Bayesian Analysis (2019) **14**, Number 4, pp. 1201–1219

Stochastic Approximations to the Pitman–Yor Process

Julyan Arbel*, Pierpaolo De Blasi†, and Igor Prünster‡

Abstract. In this paper we consider approximations to the popular Pitman–Yor process obtained by truncating the stick-breaking representation. The truncation is determined by a random stopping rule that achieves an almost sure control on the approximation error in total variation distance. We derive the asymptotic distribution of the random truncation point as the approximation error ϵ goes to zero in terms of a polynomially tilted positive stable random variable. The practical usefulness and effectiveness of this theoretical result is demonstrated by devising a sampling algorithm to approximate functionals of the ϵ-version of the Pitman–Yor process.

MSC 2010 subject classifications: Primary 62E20; secondary 62C10.

Keywords: stochastic approximation, asymptotic distribution, Bayesian Nonparametrics, Pitman–Yor process, random functionals, random probability measure, stopping rule.

1 Introduction

The Pitman–Yor process defines a rich and flexible class of random probability measures used as prior distribution in Bayesian nonparametric inference. It originates from the work of Perman et al. (1992), further investigated in Pitman (1995); Pitman and Yor (1997), and its use in nonparametric inference was initiated by Ishwaran and James (2001). Thanks to its analytical tractability and flexibility, it has found applications in a variety of inferential problems which include species sampling (Lijoi et al., 2007; Favaro et al., 2009; Navarrete et al., 2008), survival analysis and gene networks (Jara et al., 2010; Ni et al., 2018), linguistics and image segmentation (Teh, 2006; Sudderth and Jordan, 2009), curve estimation (Canale et al., 2017) and time-series and econometrics (Caron et al., 2017; Bassetti et al., 2014). The Pitman–Yor process is a discrete probability measure

$$P(\mathrm{d}x) = \sum_{i\geq 1} p_i \delta_{\xi_i}(\mathrm{d}x), \tag{1}$$

where $(\xi_i)_{i\geq 1}$ are independent and identically distributed (i.i.d.) random variables with common distribution P_0 on a Polish space $\mathcal{X}$, and $(p_i)_{i\geq 1}$ are random frequencies, i.e. $p_i \geq 0$ and $\sum_{i\geq 1} p_i = 1$, independent of $(\xi_i)_{i\geq 1}$. The distribution of the frequencies of the

*Inria Grenoble Rhône-Alpes, 655 Avenue de l'Europe, 38330 Montbonnot, France, julyan.arbel@inria.fr

†ESOMAS Department and Collegio Carlo Alberto, University of Torino, c.so Unione Sovietica 218/b, 10134 Torino, Italy, pierpaolo.deblasi@unito.it

‡Department of Decision Sciences and Bocconi Institute for Data Science and Analytics, Bocconi University, via Röntgen 1, 20136 Milano, Italy, igor@unibocconi.it

 https://doi.org/10.1214/18-BA1127

Pitman–Yor process is known in the literature as the two-parameter Poisson–Dirichlet distribution. Its distinctive property is that the frequencies in *size-biased order*, that is the random arrangement in the order of appearance in a simple random sampling without replacement, admit the *stick-breaking representation*, or residual allocation model,

$$p_i \stackrel{d}{=} V_i \prod_{j=1}^{i-1}(1-V_j), \quad V_j \stackrel{\text{ind}}{\sim} \text{beta}(1-\alpha, \theta+j\alpha) \tag{2}$$

for $0 \leq \alpha < 1$ and $\theta > -\alpha$, see Pitman and Yor (1997). By setting $\alpha = 0$ one recovers the Dirichlet process of Ferguson (1973). Representation (2) turns out very useful in devising finite support approximation to the Pitman–Yor process obtained by truncating the summation in (1). A general method consists in setting the truncation level n by replacing p_{n+1} with $1-(p_1+\cdots+p_n)$ in (1). The key quantity is the *truncation error* of the infinite summation (1),

$$R_n = \sum_{i>n} p_i = \prod_{j\leq n}(1-V_j), \tag{3}$$

since the resulting truncated process, say $P_n(\cdot)$, will be close to $P(\cdot)$ according to $|P(A)-P_n(A)| \leq R_n$ for any measurable $A \subset \mathcal{X}$. It is then important to study the distribution of the truncation error R_n as n gets large in order to control the approximation error. Ishwaran and James (2001) proposes to determine the truncation level based on the moments of R_n. Cf. also Ishwaran and Zarepour (2002); Gelfand and Kottas (2002). In this paper we propose and investigate a random truncation by setting n such that R_n is smaller than a predetermined value $\epsilon \in (0,1)$ with probability one. Specifically, we define

