IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Nonparametric bayesian multiarmed bandits for single-cell experiment design

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1810645 since 2021-10-08T11:46:24Z

Published version:

DOI:10.1214/20-AOAS1370

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

20 April 2024

Submitted to the Annals of Applied Statistics

arXiv:1910.05355v2 [stat.AP] 20 Sep 2020

NONPARAMETRIC BAYESIAN MULTI-ARMED BANDITS FOR SINGLE CELL EXPERIMENT DESIGN

By Federico Camerlenghi[*,†], Bianca Dumitrascu[*], Federico Ferrari[*], Barbara E. Engelhardt and Stefano Favaro[‡]

University of Milano - Bicocca, Institute for Advanced Study, Duke University, Princeton University, University of Torino and Collegio Carlo Alberto

The problem of maximizing cell type discovery under budget constraints is a fundamental challenge for the collection and analysis of single-cell RNA-sequencing (scRNA-seq) data. In this paper, we introduce a simple, computationally efficient, and scalable Bayesian nonparametric sequential approach to optimize the budget allocation when designing a large scale experiment for the collection of scRNA-seq data for the purpose of, but not limited to, creating cell atlases. Our approach relies on the following tools: i) a hierarchical Pitman-Yor prior that recapitulates biological assumptions regarding cellular differentiation, and ii) a Thompson sampling multi-armed bandit strategy that balances exploitation and exploration to prioritize experiments across a sequence of trials. Posterior inference is performed by using a sequential Monte Carlo approach, which allows us to fully exploit the sequential nature of our species sampling problem. We empirically show that our approach outperforms state-of-the-art methods and achieves near-Oracle performance on simulated and scRNA-seq data alike. HPY-TS code is available at https://github.com/fedfer/HPYsinglecell.

1. Introduction. Technological developments in high-throughput genomics have generated a wealth of data allowing researchers to measure and quantify RNA levels of individual cells (Macosko et al., 2015; Zheng et al., 2017). Benefiting from experimental and computational advances alike, single-cell RNA-seq (scRNA-seq) allows the characterization of cell types and cellular diversity, offering invaluable insights at scales unattainable in previous bulk gene expression studies (Zhu et al., 2018). In order to understand the diversity of the thousands of cell types and subtypes across

[*]These authors contributed equally to this work.

[†]Also affiliated to Collegio Carlo Alberto, Torino and BIDSA, Bocconi University, Milano, Italy.

[‡]Also affiliated to IMATI-CNR "Enrico Magenes" (Milan, Italy).

Keywords and phrases: cell type discovery, experimental sampling design, hierarchical Pitman-Yor model, multi-armed bandits, scRNA-seq, sequential Monte Carlo, Thompson sampling

different organisms, recent initiatives aim for molecular profiling of all cell types of complex organisms such as mouse or human (Regev et al., 2017; Han et al., 2018). Despite the decreasing cost of technologies for single-cell sequencing, cell atlases are expensive to collect and hard to coordinate across species, cells, tissues, organs, diseases, technologies, and labs. A principled way of collecting data is therefore paramount: given the experimental cost limiting the number of cells to be sequenced, and given multiple related experimental scenarios (e.g., developmental time, biological region, tumor site), how can one allocate the cellular sequencing budget in order to minimize experimental cost and to maximize the number of distinct cells types obtained? In this paper, we present an effective Bayesian nonparametric approach to address this fundamental question.

Recent work (Bubeck, Ernst and Garivier, 2013; Battiston, Favaro and Teh, 2018; Dumitrascu, Feng and Engelhardt, 2018a) proposed the use of classical multi-armed bandit strategies—upper confidence bounds (UCB) (Lai and Robbins, 1985; Auer, Cesa-Bianchi and Fischer, 2002) and Thompson sampling (TS) (Thompson, 1933)—for devising sequential approaches to maximize the number of distinct species discovered by sampling over multiple populations. In particular, these sampling strategies balance the exploration of the experimental choices—which populations are sampled—with the exploitation of populations that maximize current estimates of the expected rewards–the observed species diversity within a population. In the classical multi-armed bandit setting, a gambler is presented with slot machines (*one-armed bandits* is the colloquial term for a slot machine in American slang) that each pay out a random reward sampled from an arm-specific probability distribution. The gambler commits to querying a given arm for a single trial before switching to another arm, and her goal is to select a sequence of arms to play in order to maximize her rewards over subsequent trials. At each step, the gambler estimates the expected rewards of a single trial from each machine's arm, both queried and not. She must then balance exploiting the arm with the current highest estimate of expected rewards and exploring undersampled arms to improve estimates of the arms' expected rewards.

A natural variation of the above multi-armed bandit setting is when the gambler commits to querying a given arm for a pre-determined number of consecutive trials before switching to another arm. This variation is readily applicable to the experimental design problem of guiding the sequential selection of samples through single cell sequencing technologies: We may sequence some number of cells from one of multiple tissues or sample sites. In detail, we consider this problem as a set of sequential trials where a sci-

entist may choose a subset of tissue samples to assay. Each organ, tissue type, sample site, or experimental condition represents an arm to be pulled. When choosing a specific arm, the scientist commits to sequencing a number of cells proportional to the maximum number of new cell type discoveries expected in a future sample from the given experimental condition. The reward of each experimental trial is given by the number of new cell types uncovered in the sequenced sample. A first attempt to address this problem, within the context of scRNA-seq data, was proposed in Dumitrascu, Feng and Engelhardt (2018b) by combining a class of Good-Toulmin (GT) estimators (Good, 1953; Good and Toulmin, 1956; Efron and Thisted, 1976; Orlitsky, Suresh and Wu, 2016) with the TS strategy.

In this paper, we follow ideas from Battiston, Favaro and Teh (2018) and Dumitrascu, Feng and Engelhardt (2018b) to introduce a Bayesian nonparametric counterpart of the previous Good-Toulmin Thompson sampling (GT-TS) approach (Dumitrascu, Feng and Engelhardt, 2018b). Because of the purely nonparametric nature of smoothed GT estimators, the GT-TS approach does not allow us to take into account the structure of cell type diversity. As cell types arise through *cellular differentiation* (Rizvi et al., 2017), they organize themselves in developmental landscapes (Waddington et al., 1957). Hierarchical structures can be imposed on the cell types through Bayesian nonparametric priors, as was done for cell trajectory reconstruction and Bayesian inference on developmental lineages (Heaukulani, Knowles and Ghahramani, 2014; Shiffman et al., 2018).

A natural choice for a nonparametric prior to model cell type diversity is the hierarchical Pitman-Yor process (HPY) (Teh, 2006; Teh and Jordan, 2010). The HPY process has previously been used in the context of species discovery problems in multiple populations, and it has been shown to have good performance in small data sets (Camerlenghi et al., 2019; Bassetti, Casarin and Rossini, 2018). Yet, species sampling problems considered in these recent studies are not sequential problems: a Bayesian nonparametric model with a HPY prior is fit to the data *de novo* each time new data become available. This makes current posterior sampling procedures designed for the HPY prior infeasible for our sequential species sampling problem of rapidly-growing single cell data sets.

