	0	0	0	0
10	2	6	4	2	7	5	2	2	16	14	0	0	0	0	0	0

Table 3: Simulations: Pure Exploration. Abundance data.

	HPY-TS								Good-Turing							
Runs/Zipf	1.3	1.3	1.3	1.3	1.3	1.3	2	2	1.3	1.3	1.3	1.3	1.3	1.3	2	2
1	43	47	41	68	30	71	0	0	2	4	1	1	1	290	1	0
2	1	61	37	15	100	86	0	0	0	2	295	1	1	1	0	0
3	11	40	148	57	40	3	0	1	2	287	4	2	2	1	1	1
4	46	114	15	47	64	14	0	0	298	0	0	0	2	0	0	0
5	4	112	1	43	37	103	0	0	2	286	1	3	2	5	0	1
6	56	47	48	21	100	28	0	0	2	2	289	1	5	1	0	0
7	42	53	9	67	98	31	0	0	6	2	1	3	286	2	0	0
8	58	25	31	51	34	101	0	0	285	2	7	3	1	1	1	0
9	36	68	107	37	1	51	0	0	1	3	293	1	1	1	0	0
10	36	35	44	68	41	76	0	0	2	2	1	2	280	12	1	0
	Uniform								Oracle							
Runs/Zipf	1.3	1.3	1.3	1.3	1.3	1.3	2	2	1.3	1.3	1.3	1.3	1.3	1.3	2	2
1	31	43	31	34	38	49	30	44	83	41	30	58	54	34	0	0
2	45	24	29	36	32	40	50	44	73	41	74	77	5	30	0	0
3	43	47	32	36	38	28	35	41	38	31	61	78	48	44	0	0
4	27	44	37	45	41	30	41	35	46	30	60	42	54	68	0	0
5	33	40	49	44	34	30	35	35	52	45	42	49	35	77	0	0
6	39	36	42	40	33	36	33	41	75	57	75	36	10	47	0	0
7	31	36	43	54	30	42	30	34	38	60	108	19	51	24	0	0
8	49	32	31	36	39	42	35	36	37	72	19	86	53	33	0	0
9	33	39	42	36	36	40	34	40	53	52	51	17	69	58	0	0
10	40	33	30	33	40	45	39	40	55	73	47	36	61	28	0	0

Table 4: Simulations: Pure Exploration. Incidence data.

	HPY-TS								Good-Turing							
Runs/Zipf	1.3	1.3	1.3	1.3	1.3	1.3	2	2	1.3	1.3	1.3	1.3	1.3	1.3	2	2
1	0	14	5	6	2	3	0	0	2	2	5	15	2	2	1	1
2	6	8	5	5	6	0	0	0	2	5	15	2	2	2	1	1
3	1	3	4	7	7	8	0	0	3	1	19	1	1	3	1	1
4	7	4	0	5	4	10	0	0	5	2	2	2	8	9	1	1
5	6	8	7	7	2	0	0	0	2	7	6	7	2	3	2	1
6	1	16	2	5	2	4	0	0	1	1	1	25	1	1	0	0
7	10	1	6	1	9	3	0	0	1	1	24	1	1	2	0	0
8	5	12	1	2	2	8	0	0	3	1	1	1	2	21	0	1
9	7	5	10	2	0	6	0	0	4	4	2	1	1	16	1	1
10	3	0	15	0	6	6	0	0	1	1	3	3	18	2	1	1
	Uniform								Oracle							
Runs/Zipf	1.3	1.3	1.3	1.3	1.3	1.3	2	2	1.3	1.3	1.3	1.3	1.3	1.3	2	2
1	4	4	3	5	4	3	3	4	6	5	1	6	7	5	0	0
2	3	4	4	3	6	4	3	3	5	8	4	4	4	5	0	0
3	4	4	3	5	3	5	2	4	6	10	3	7	2	2	0	0
4	7	0	4	3	6	5	2	3	4	2	5	6	9	4	0	0
5	3	5	3	2	3	6	3	5	3	4	4	3	10	6	0	0
6	4	6	3	3	8	2	1	3	5	5	5	5	6	4	0	0
7	3	3	1	3	5	7	6	2	8	4	3	3	4	8	0	0
8	3	7	5	2	4	3	1	5	3	4	3	6	6	8	0	0
9	3	4	2	7	5	4	2	3	3	7	6	4	6	4	0	0
10	6	3	2	3	6	2	6	2	5	5	7	4	4	5	0	0

Table 5: Simulations: Exploration-Exploitation. Abundance data.