$$\tau(\epsilon) = \min\{n \geq 1 : \ R_n < \epsilon\} \tag{4}$$

as the stopping time of the multiplicative process $(R_n)_{n\geq 1}$ and, following Ghosal and van der Vaart (2017, Section 4.3.3), we call *ϵ-Pitman–Yor (ϵ-PY) process* the Pitman–Yor process truncated at $n = \tau(\epsilon)$, namely

$$P_\epsilon(\mathrm{d}x) = \sum_{i=1}^{\tau(\epsilon)} p_i \delta_{\xi_i}(\mathrm{d}x) + R_{\tau(\epsilon)}\delta_{\xi_0}(\mathrm{d}x), \tag{5}$$

where ξ_0 has distribution P_0, independent of the sequences $(p_i)_{i\geq 1}$ and $(\xi_i)_{i\geq 1}$. By construction, P_ϵ is the finite stick-breaking approximation to P with the smallest number of support points given a predetermined approximation level. In fact $\tau(\epsilon)$ controls the error of approximation according to the total variation bound

$$d_{TV}(P_\epsilon, P) = \sup_{A\subset\mathcal{X}} |P(A) - P_\epsilon(A)| \leq \epsilon \tag{6}$$

almost surely (a.s.). As such, it also guarantees the almost sure convergence of measurable functionals of P by the corresponding functionals of P_ϵ as $\epsilon \to 0$, cf. Ghosal and van der Vaart (2017, Proposition 4.20). A typical application is in Bayesian nonparametric inference on mixture models where the Pitman–Yor process is used as prior

distribution on the mixing measure. The approximation P_ϵ can be applied to the posterior distribution given the latent variables, cf. Section 2.2 for details. In the Dirichlet process case, P_ϵ has been studied by Muliere and Tardella (1998). In this setting $\tau(\epsilon)-1$ is Poisson distributed with parameter $\theta \log 1/\epsilon$, which makes an exact sampling of the ϵ-approximation (5) feasible. This has been implemented in the highly popular R software DPpackage, see (Jara, 2007; Jara et al., 2011), to draw posterior inference on the random effect distribution of linear and generalized linear mixed effect model. Finally, in Al Labadi and Zarepour (2014) a different type of finite dimensional truncation of the Pitman–Yor process based on decreasing frequencies has been proposed, see Section 5 for a discussion.

The main theoretical contribution of this paper is the derivation of the asymptotic distribution of $\tau(\epsilon)$ as $\epsilon \to 0$ for $\alpha > 0$. As (4) suggests, the asymptotic distribution of $\tau(\epsilon)$ is related to that of R_n in (3) as $n \to \infty$. According to Pitman (2006, Lemma 3.11), the latter involves a polynomially tilted stable random variable $T_{\alpha,\theta}$, see Section 2 for a formal definition. The main idea is to work with $T_n = -\log R_n$ so to deal with sums of the independent random variables $Y_i = -\log(1-V_i)$. The distribution of $\tau(\epsilon)$ can be then studied in terms of the allied *renewal counting process* $N(t) = \max\{n : \ T_n \leq t\}$, according to the relation $\tau(\epsilon) = N(\log 1/\epsilon)+1$. The problem boils down to the derivation of an appropriate a.s. convergence of $N(t)$ as $t \to \infty$, which, in turn, is obtained from the asymptotic distribution of T_n by showing that $N(t) \to \infty$ a.s. as $t \to \infty$ together with a (non standard) application of the law of large numbers for randomly indexed sequences. This strategy proves successful in establishing the almost sure convergence of $\tau(\epsilon)-1$ to $(\epsilon T_{\alpha,\theta}/\alpha)^{-\alpha/(1-\alpha)}$ as $\epsilon \to 0$. The form of the asymptotic distribution reveals how large the truncation point $\tau(\epsilon)$ is as ϵ gets small in terms of the model parameters α and θ. In particular, it highlights the power law behavior of $\tau(\epsilon)$ as $\epsilon \to 0$, namely the growth at the polynomial rate $1/\epsilon^{\alpha/(1-\alpha)}$ compared to the slower logarithmic rate $\theta \log 1/\epsilon$ in the Dirichlet process case. This is further illustrated by a simulation study in which we generate from the asymptotic distribution of $\tau(\epsilon)$ by using Zolotarev's integral representation of the positive stable distribution as in Devroye (2009). As far as the simulation of the ϵ-PY process is concerned, exact sampling is feasible by implementing the stopping rule in (4), that is by simulating the stick breaking frequencies p_j until the error R_n crosses the approximation level ϵ. As this can be computationally expensive when ϵ is small, as an alternative we propose to use the asymptotic distribution of $\tau(\epsilon)$ by simulating the truncation point first, then run the stick breaking procedure up to that point. It results in an approximate sampler of the ϵ-PY process that we compare with the exact sampler in a simulation study involving moments and mean functionals.