We propose a simple, computationally efficient, and scalable Bayesian nonparametric sequential approach for guiding the selection of samples for single cell sequencing technologies with the goal of maximizing the diversity of cell types discovered. Our approach has two main contributions. First, we introduce a multi-armed bandit strategy that combines the TS approach with a Bayesian nonparametric counterpart of the GT estimator under the

HPY prior, extending previous work that allowed only a single trial to allow a pre-determined number of consecutive trials before switching arms (Battiston, Favaro and Teh, 2018). The TS strategy encodes the sequential exploration-exploitation process associated with data collection from any given region, whereas the use of the HPY prior incorporates biologically-relevant information regarding the relationships among cell types to guide the allocation of resources. Second, we devise an efficient posterior sampling scheme that relies on sequential Monte Carlo methods (West, 1993a; Liu and West, 2001). Sequential Monte Carlo (SMC) allows us to fully exploit the sequential nature of our species sampling problem, thus avoiding the overwhelming computational burden of the Markov chain Monte Carlo (MCMC) scheme proposed in Battiston, Favaro and Teh (2018). We compare our method to the previous method (GT-TS) and to an Oracle in simulations and in a data set based on the Mouse Cell Atlas (Han et al., 2018). Since our motivation lies in the realm of single cell experimental design, we illustrate how, given a per-trial budget, the resulting algorithm leverages information across tissues to inform subsequent experiments in order to maximize cell type discovery in the Mouse Cell Atlas (Han et al., 2018).

The paper is structured as follows. Section 2 contains preliminaries on i) the multi-armed bandit setting within the context of prioritizing single cell sampling across populations, i.e., organs, tissues, regions, and experimental conditions; ii) the definition of the HPY prior, and some of its marginal sampling properties. In Section 3, we introduce our Bayesian nonparametric sequential approach, referred to as the HPY-TS strategy, for guiding the selection of samples through single cell sequencing technologies. A detailed description of the sequential Monte Carlo approach for posterior sampling is presented in Section 4. Section 5 highlights results of our approach through a simulation study and an application to a data set derived from the Mouse Cell Atlas. In Section 6, we summarize our work and briefly discuss extensions to our HPY-TS strategy. Additional simulation studies, posterior diagnostics, and proofs are deferred to the Supplementary Material (Camerlenghi et al., 2020).

2. Preliminaries. Let $\mathcal{Y}$ denote the set of labels representing the cell types of an organism being studied. The cell type composition within each of the J possible populations (arms, experiments) is characterized by a probability distribution over $\mathcal{Y}$, such that cell types are shared across the J populations. Precisely, we denote by P_j the probability distribution on $\mathcal{Y}$ in population j, for $j = 1, \ldots, J$. Let n_j be the number of cells (pulls) observed from the jth population, let $\boldsymbol{Y}_j = (Y_{j,1}, \cdots, Y_{j,n_j}) \in \mathcal{Y}^{n_j}$ be the vector of

n_j observations from the jth population, whereas $\boldsymbol{Y} = (\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_J)$ is the joint sample corresponding to a budget $n_1 + \cdots + n_J$. Assume now to have an additional budget M that constraints the number of cells that can be collected per trial in a future experiment. If J populations are available, a multi-armed bandit iteratively selects a subset of populations to sample from as well as the appropriate number of cells to sample in each population. In other words, at each step, the arm is chosen with the goal of maximizing the number of novel cell types observed. Therefore, in order to set up our strategy, the first step consists in estimating the number of thus far unseen species (cells) that are going to be sampled for every possible arm j, as $j = 1, \ldots, J$.

A possible strategy to address this sequential problem was first proposed in the work of Dumitrascu, Feng and Engelhardt (2018a). This approach relies on a smoothed version of the Good-Toulmin estimator of the number of unseen species (Orlitsky, Suresh and Wu, 2016). However, while the smoothed Good-Toulmin estimator presents attractive statistical properties and provable guarantees in terms of minimax optimality (Orlitsky, Suresh and Wu, 2016), it is designed for a single population scenario. In this paper, we focus on data coming from multiple related populations. Indeed, we discover cell types across diverse tissue types assayed in scRNA-seq experiments. It is then important to guarantee two key properties in our model: i) the model preserves data heterogeneity for different tissues; and ii) the model allows borrowing of information across the different tissues. Hierarchical Bayesian nonparametric priors are tailored for such situations: the data are divided into distinct populations (according to the tissue they are derived from), and at the same time the hierarchical construction allows a borrowing of information across the diverse populations of cell types.

2.1. *The Pitman-Yor process.* The Bayesian nonparametric (BNP) approach relies on the choice of a prior distribution for the cell type labels. The Dirichlet process (DP) (Ferguson, 1973) is a well-known Bayesian nonparametric distribution. In this paper, we make use of a generalization of the DP, the Pitman-Yor (PY) process (Pitman and Yor, 1997). The PY process P is a random probability measure that depends on two parameters (σ, θ), respectively called the *concentration* and the *mass* parameter, with a base measure P_0 on the space of labels $\mathcal{Y}$. The admissible values we consider here for these parameters are $\sigma \in (0, 1)$ and $\theta > 0$. The most simple way to define the PY process uses a stick-breaking procedure (Sethuraman, 1994). More specifically, P is a discrete random probability measure $P = \sum_{k \geq 1} \pi_k \delta_{y_k}$

such that

$$\pi_1 = V_1$$

and

$$\pi_k = V_k \prod_{h=1}^{k-1} (1 - V_h), \text{ for } h \geq 2,$$

where $(y_k)_{k\geq 1}$ is a sequence of i.i.d. random variables as P_0, and $(V_k)_{k\geq 1}$ is a collection of independent beta-distributed random variables with parameters $(\theta + k\sigma, 1 - \sigma)$. The two sequences $(y_k)_{k\geq 1}$ and $(V_k)_{k\geq 1}$ are assumed to be independent. We write $P \sim PY(\sigma, \theta; P_0)$ to denote the distribution of P. The classical DP prior can be found as a limiting case of the PY process, letting $\sigma \to 0$.

It is worthwhile to highlight the differences between the PY process and the DP with respect to the predictive distributions they induce. In both cases, the predictive distribution may be represented in terms of the celebrated Chinese restaurant process (CRP); see Pitman and Yor (1997) for a comprehensive account and references. Consider a random sample $Y_1, \ldots, Y_n | P \overset{iid}{\sim} P$ of size n from the PY process. The almost sure discreteness of the random probability measure P allows for ties within the sample. Then, let K_n be the number of distinct values within the sample $(Y_1, \ldots, Y_n)$, denoted as $(Y_1^*, \ldots, Y_{K_n}^*)$ and having multiplicities $(n_1, \ldots, n_{K_n})$. Then, the predictive distribution of the $n+1$st observation Y_{n+1} given past observations is

$$Y_{n+1} | (Y_1, \ldots, Y_n), P_0, \sigma, \theta \sim \sum_{k=1}^{K_n} \frac{n_k - \sigma}{\theta + n} \delta_{Y_k^*} + \frac{\theta + K_n \sigma}{\theta + n} P_0. \tag{2.1}$$

In other words, Equation (2.1) tells us that the probability of observing a previously observed value Y_k^* is proportional to $n_k - \sigma$. Intuitively, the more samples of a species we observe, the higher the probability of sampling it again in future trials; this is referred to as "the rich get richer" behavior. Alternatively, the probability of sampling a new observation from the base measure P_0 is proportional to $\theta + K_n \sigma$. Notice that the clustering structure of the PY process depends on two parameters, σ and θ, whereas in the DP it is governed only by θ. This more complex parametrization offers more flexible clustering rates and cluster size tail behaviors for the PY process (Ishwaran and James, 2001).