	HPY-TS								Good-Turing							
Runs/Zipf	1.3	1.3	1.3	1.3	2	2	2	2	1.3	1.3	1.3	1.3	2	2	2	2
1	10	159	74	57	0	0	0	0	0	0	300	0	0	0	0	0
2	81	76	63	80	0	0	0	0	297	1	1	1	0	0	0	0
3	72	60	85	83	0	0	0	0	1	2	4	292	1	0	0	0
4	72	30	129	69	0	0	0	0	1	1	2	294	1	0	0	1
5	107	42	59	90	0	0	2	0	298	0	1	1	0	0	0	0
6	63	105	112	20	0	0	0	0	298	2	0	0	0	0	0	0
7	92	86	42	78	0	1	1	0	4	2	3	290	0	0	1	0
8	79	120	27	73	0	0	1	0	293	5	1	1	0	0	0	0
9	61	49	108	81	0	0	0	1	296	1	1	2	0	0	0	0
10	100	35	57	108	0	0	0	0	300	0	0	0	0	0	0	0
	Uniform								Oracle							
Runs/Zipf	1.3	1.3	1.3	1.3	2	2	2	2	1.3	1.3	1.3	1.3	2	2	2	2
1	37	35	36	54	36	29	38	35	68	62	82	88	0	0	0	0
2	39	41	28	42	44	30	47	29	33	94	109	64	0	0	0	0
3	42	28	26	48	35	46	33	42	46	103	45	106	0	0	0	0
4	37	39	35	38	41	21	48	41	52	81	66	101	0	0	0	0
5	32	34	41	37	38	28	50	40	55	31	146	68	0	0	0	0
6	35	38	32	35	46	43	36	35	87	63	74	76	0	0	0	0
7	45	33	37	34	45	24	38	44	97	48	61	94	0	0	0	0
8	41	32	36	37	33	32	44	45	62	55	78	105	0	0	0	0
9	30	41	37	47	43	33	33	36	88	107	57	48	0	0	0	0
10	31	44	36	33	44	32	43	37	64	86	97	53	0	0	0	0

Table 6: Simulations: Exploration-Exploitation. Incidence data.

	HPY-TS								Good-Turing							
Runs/Zipf	1.3	1.3	1.3	1.3	2	2	2	2	1.3	1.3	1.3	1.3	2	2	2	2
1	5	5	6	14	0	0	0	0	3	8	4	9	2	1	2	1
2	5	14	7	4	0	0	0	0	18	2	2	4	1	1	1	1
3	6	6	9	9	0	0	0	0	2	6	2	16	1	1	1	1
4	9	8	1	12	0	0	0	0	3	3	1	18	2	1	1	1
5	1	6	8	15	0	0	0	0	3	3	5	12	2	1	2	2
6	1	9	4	16	0	0	0	0	3	16	3	3	2	1	1	1
7	4	9	9	8	0	0	0	0	2	1	2	21	1	1	1	1
8	5	6	12	7	0	0	0	0	7	2	14	3	1	1	1	1
9	6	7	9	8	0	0	0	0	27	1	1	1	0	0	0	0
10	4	6	8	12	0	0	0	0	4	3	2	16	1	2	1	1
	Uniform								Oracle							
Runs/Zipf	1.3	1.3	1.3	1.3	2	2	2	2	1.3	1.3	1.3	1.3	2	2	2	2
1	7	2	0	1	7	7	2	4	8	8	7	7	0	0	0	0
2	5	2	4	2	4	3	5	5	7	7	11	5	0	0	0	0
3	4	7	3	3	2	5	6	0	7	5	12	6	0	0	0	0
4	6	4	4	1	2	3	6	4	7	3	10	10	0	0	0	0
5	4	3	4	6	3	3	3	4	8	8	9	5	0	0	0	0
6	4	2	2	6	5	6	4	1	10	6	8	6	0	0	0	0
7	1	3	2	5	9	4	5	1	10	5	10	5	0	0	0	0
8	2	5	2	0	3	8	5	5	5	11	6	8	0	0	0	0
9	4	4	3	5	4	5	2	3	5	8	10	7	0	0	0	0
10	7	4	3	1	2	4	3	6	5	9	8	8	0	0	0	0