The rest of the paper is organized as follows. In Section 2, we derive the asymptotic distribution of $\tau(\epsilon)$ and explain how to use it to simulate from the ϵ-PY process. Section 3 reports a simulation study on the distribution of $\tau(\epsilon)$ and on functionals of the ϵ-PY process. In Section 4, to help the understanding and gain additional insight on the asymptotic distribution, we highlight the connections of $\tau(\epsilon)$ with Pitman's theory on random partition structures. We conclude with a discussion of open problems in Section 5. The details of Devroye's algorithm for generating from a polynomially tilted positive stable random variable are given in Supplementary Material (Arbel etal., 2018).

2 Theory and algorithms

2.1 Asymptotic distribution of $\tau(\epsilon)$

In this section we derive the asymptotic distribution of the stopping time $\tau(\epsilon)$ and show how to simulate from it. We start by introducing the renewal process interpretation which is crucial for the asymptotic results. As explained in the previous section, in order to study the distribution of $\tau(\epsilon)$ it is convenient to work with the log transformation of the truncation error R_n in (3), that is

$$T_n = \sum_{i=1}^{n} Y_i, \quad Y_i = -\log(1 - V_i), \tag{7}$$

with $V_j \overset{\text{ind}}{\sim} \text{beta}(1-\alpha, \theta + j\alpha)$ as in (2). Being a sum of independent and nonnegative random variables, $(T_n)_{n\geq 1}$ takes the interpretation of a (generalized) renewal process with independent waiting times Y_i. For $t \geq 0$ define

$$N(t) = \max\{n : \ T_n \leq t\}, \tag{8}$$

to be the *renewal counting process* associated to $(T_n)_{n\geq 1}$, which is related to $\tau(\epsilon)$ via $\tau(\epsilon) = N(\log 1/\epsilon) + 1$. Classical renewal theory pertains to iid waiting times while here there is no identity in distribution unless $\alpha = 0$, i.e. the Dirichlet process case. In the latter setting, one gets $Y_i \overset{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$ so that T_n has gamma distribution with scale parameter n. We immediately get from the relation $\{T_n \leq t\} = \{N(t) \geq n\}$ that $N(t) \sim \text{Pois}(\theta t)$ and, in turn, that $\tau(\epsilon)-1$ has $\text{Pois}(\theta \log(1/\epsilon))$ distribution. As far as asymptotics is concerned, T_n satisfies the central limit theorem (CLT) with $(T_n - n/\theta)/(\sqrt{n}/\theta) \to_d Z$ where $Z \sim \text{N}(0,1)$. The asymptotic distribution of $N(t)$ can be obtained via Ascombe theorem, cf. Gut (2013, Theorem 7.4.1), to get $(N(t) - \theta t)/(\sqrt{\theta t}) \to_d Z$, as $t \to \infty$, in accordance with the standard normal approximation of the Poisson distribution with large rate parameter.

In the general Pitman–Yor case $\alpha > 0$, the waiting times Y_i are no more identically distributed. More importantly, generalizations of the CLT such as the Lindeberg–Feller theorem do not apply for T_n, hence we cannot resort to Anscombe's theorem to derive the asymptotic distribution of $N(t)$ and, in turn, of $\tau(\epsilon)$. Nevertheless, the limit exists but is not normal as stated in Theorem 1 below. To this aim, let T_α be a positive stable random variable with exponent α, that is $\text{E}(\text{e}^{-sT_\alpha}) = \text{e}^{-s^\alpha}$, and denote its density by $f_\alpha(t)$. A polynomially tilted version of T_α is defined as the random variable $T_{\alpha,\theta}$ with density proportional to $t^{-\theta} f_\alpha(t)$, that is

$$f_{\alpha,\theta}(t) = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)} t^{-\theta} f_\alpha(t), \quad t > 0. \tag{9}$$

The random variable $T_{\alpha,\theta}$ is of paramount importance in the theory of random partition structures associated to the frequency distribution of the Pitman–Yor process, see Section 4 for details. In particular, the convergence of R_n can be expressed in terms of $T_{\alpha,\theta}$. In Theorem 1 the a.s. limit of $\log N(t)$ as $t \to \infty$ is obtained from that of $T_n = -\log R_n$ as $n \to \infty$ by showing that $N(t) \to \infty$ a.s. as $t \to \infty$ and by an application of the law of large numbers for randomly indexed sequences.