2.2. *Hierarchies of Pitman-Yor processes.* When considering observations sampled from multiple populations, it is natural in the Bayesian framework to model the group structure with a hierarchical framework. Here we

use a hierarchical structure based on the PY process. We denote the distribution of cell type labels across all of the populations (experimental design conditions) by P. The probability distribution P is almost surely discrete with an unknown number of atoms, and we select a PY process prior to model the distribution of cell type labels with parameters (σ, θ) and non-atomic base measure P_0 on the space of labels. Then each population-specific distribution P_j is modeled using a PY process prior with parameters (σ_j, θ_j), and we further suppose that the common base measure for all the P_js is the PY process P. Summing up, we have specified the following hierarchical PY (HPY) process prior:

$$
\begin{aligned}
&P|\sigma, \theta, P_0 \sim PY(\sigma, \theta; P_0) \\
&P_j|\sigma_j, \theta_j, P \overset{ind}{\sim} PY(\sigma_j, \theta_j; P) \quad \forall j = 1, 2, \ldots, J \\
&Y_{j,i}|P_j \overset{iid}{\sim} P_j \quad \forall j = 1, 2, \ldots, J, \quad \forall i = 1, 2, \ldots, n_j.
\end{aligned} \tag{2.2}
$$

The above hierarchical specification introduces dependencies among different populations (experimental conditions or arms), thus allowing the sharing of information across populations since the base measure P is common to the different collections of observations (Teh, 2006; Camerlenghi et al., 2019). In particular, conditional on the *base measure* P, the P_js are independent PY processes with base measure P. In particular, the interpretation of the parameters (σ_j, θ_j) is the same as in the single population case described above.

The predictive distribution and the combinatorial structure induced by hierarchical processes can be thought of in terms of the Chinese restaurant franchise (CRF) metaphor (Teh and Jordan, 2010). According to this culinary metaphor, each sample $\boldsymbol{Y}_j := (Y_{j,1}, \ldots, Y_{j,n_j})$ identifies the dishes chosen by the n_j customers of restaurant (group) j, for any $j = 1, \ldots, J$. People sitting at the same table eat the same dish, and the same dish can be served within the same restaurant or across different restaurants, since we use the same a.s. discrete base measure P for all of the groups. We denote by $Y_1^{**}, \ldots, Y_K^{**}$ the K distinct dishes across the J samples, whereas $n_{j,k} \geq 0$ represents the number of customers in restaurant j eating dish k. Finally, the vector $\boldsymbol{n}_j := (n_{j,1}, \ldots, n_{j,n_j})$ encodes all of the frequencies for a specific population j.

The combinatorial structure induced by the HPY process is formally described by the so called *partially exchangeable partition probability function*

(pEPPF) defined by

$$\Pi_k^{(n)}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J) := \mathbb{E} \int_{\mathcal{Y}^K} \prod_{j=1}^{J} \prod_{k=1}^{K} P_j^{n_{j,k}}(\mathrm{d}Y_k^{**}). \tag{2.3}$$

In other words, this is the probability of observing a specific configuration of the dishes across the restaurants. A tractable expression of the pEPPF (2.3) was found in Camerlenghi et al. (2019), resorting to auxiliary latent variables, which may be seen as tables in the CRF language. More specifically, each observation (customer) is associated with a latent tag identifying the table of the restaurant at which the specific customer is seated. We have the constraints

$$n_{j,k} = \sum_{t=1}^{m_{j,k}} n_{j,t,k},$$

where $m_{j,k}$ is the number of tables in restaurant j serving dish k, i.e., Y_k^{**}, and $n_{j,t,k}$ is the number of customers in restaurant j sitting at table t, eating dish k. In the sequel it will be useful to denote by K_j the number of distinct values in the jth group $\boldsymbol{Y}_j$, indicated by $(Y_{j,1}^*, \ldots, Y_{j,K_j}^*)$, which is a subset of $\{Y_1^{**}, \ldots, Y_K^{**}\}$.

The introduction of auxiliary variables leads to a refinement of the partition of the observations $\boldsymbol{Y}$ defined in Equation (2.3). Indeed, now we can look for the probability that the observations are partitioned into a set of $m_{\cdot,\cdot}$ distinct groups according to both tables and dishes. Such a probability coincides with an augmented version of Equation (2.3) derived in Camerlenghi et al. (2019), i.e.,

$$\begin{aligned} &\Pi_k^{(n)}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J; (n_{j,t,k})_{j,t,k}, (m_{j,k})_{j,k}) \\ &\qquad = \Phi_K^{m_{\cdot,\cdot}}(m_{\cdot,1}, \ldots, m_{\cdot,K}) \prod_{j=1}^{J} \Phi_{m_{j,\cdot},j}^{(n_j)}(n_{j,\cdot,1}, \ldots, n_{j,\cdot,K}), \end{aligned} \tag{2.4}$$

where the functions $\Phi_K^{m_{\cdot,\cdot}}$ and $\Phi_{m_{j,\cdot},j}^{(n_j)}$ denote the so-called *exchangeable partition probability function* (EPPF) induced by P and P_j, respectively. Then we have

$$\Phi_K^{m_{\cdot,\cdot}}(m_{\cdot,1}, \ldots, m_{\cdot,K}) = \frac{\prod_{i=1}^{K-1}(\theta + \sigma i)}{(\theta)_{m_{\cdot,\cdot}}} \prod_{i=1}^{K} (1-\sigma)_{m_{\cdot,i}-1},$$

where $(a)_n := \Gamma(n+a)/\Gamma(a)$ is the Pochhammer symbol for the rising factorial, and where $m_{j,\cdot}$ represents the total number of tables in group j, and

$m_{\cdot,\cdot}$ is the number of tables across restaurants. An analogous explicit formula holds for the probability $\Phi^{(n_j)}_{m_{j,\cdot},j}$ as well. Furthermore, one can then obtain an expression for Equation (2.3) by integrating out the tables in Equation (2.4).

The CRF provides a simple and meaningful interpretation of the predictive distributions for observed species within and across populations. In particular, conditional on P_j, the predictive distribution for a new observation Y_{j,n_j+1} of the jth population is the same as the CRP in the single population case. On the other hand, by integrating out P_j, we obtain the predictive distribution for the new species in population j with respect to the unique species in the joint sample (across populations). That is, we can write

$$Y^*_{j,m_{j,\cdot}+1}|Y^*_{1,1},\cdots,Y^*_{J,m_{J,\cdot}},P \sim \sum_{k=1}^{K} \frac{m_{\cdot,k}-\sigma}{\theta+m_{\cdot,\cdot}}\delta_{Y^{**}_k} + \frac{\theta+K\sigma}{\theta+m_{\cdot,\cdot}}P,$$

where $(Y^{**}_1,\cdots,Y^{**}_K)$ are the distinct species in the joint sample from J populations, and $m_{j,k}$ is the number of observations in population j from species Y^{**}_k. Notice that $m_{\cdot,k}$ is the number of times that species Y^{**}_k has been observed in the joint sample. Intuitively, a high value of $m_{\cdot,k}$ leads to a high probability of observing Y^{**}_k in all populations, even if Y^{**}_k has not yet been sampled in some of the J populations. In particular, this probability is proportional to $m_{\cdot,k}-\sigma$: the number of times that we observe Y^{**}_k minus the discount parameter of the base distribution P. In other words, the pair of parameters (θ,σ) allow us to control the total number of species in the joint sample and the extent of sharing of species across different populations. More precisely, we have that: i) if θ is low then, in expectation, the total number of distinct species in the joint sample will be low in expectation; ii) if σ is high then, in expectation, the distinct populations will share fewer species.

3. The HPY-TS strategy. In this section, we present our Bayesian nonparametric sequential approach, referred to as HPY-TS, for guiding the selection of samples for single cell sequencing technologies. HPY-TS is a multi-armed bandit strategy that combines the TS strategy with a Bayesian nonparametric counterpart of the GT estimator under the HPY process prior. Our HPY-TS strategy may be viewed as an extension of the strategy that has been recently proposed by Battiston, Favaro and Teh (2018) from a single trial before switching arms to a pre-determined number of consecutive trials before switching arms. It also may be viewed as a Bayesian nonparametric counterpart of the GT-TS strategy proposed in Dumitrascu, Feng

and Engelhardt (2018b), where the smoothed GT estimator is replaced by the Bayesian nonparametric alternative under the HPY prior including a hierarchical structure on the species.

Consider M cells that are simultaneously observed from multiple populations. Populations, i.e., organs, tissues, regions, or experimental conditions, represent arms to be selected for experimentation. Under the HPY prior assumption for the unknown composition of the populations, the HPY-TS strategy prescribes to select the population in such a way as to maximize the number of new distinct cell types that we expect to observe in M additional cells from the selected population. We define the set of hitherto unobserved cells as $A = \{y \in \mathcal{Y} : y \notin \boldsymbol{Y}\}$, and we denote by $K_j^{(M)}|\boldsymbol{Y}$ the random number of new distinct cell types that will be observed in an additional sample of size M collected from population (or arm) j. In such a situation the reward distribution for each arm j is the distribution of the random variable $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y})$, whose randomness is due to the fact that P_j is random. We remark that, conditioning on $P_j|\boldsymbol{Y}$, then $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y})$ becomes a number. The HPY-TS strategy computes draws from the posterior distribution of $(\mathbb{E}(K_1^{(M)}|\boldsymbol{Y}), \ldots, \mathbb{E}(K_J^{(M)}|\boldsymbol{Y}))$, and to select the arm j that corresponds to the maximum value of $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y})$. This strategy usually outperforms with respect to the so-called greedy strategy, which selects the arm with the highest posterior point estimate of $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y})$. Indeed, the HPY-TS better balances the exploration step as clearly explained in Battiston, Favaro and Teh (2018).

Under the HPY process prior, $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y})$ provides the natural Bayesian nonparametric counterpart of the smoothed GT estimator proposed in Dumitrascu, Feng and Engelhardt (2018b). An explicit expression for the posterior expectation $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y})$ appeared in Proposition 2 of Battiston, Favaro and Teh (2018). In the next proposition, we simplify this expression. We denote by $\text{beta}(\cdot \,|\, a, b)$ the beta distribution with parameters (a, b). Let P_j be the unknown cell type proportions of population j. Let $P_j(A)$ represent the unknown cell type proportions for the collection of cells that have not yet been sampled A from population j.

PROPOSITION 3.1. *Let the unknown cell type proportions P_j of population j be modeled according to the HPY process (Equation (2.2)). Conditional on random variables $\beta_0|\boldsymbol{Y} \sim \text{beta}(\beta_0|\theta+K\sigma, m_{\cdot\cdot}-\sigma K)$ and $P_j(A)|\boldsymbol{Y}, \beta_0 =$*

p_j, where

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

one has

$$\mathbb{E}(K_j^{(M)}|\boldsymbol{Y},\beta_0,p_j) = \frac{\theta+K\sigma}{\sigma}\Big[\sum_{i=1}^{M}\binom{M}{i}p_j^i(1-p_j)^{M-i} \times \mathbb{E}\left[\frac{(\theta+K\sigma+\sigma)_{J_i}}{(\theta+K\sigma)_{J_i}}\right] - (1-(1-p_j)^M)\Big] \tag{3.1}$$

with

$$\mathbb{E}\left[\frac{(\theta+K\sigma+\sigma)_{J_i}}{(\theta+K\sigma)_{J_i}}\right] = \sum_{\tilde{m}=1}^{i} F(i,\tilde{m},\sigma,(\theta+m_{j\cdot})\beta_0)\frac{(\theta+K\sigma+\sigma)_{\tilde{m}}}{(\theta+K\sigma)_{\tilde{m}}}.$$

Here, random variable J_i, for any $i = 1,\ldots,M$, counts the number of distinct values in a random sample of size i from a PY process with updated parameters $(\sigma,(\theta+m_{j,\cdot})\beta_0)$, and $F(n,k,\sigma,\theta)$ is the probability that $\{J_i = \tilde{m}\}$.

The proof of Proposition 3.1 is deferred to the Supplementary Material. Based on Proposition 3.1, one can recover an explicit formula for $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y})$ by simply integrating Equation (3.1) with respect to the distribution of p_j and the distribution of β_0. Then, we can infer that the computational complexity of computing the formula for $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y},\beta_0,p_j)$ is proportional to M. Having found the posterior expectation of $K_j^{(M)}|\boldsymbol{Y}$ for all populations j in Proposition 3.1, our HPY-TS strategy selects the population with the highest expected rewards, computed from a posterior sample. More specifically, we sample $\beta_0|\boldsymbol{Y}$ and $P_j(A)|\boldsymbol{Y},\beta_0$ from the distribution described in Proposition 3.1. Then, conditional on these realizations, we compute $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y},\beta_0,p_j)$ according to Equation (3.1). Finally, we select the population with the highest realized value. Details of the HPY-TS strategy are described in Section 4. Our HPY-TS strategy is based on the Thompson's sampling approach (Algorithm 1), with parameters updated sequentially according to Algorithm 2.

With regards to the choice of the prior distributions for the hyperparameters of the HPY process, we assume a Uniform prior on $(0,1)$ for both parameters σ and σ_j. Moreover, we assume a Gamma prior with parameters $(1,1)$ for both parameters θ and θ_j. All prior distributions are assumed to

be independent. Note that Algorithm 1 depends on the following collection of parameters

$$\eta = (\theta, \sigma, \sigma_j, \theta_j;\ j = 1, \cdots, J),$$

and on the table counts of the CRF, which are encoded by the vector $\boldsymbol{m}_J = (m_{j,\cdot};\ j = 1, \cdots, J)$. Here it is worth stressing that the collection of table counts $\boldsymbol{m}_J$ are latent variables, that is quantities that have not been observed in the initial sample. Therefore, before running Algorithm 1, we estimate these latent variables. This is done by using a Gibbs sampling algorithm that relies on the explicit expression of the pEPPF from the work of Camerlenghi et al. (2019). Specifically, we exploit the sequential structure of our species sampling problem to update the vector of parameters η: we describe the novel and efficient algorithm for the updating of η in the next section.

4. Sequential parameter updates. The multi-armed bandit problem is described as a sequential allocation problem, where the goal is to find the best allocation strategy to sample new observations from J different populations at every experimental time step. Whenever the new M cells are sampled from a population, one has to update the parameters of the HPY process in a computationally feasible way. A possible approach to this problem was first suggested in Battiston, Favaro and Teh (2018), where the authors propose a Markov chain Monte Carlo (MCMC) in order to estimate the posterior distributions of the hyperparameters of the HPY. However, such an approach does not take advantage of the sequential nature of the species sampling problem and, more importantly, is not computationally feasible with large data sets. The computational burden of the approach of Battiston, Favaro and Teh (2018) makes its direct application almost impossible, except for toy examples with small numbers of arms. In this section, we suggest a computationally tractable approach that leverages the sequential structure of the problem (Algorithm 1) and is based on a filtering algorithm of Liu and West (2001).