Theorem 1. *Let $N(t)$ be defined in (7)–(8) and $T_{\alpha,\theta}$ be the random variable with density in (9). Then $t-(1/\alpha-1)\log N(t)+\log\alpha \to_{a.s.} \log T_{\alpha,\theta}$ as $t\to\infty$.*

Proof. By definition (8), the renewal process $N(t)$ is related to the sequence of renewal epochs T_n through

$$\{T_n \le t\} = \{N(t) \ge n\}. \tag{10}$$

Since $N(T_n)=n$, we have $T_{N(t)}=T_n$ when $t=T_n$, thus $0=t-T_{N(t)}$ for $t=T_n$. Moreover, since $N(t)$ is increasing, when $T_n<t<T_{n+1}$, $N(T_n)<N(t)<N(t)+1$, hence $T_{N(t)}<t<T_{N(t)+1}$, i.e. $0<t-T_{N(t)}<T_{N(t)+1}-T_{N(t)}=Y_{N(t)+1}$. Together the two relations above yield

$$0 \le t-T_{N(t)} < Y_{N(t)+1}. \tag{11}$$

From Lemma 3.11 of Pitman (2006) and an application of the continuous mapping theorem (see Theorem 10.1 in Gut, 2013) the asymptotic distribution of T_n is obtained as

$$T_n-(1/\alpha-1)\log n+\log\alpha \to_{a.s.} \log T_{\alpha,\theta} \quad \text{as } n\to\infty.$$

Now we would like to take the limit with respect to $n=N(t)$ as $t\to\infty$, that is apply the law of large numbers for randomly indexed sequence (see Theorem 6.8.1 in Gut, 2013). To this aim, we first need to prove that $N(t)\to_{a.s.}\infty$ as $t\to\infty$. Since $N(t)$ is non decreasing, by an application of Theorem 5.3.5 in Gut (2013), it is sufficient to prove that $N(t)\to\infty$ in probability as $t\to\infty$, that is $\mathrm{P}(N(t)\ge n)\to 1$ as $t\to\infty$ for any $n\in\mathbb{N}$. But this is an immediate consequence of the inversion formula (10). We have then established that

$$T_{N(t)}-(1/\alpha-1)\log N(t)+\log\alpha \to_{a.s.} \log T_{\alpha,\theta} \quad \text{as } t\to\infty.$$

To conclude the proof, we need to replace $T_{N(t)}$ with t in the limit above. Note that, from (11), $|t-T_{N(t)}|\le Y_{N(t)+1}$ so it is sufficient to show that the upper bound goes to zero a.s.. Actually, by a second application of Theorem 6.8.1 in Gut (2013) it is sufficient to show that $Y_n\to_{a.s.}0$ as $n\to\infty$. This last result is established as follows. Recall that $Y_j=-\log(1-V_j)$ for $V_j\overset{\text{ind}}{\sim}\text{beta}(1-\alpha,\theta+j\alpha)$. For $\epsilon>0$,

$$\begin{aligned}
\mathrm{P}(1-V_n<\mathrm{e}^{-\epsilon}) &= \int_0^{\mathrm{e}^{-\epsilon}} \frac{\Gamma(\theta+n\alpha+1-\alpha)}{\Gamma(\theta+n\alpha)\Gamma(1-\alpha)} v^{\theta+n\alpha-1}(1-v)^{-\alpha}\mathrm{d}x \\
&\le \frac{(1-\mathrm{e}^{-\epsilon})^{-\alpha}}{\Gamma(1-\alpha)} \frac{\Gamma(\theta+n\alpha+1-\alpha)}{\Gamma(\theta+n\alpha)} \frac{\mathrm{e}^{-\epsilon(\theta+n\alpha)}}{\theta+n\alpha} \\
&= \frac{(1-\mathrm{e}^{-\epsilon})^{-\alpha}}{\Gamma(1-\alpha)} (\theta+n\alpha)^{-\alpha}\mathrm{e}^{-\epsilon(\theta+n\alpha)}\Big(1+O\Big(\frac{1}{\theta+n\alpha}\Big)\Big), \qquad (12)
\end{aligned}$$