The HPY-TS strategy selects the arm to sample from, then one sequentially samples the batch of M cells from the selected arm. After that, one updates the model parameters with the new observation encoded by η to select the new arm to sample from. We then consider discrete time points $t = 1, 2, \ldots$, and we clarify how to sequentially update the parameters of our model in a computationally feasible way. Let us fix some notation: y_t is the vector of observations from the arm selected at time t, and $D_t = \{y_t, D_{t-1}\}$ is the set of observations available at time t. Thus, we can think of a model that is described by a distribution $p(y_t|\eta)$ evolving in time and depending on a vector of model parameters η. At each iteration, we select an arm and

Algorithm 1 HPY-TS

for $i \in$ 1:number of new samples **do**
  draw $\beta_0 \sim \text{beta}(\theta + \sigma K, m_{\cdot,\cdot} - \sigma K)$
  for $j \in 1 : J$ **do**
    draw $p_j \sim \text{beta}((\theta_j + m_{j,\cdot}\sigma_j)\beta_0, (\theta_j + m_{j,\cdot}\sigma_j)(1-\beta_0) + n_{j,\cdot,\cdot} - \sigma_j m_{j,\cdot})$
    Compute $\mathbb{E}(K_j^{(M)}|\boldsymbol{Y}, \beta_0, p_j)$ according to Proposition 3.1
  end for
  Compute $j^* = argmax\{\mathbb{E}(K_j^{(M)}|\boldsymbol{Y}, \beta_0, p_j),\ j = 1, \cdots, J\}$;
  Draw the next sample from population j^*;
  Update the HPY parameters according to Algorithm 2;
end for

observe y_{t+1}, and we sample the updated parameters from the posterior distribution $p(\eta|D_{t+1})$, as $t = 1, 2, \ldots$. Note that this posterior distribution is proportional to

$$p(y_{t+1}|\eta, D_t)p(\eta|D_t),$$

due to Bayes' theorem. We can think of $p(\eta|D_t)$ as the density function of η at time t. Our aim is to sample a new set of parameters from the posterior distribution of η in the presence of a new observation y_{t+1}. The key idea is to approximate the distribution of $\eta|D_t$ with a mixture of N Gaussian kernels, i.e.,

$$p(\eta|D_t) \approx \sum_{i=1}^{N} \omega_t^{(i)} \mathcal{N}(\eta|m_t^{(i)}, h^2 V_t),$$

where $\{\omega_t^{(1)}, \ldots, \omega_t^{(N)}\}$ is the set of importance sampling weights for $\{\eta_t^{(i)} : i = 1, 2, \ldots, N\}$ at time t, N is the number of importance samples at each time step, and $\mathcal{N}(\eta|m, V)$ is the density function of a Gaussian distribution with mean m and covariance V. Moreover, V_t is the estimate of the covariance with respect to the Monte Carlo posterior, and h is a smoothing parameter. Liu and West (2001) suggest to choose h as a decreasing function of the number of importance samples. In our simulations, we set $h = \frac{1}{N}$. In order to avoid "loss of information" over time, earlier work (West, 1993a,b) proposes shrinkage kernel locations and suggests setting $m_t^{(i)} = a\eta_t^{(i)} + (1-a)\bar{\eta}_t$ and $a = \sqrt{1-h^2}$, where $\bar{\eta}_t$ is the mean of the Monte Carlo sample of size N at time t. With these choices, we preserve the covariance V_t over time.

In our framework, one only needs to evaluate the conditional distribution $p(y_{t+1}|\eta, D_t)$, which may be recovered from the expression of the pEPPF in

Algorithm 2 Filtering algorithm

Evaluate $m_t^{(i)}$ for each $i = 1, 2, \ldots, N$:

$$m_t^{(i)} = a\eta_t^{(i)} + (1-a)\bar{\eta}_t,$$

which are the prior point estimates of η. Construct a posterior approximation of $p(\Theta|D_{t+1})$ with weights $\omega_{t+1}^{(i)}$ and samples $\eta_{t+1}^{(i)}$, for $i = 1, \ldots, N$, as follows.

for $i \in 1 : N$ **do**

1) Sample an auxiliary integer variable k from the set $\{1, 2, \ldots, N\}$ with probability proportional to:

$$g_{t+1}^{(i)} \propto \omega_t^{(i)} p(y_{t+1}|m_t^{(i)}, D_t)$$

2) Sample a new parameter vector $\eta_{t+1}^{(k)}$ from the kth normal component of the kernel density, namely:

$$\eta_{t+1}^{(k)} \sim N(m_t^{(k)}, h^2 \boldsymbol{V}_t)$$

3) Evaluate the corresponding weights

$$\omega_{t+1}^{(k)} \propto \frac{p(y_{t+1}|\eta_{t+1}^{(k)}, D_t)}{p(y_{t+1}|m_t^{(k)}, D_t)}$$

where

$$p(y_{t+1}|\eta, D_t) \propto \Pi_k^{(n)}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J; (n_{j,t,k})_{j,t,k}, (m_{j,k})_{j,k}).$$

In other words $p(y_{t+1}|\eta, D_t)$ is proportional to the pEPPF defined in (2.4), depending on the parameters η and on information available up to time $t+1$.

end for

Resample according to the importance weights $\omega_{t+1}^{(k)}$ to obtain a set of parameters with equal weights–in other words, a Monte Carlo approximation of the posterior.

Equation (2.4). In particular, we initially run a Gibbs sampler (Camerlenghi et al., 2019) to obtain a collection of random samples for the latent table counts and parameters η. Then, the output of the initial sample may be regarded as an importance sample of η with equal weights at time $t = 0$. More precisely, we used the Gibbs sampling scheme described in Camerlenghi et al. (2019) to initialize the values of all of the parameters η, conditionally on the data D_0, which contains the observations $\boldsymbol{Y}$. In fact, we run the MCMC procedure of Camerlenghi et al. (2019), and we used the output of the last N runs to initialize all the parameters for the particle filtering algorithm. We also assign uniform importance sampling weights to all of the initial particles, i.e., $\omega_0^{(i)} = 1/N$ as $i = 1, \ldots, N$. Then, we use Algorithm 2 to sequentially update the parameters of the HPY.

5. Applications.

5.1. *Simulation study.* We first demonstrate the performance of our HPY-TS algorithm in the context of simulated data. Additional simulation studies are presented in Section **??** of the Supplementary Material. We consider a setup with 100 arms, representing a sample corresponding to 20000 different species. The true distribution of each arm follows Zipf's law, such that the mass assigned to the kth most common species in a population j is

$$p_j(k; s_j) = \frac{1/k^{s_j}}{\sum_{i=1}^{N_j} 1/i^{s_j}},$$

where $N_j > 0$ is the number of species in population j, and $s_j > 1$ is a real parameter that controls the distribution of mass among the support– a large s_j indicates that the total mass is concentrated on a few points, and a small value indicates that the mass is shared across many points. Hence, an arm with a low s_j is a 'winning arm,' or an arm with high species diversity. Among the 100 arms, we consider 4 winning arms (Zipf parameter $s_j = 1.3$), and 96 less diverse arms (Zipf parameter $s_j = 2$). An optimal strategy should balance exploration and exploitation, and query the less diverse arms occasionally, while focusing on the winning arms.

We evaluate the performance of our TS strategy by comparing it with three baselines: the Oracle strategy, the Uniform strategy, and the Good-Toulmin Thompson sampling (GT-TS) strategy proposed by Dumitrascu, Feng and Engelhardt (2018a). In particular, the Oracle strategy is used to compare the performance with the optimal behavior; the Oracle strategy is allowed to see into the future or pre-sample from all the arms and make the optimal decision at every iteration. More precisely, the Oracle strategy

selects the arm having the highest probability of observing a new cell, where such a probability is evaluated numerically assuming knowledge of the true distribution of the data (i.e., a Zipf distribution in our experiments). The Uniform strategy allocates the budget uniformly across the arms, whereas the GT-TS strategy is based on a smoothed version of the Good-Toulmin estimator. More precisely, the smoothed Good-Toulmin estimator (Orlitsky, Suresh and Wu, 2016) estimates the number of new species that will be sampled in an additional sample of size M for a fixed population j. This estimator is defined as

$$\widehat{U}_j^{(M)}(\boldsymbol{Y}_j) = -\sum_{i=1}^{\infty}(-M/n_j)^i \mathbb{P}(L > c_j)\Phi_{i_j},$$

where M/n_j is referred to as the *extrapolation factor*, Φ_{i_j} denotes the number of species occurring with frequency i in $\boldsymbol{Y}_j$, the random sample from the jth population, and L is an independent random nonnegative integer. Common choices for the distribution of the random variable L include the Poisson distribution and the binomial distribution (Orlitsky, Suresh and Wu, 2016). The Good-Toulmin diversity estimator can be incorporated into the multi-armed bandit framework as follows. At each sampling step, an arm is chosen based on its probability of yielding the greatest number of novel species. The probability that the jth population is chosen during a trial is based on the weight of its Good-Toulmin estimator $\widehat{U}_j^{(M)}(\boldsymbol{Y}_j)$. Upon collecting M new samples from the chosen arm, the reward (the number of novel cell types) is observed, and the parameters of the Good-Toulmin estimator for the chosen population are reestimated with the new samples and reward (Dumitrascu, Feng and Engelhardt, 2018a).

In implementing our HPY-TS strategy, we use an initial sample of 20 observations from each of the 100 arms, with $M = 50$ observations sampled at each iteration. We use 500 sampling steps, and the results are averaged over 50 runs. The computations are performed in parallel, and the code is available at `https://github.com/fedfer/HPYsinglecell`. We observe that the HPY-TS algorithm performs better than the GT-TS strategy and the Uniform strategy; the latter two methods explore, but fail to exploit the most diverse arms (Fig. 1). As expected, the HPY algorithm discovers fewer new species than the Oracle strategy, but the HPY approach comes close to Oracle behavior. The results show similarities with the performance previously reported in the work of Battiston, Favaro and Teh (2018) for a simulation scenario with a small number of arms (8 arms), with the added benefit of *scalability* to an order of magnitude more arms. The good performance of the algorithm has been assessed in Section **??** of the Supplementary Mate-

rial through posterior diagnostics for the simulation scenario considered in this section.

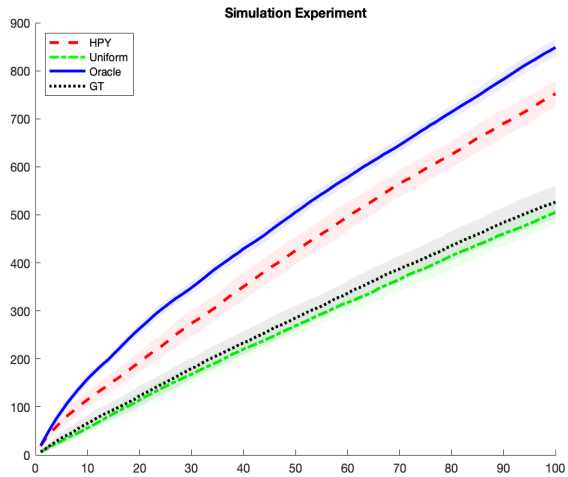

FIG 1. ***Simulation results.*** *We consider a multi-armed bandit setting for population sampling with* 100 *arms in which the species diversity follows Zipf's law with parameters* 1.3 *(*4 *high diversity, winning arms) and* 2 *(*96 *low diversity arms). An initial sample of* 20 *cells were collected from each of the* 100 *arms, with* 50 *additional cells sampled at each iteration. We used* 100 *sampling steps and averaged the results over* 50 *runs. We compared HPY-TS (red, dashed) to two baselines—the GT-TS sampler (black, dotted) and a Uniform sampling strategy (green, dot-dashed line)—and to an Oracle estimator (blue, solid). The shaded bands are within one standard deviation of the average performance, computed as the mean across simulations.*

5.2. *Application to single cell RNA-seq experimental design.* We further illustrate the advantage of our approach in the context of a simulation study based on the Mouse Cell Atlas data (Han et al., 2018). The Mouse Cell Atlas aims to provide the first high-throughout transcriptome-based single-cell atlas in a mammalian system. The project assayed over 400,000 cells from all of the major mouse organs and identified previously uncharacterized cell populations (Fig. 2). Following technical noise correction, 60,000 high-quality cells were sequenced, representing 43 distinct tissues and 98 major cell types across four developmental stages – embryo, fetal, newborn, and adult. In the collection process, equal numbers of cells were sampled uniformly across organs and developmental stages. We show that our experimental design approach achieves similar cell type diversity while requiring substantially fewer samples when compared to related methods. We follow

the simulation setup developed in prior work (Dumitrascu, Feng and Engelhardt, 2018a), outlined below.

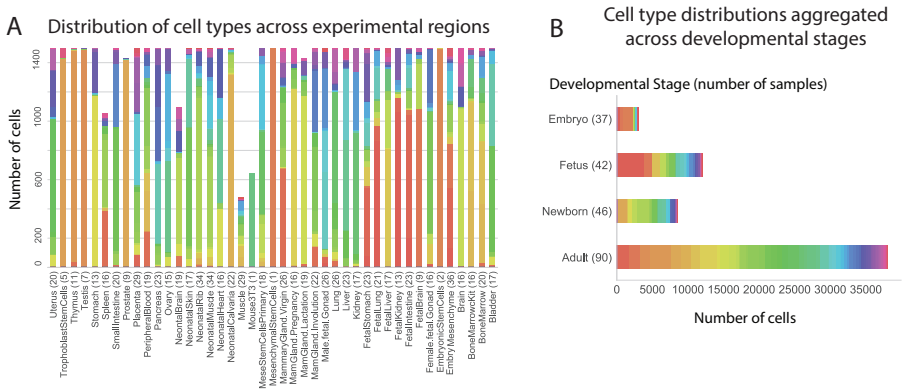

FIG 2. *Summary figure from Dumitrascu, Feng and Engelhardt (2018a) of the single cell RNA-seq data from the Mouse Cell Atlas (Han et al., 2018). The different colors represent different cell types. Panel A: Cell type distributions across tissues together with the corresponding cells and specimens. Panel B: Cell type distributions per arm: aggregated tissue types and developmental stages.*

In our simulation study, we envision a setting in which the cells were assayed in smaller batches than in the actual experiments. In particular, smaller batches are common in single cell experiments that use technologies that are less noisy but more expensive; thus experimental design plays an important role in minimizing cost (Angerer et al., 2017). Moreover, larger batches would quickly saturate the available data, so we evaluate on batch sizes that are typically smaller than are used. The 43 mouse organs were aggregated across the four developmental stages—embryo, fetal, newborn, and adult—resulting in a heterogeneous data set. Cells were sampled with replacement from each of the four experimental categories (arms), representing the four developmental stages. An experimental round corresponds to an allocation step in which the cell budget is distributed across the four experimental conditions.

We consider two ways of allocating samples: the *incidence case* and the *delayed abundance* case. In the incidence case (see Fig. 3), a single most informative experimental condition is chosen, and M samples come from that single condition. In the delayed abundance case (see Fig. 4), samples are allocated across all of the available experimental conditions in parallel. In both cases, the budget allocation step is applied using the HPY-TS strategy as follows. In the incidence case, we allocate more cells to the experiment

(i.e., developmental stage) with a higher probability of yielding new cell types based on previous trials. Following the initial sampling step with $M = 50$ samples from each arm, 20 additional trials were performed. At each time step, all $M = 25$ cells were sampled from one chosen experiment. In the delayed abundance case, after the initial $M = 50$ samples from each arm, a budget of $M = 100$ cells were distributed across arms according to the HPY-TS estimated probabilities, across 20 sequential trials. The results were averaged over 100 runs for each algorithm, and the HPY-TS sequential Monte Carlo strategy used 500 sampling steps. We compare HPY-TS to three other approaches–the GT-TS sampler, a Uniform sampling strategy, and an Oracle estimator.

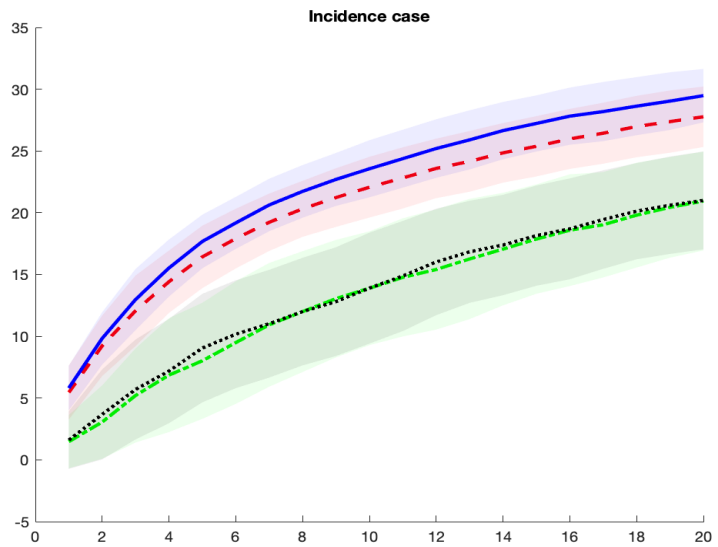

FIG 3. ***Performance of HPY-TS on the Mouse Cell Atlas data (incidence case).*** *An initial sample of $M = 50$ cells were collected from each of four populations: embryo, fetal, newborn and adult. Following the initial sampling step with $M = 50$ samples, 20 sequential trials were performed. At each time step, all $M = 25$ cells were sampled from one chosen experiment. The results were averaged over 100 runs of each algorithm. We compared HPY-TS (red, dashed) to two baselines—the GT-TS sampler (black, dotted) and a Uniform sampling strategy (green, dot-dashed line)—and to an Oracle estimator (blue, solid). The shaded bands are within one standard deviation of the average performance, computed as the mean across simulations.*

Our results show that the HPY-TS approach achieves substantial improvement in efficiency as compared to the baseline GT-TS estimator and

to the Uniform sampling strategy (Fig. 3). Moreover, the HPY-TS approach shows nearly optimal performance, as compared with the performance of the Oracle strategy. When compared to the Uniform strategy, our HPY-TS approach leads to, on average, as much as 50% more distinct cell types identified, with an average consistent margin of 10 additional distinct cell types identified across trials (Fig. 3, 4). The baseline GT-TS approach approximates the probability of observing a new cell type according to a model that assumes the cell types are distributed according to a Poisson process (Orlitsky, Suresh and Wu, 2016). In contrast, the HPY-TS algorithm assumes that all arms share a baseline distribution given by the base measure, information that is diffused across the developmental landscape to generate the developmental stage-specific cell type distributions. Sharing information across experiments using this prior appears to substantially improve performance by allowing updates of the parameters governing experiments similar the chosen experiment at each iteration, instead of only updating the chosen experiment's parameters.

6. Discussion. We propose the HPY-TS multi-armed bandit strategy, which uses the Thompson sampling strategy and a hierarchical Pitman-Yor process prior to optimize species discovery in experimental design. The HPY-TS strategy was shown to substantially improve cell type discovery in the setting of experimental design for single cell sequencing experiments. In particular, the HPY-TS strategy readily applies to cases where the number of arms corresponding to experimental conditions have substantial structure across those conditions. In particular, as cell atlases emerge, the strategy developed here is crucial to efficiently and effectively study cell type variability across new and growing experimental conditions including many thousands of simultaneous cellular perturbations (e.g., Perturb-seq (Dixit et al., 2016)) and combinatorial interventions (Horlbeck et al., 2018). The improvements that the HPY-TS strategy achieves over uniform experimental design strategies in both simulated and real data justify incorporating these types of methods in the data collection pipeline during the experimental process.

From a statistical standpoint, our work proposed a sequential Monte Carlo scheme that, unlike the previous work of Battiston, Favaro and Teh (2018), scales to a multi-sample setting and allows for inference across a large number of experiments as one finds in cell atlas development or Perturb-seq experiments. This makes our HPY-TS strategy appropriate for experimental setups with a large and growing number of arms. In this paper we demonstrate a number of advantages in using Bayesian experimental design to maximize cell type discovery within a budget during single cell RNA-sequencing

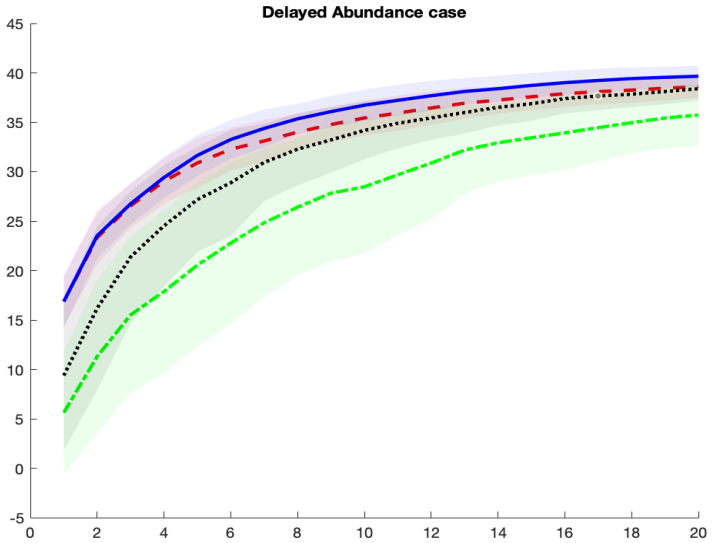

FIG 4. ***Performance of HPY-TS on the Mouse Cell Atlas data (delayed abundance).*** *An initial sample of $M = 50$ cells were collected from the four populations: embryo, fetal, newborn and adult. Following the initial sampling step, 20 additional trials were performed. At each time step, $M = 100$ samples were distributed across the arms following a diversity estimation step. The results were averaged over 100 runs of each algorithm. We compared HPY-TS (red, dashed) to two baselines—the GT-TS sampler (black, dotted) and a Uniform sampling strategy (green, dot-dashed line)—and to an Oracle estimator (blue, solid). The shaded bands are within one standard deviation of the average performance, computed as the mean across simulations.*

experiments. We further show evidence that modeling the cell type structure of single cell data using an HPY prior captures the developmental constraints guiding cell type diversity and allows each sample to inform all of the arms, leading to near-Oracle behavior.

As mentioned in Section 5.2, the optimization of cell type discovery in a multi-tissue setting was proposed in Dumitrascu, Feng and Engelhardt (2018a), which uses a strategy based on the Good-Toulmin estimator. However, the approach of Dumitrascu, Feng and Engelhardt (2018a) is empirical rather than model-based. Indeed, some paramount statistical challenges remain unsolved: i) how to model both dependence and heterogeneity across tissues in a principled statistical way? ii) how to incorporate uncertainty quantification across experimental conditions (arms) to guide arm selection at each step? iii) does a suitable statistical model, answering i)–ii), funda-

mentally improve performance? Our Bayesian nonparametric approach takes into account all these challenges: the hierarchical structure allows information to be shared across similar tissues or experimental conditions, and the uncertainty in estimated rewards across experimental conditions is incorporated in the strategy through fast posterior computations (see Proposition 3.1). Finally, the experiments in Section 5 and Section 2 of the Supplementary Material show that our model, which answers i)–ii), is able to discover more cell types sampling fewer cells with respect to the competing strategies.

In conclusion, the proposed HPY-TS strategy outperforms the current state-of-the-art strategies, and our contribution paves the way for future research in the field. We first emphasize that the number of cells that can be collected per trial M has been assumed to be fixed, since this is typically the case in cell experiments. A possible alternative, which we do not consider in our paper, would focus on optimizing each time for how many samples M should be collected over a total fixed available budget. Secondly, a natural question stemming from our analysis is understanding the effect that batch correction and cell type matching have on optimal budget allocation. In order to distinguish new cell types in a true online fashion, appropriate algorithms are needed to cluster data from new experiments, as well as reconcile the identified clusters with previously discovered ones (batch correction), in the likely presence of experiment-specific noise. In this paper, we focus on optimal experimental design under the assumption that a precise label is available at the time of the experiment. Understanding the effect of a suboptimal, possibly incorrect, or time-delayed label has on optimal experimental design is an additional area of focus for future work.

Acknowledgements. The authors would like to thank the Associate Editor and the Referees for their valuable comments and suggestions which led to a substantial improvement of the paper. The authors are indebted to David B Dunson, Jordan Bryan for helpful comments. Federico Camerlenghi and Stefano Favaro received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 817257. Federico Camerlenghi and Stefano Favaro gratefully acknowledge the financial support from the Italian Ministry of Education, University and Research (MIUR), "Dipartimenti di Eccellenza" grant 2018-2022. Bianca Dumitrascu is supported by the National Science Foundation (NSF) under grant DMS-1638352 and completed part of this work at Princeton University and while visiting The Statistical and Applied Mathematical Sciences Institute in Durham, NC, under the kind support of the NSF grant DMS-1638521. Federico Ferrari is partially

supported by grant 1R01ES028804-01 of the National Institute of Environmental Health Sciences of the United States Institutes of Health. Barbara E Engelhardt is supported by NIH R01 HL133218 and an NSF CAREER IIS-1750729.

Conflict of Interest. Barbara E Engelhardt is on the SAB for Celsius Therapeutics and Freenome, and is currently employed by Genomics plc and Freenome and on a year leave-of-absence from Princeton University.

SUPPLEMENTARY MATERIAL

Supplementary material to "Nonparametric Bayesian multi-armed bandits for single cell experiment design".
(). The supplementary material contains proofs, posterior diagnostics. and additional simulation studies.

References.

Angerer, P., Simon, L., Tritschler, S., Wolf, F. A., Fischer, D. and Theis, F. J. (2017). Single cells make big data: new challenges and opportunities in transcriptomics. *Current Opinion in Systems Biology* **4** 85–91.

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

Bassetti, F., Casarin, R. and Rossini, L. (2018). Hierarchical Species Sampling Models. *Preprint arXiv:1803.05793*.

Battiston, M., Favaro, S. and Teh, Y. W. (2018). Multi-armed bandit for species discovery: a Bayesian nonparametric approach. *Journal of the American Statistical Association* **113** 455–466.

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.

Camerlenghi, F., Lijoi, A., Orbanz, P. and Prünster, I. (2019). Distribution theory for hierarchical processes. *The Annals of Statistics* **47** 67–92.

Camerlenghi, F., Dumitrascu, B., Ferrari, F., Engelhardt, B. E. and Favaro, S. (2020). Supplementary material to "Nonparametric Bayesian multi-armed bandits for single cell experiment design".

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R. et al. (2016). Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167** 1853–1866.

Dumitrascu, B., Feng, K. and Engelhardt, B. E. (2018a). GT-TS: Experimental design for maximizing cell type discovery in single-cell data. *bioRxiv* 386540.

Dumitrascu, B., Feng, K. and Engelhardt, B. (2018b). Pólya-Gamma Thompson Sampling for Logistic Contextual Bandits. In *International Conference on Machine Learning*.

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

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

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

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

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F. et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172** 1091–1107.

Heaukulani, C., Knowles, D. and Ghahramani, Z. (2014). Beta diffusion trees. In *International Conference on Machine Learning* 1809–1817.

Horlbeck, M. A., Xu, A., Wang, M., Bennett, N. K., Park, C. Y., Bogdanoff, D., Adamson, B., Chow, E. D., Kampmann, M., Peterson, T. R. et al. (2018). Mapping the genetic landscape of human cells. *Cell* **174** 953–967.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96** 161–173.

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

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice* 197–223. Springer.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M. et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161** 1202–1214.

Orlitsky, A., Suresh, A. T. and Wu, Y. (2016). Optimal prediction of the number of unseen species. *Proceedings of the National Academy of Sciences* **113** 13283–13288.

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

Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M. et al. (2017). Science forum: the human cell atlas. *Elife* **6** e27041.

Rizvi, A. H., Camara, P. G., Kandror, E. K., Roberts, T. J., Schieren, I., Maniatis, T. and Rabadan, R. (2017). Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nature biotechnology* **35** 551.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica sinica* 639–650.

Shiffman, M., Stephenson, W. T., Schiebinger, G., Huggins, J., Campbell, T., Regev, A. and Broderick, T. (2018). Reconstructing probabilistic trees of cellular differentiation from single-cell RNA-seq data. *arXiv preprint arXiv:1811.11790*.

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 the 44th annual meeting of the Association for Computational Linguistics* 985–992. Association for Computational Linguistics.

Teh, Y. W. and Jordan, M. I. (2010). Hierarchical Bayesian nonparametric models with applications. *Bayesian nonparametrics* **1** 158–207.

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.

Waddington, C. H. et al. (1957). The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser.*

West, M. (1993a). Approximating posterior distributions by mixture. *Journal of the Royal Statistical Society. Series B (Methodological)* 409–422.

West, M. (1993b). Mixture models, Monte Carlo, Bayesian updating, and dynamic models. *Computing Science and Statistics* 325–325.

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J. et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8** 14049.

Zhu, L., Lei, J., Devlin, B., Roeder, K. et al. (2018). A unified statistical framework for single cell and bulk RNA sequencing data. *The Annals of Applied Statistics* **12** 609–632.

Federico Camerlenghi
Department of Economics, Management and Statistics
University of Milano - Bicocca
20126 Milano, Italy
E-mail: federico.camerlenghi@unimib.it

Bianca Dumitrascu
School of Mathematics
Institute for Advanced Study
1 Einstein Drive, Princeton, NJ, 08540 USA
E-mail: biancad@ias.edu

Federico Ferrari
Department of Statistical Science
Duke University
415 Chapel Dr, Durham, NC 27705
E-mail: ff31@duke.edu

Barbara E. Engelhardt
Department of Computer Science
Center for Statistics and Machine Learning
Princeton University
35 Olden Street, Princeton, NJ 08540
E-mail: bee@princeton.edu

Stefano Favaro
Department of Economics and Statistics
University of Torino
10134 Torino, Italy
E-mail: stefano.favaro@unito.it