where in equality (12) we have used Euler's formula

$$\Gamma(z+\alpha)/\Gamma(z+\beta) = z^{\alpha-\beta}\left[1+\frac{(\alpha-\beta)(\alpha+\beta-1)}{2z}+O(z^{-2})\right]$$

for $z\to\infty$, see Tricomi and Erdélyi (1951). Since $\mathrm{P}(Y_n>\epsilon)=\mathrm{P}(1-V_n<\mathrm{e}^{-\epsilon})$, (12) implies that $\mathrm{P}(Y_n>\epsilon)$ is exponentially decreasing in n and, in turn, that $\sum_{n\ge1}\mathrm{P}(Y_n>\epsilon)<\infty$. An application of Borel–Cantelli Lemma yields $Y_n\to_{a.s.}0$ and the proof is complete. □

The asymptotic distribution of $\tau(\epsilon)$ is readily derived from Theorem 1 via the formula $\tau(\epsilon) = N(\log 1/\epsilon)+1$ and an application of the continuous mapping theorem. The proof is omitted.

Theorem 2. *Let $\tau(\epsilon)$ be defined in (4) and $T_{\alpha,\theta}$ be the random variable with density in (9). Then $\tau(\epsilon) - 1 \sim_{a.s.} (\epsilon T_{\alpha,\theta}/\alpha)^{-\alpha/(1-\alpha)}$ as $\epsilon \to 0$.*

In order to sample from the asymptotic distribution of $\tau(\epsilon)$, the key ingredient is random generation from the polynomially tilted stable random variable $T_{\alpha,\theta}$. Following Devroye (2009), we resort to Zolotarev's integral representation, so let $A(u)$ be the Zolotarev function

$$A(x) = \left(\frac{\sin(\alpha x)^{\alpha} \sin((1-\alpha)x)^{1-\alpha}}{\sin(x)} \right)^{\frac{1}{1-\alpha}}, \quad x \in [0, \pi]$$

and $Z_{\alpha,b}$, $\alpha \in (0,1)$ and $b > -1$ be a Zolotarev random variable with density given by

$$f(x) = \frac{\Gamma(1+b\alpha)\Gamma(1+b(1-\alpha))}{\pi\Gamma(1+b)A(x)^{b(1-\alpha)}}, \quad x \in [0, \pi].$$

According to Theorem 1 of Devroye (2009), for G_a a gamma distributed random variable with shape $a > 0$ and unit rate,

$$T_{\alpha,\theta} \stackrel{d}{=} \left(\frac{A(Z_{\alpha,\theta/\alpha})}{G_{1+\theta(1-\alpha)/\alpha}} \right)^{\frac{1-\alpha}{\alpha}}$$

so that random variate generation simply requires one gamma random variable and one Zolotarev random variable. For the latter, rejection sampler can be used as detailed in Devroye (2009). See Algorithm 3 in Supplementary Material.

2.2 Simulation of the ϵ-PY process

Given α, θ, ϵ and a probability measure P_0 on $\mathcal{X}$, an ϵ-PY process can be generated by implementing the stopping rule in the definition of $\tau(\epsilon)$, cf. (4). The algorithm consists in a while loop as follows:

Algorithm 1 (Exact sampler of ϵ-PY)

1. set $i = 1$, $R = 1$
2. while $R \geq \epsilon$: generate V from beta$(1-\alpha, \theta+i\alpha)$.
   set $p_i = VR$, $R = R(1-V)$, $i = i+1$
3. set $\tau = i$, $R_\tau = R$
4. generate $\tau+1$ random variates $\xi_0, \xi_1, \ldots, \xi_\tau$ from P_0
5. set $P_\epsilon(\mathrm{d}x) = \sum_{i=1}^{\tau} p_i \delta_{\xi_i}(\mathrm{d}x) + R_\tau \delta_{\xi_0}(\mathrm{d}x)$

When ϵ is small, the while loop happens to be computationally expensive since conditional evaluations at each iteration slow down computation, and memory allocation for the frequency and location vectors cannot be decided beforehand. In order to avoid these pitfalls and make the algorithm faster, one should generate the stopping time $\tau(\epsilon)$

first, and the frequencies up to that point later. We propose to exploit the asymptotic distribution of $\tau(\epsilon)$ in Theorem 2 as follows:

ALGORITHM 2 (Approximate sampler of ϵ-PY)

`1: generate` $T \overset{d}{=} T_{\alpha,\theta}$

`2: set` $\tau \leftarrow 1 + \lfloor (\epsilon T/\alpha)^{-\alpha/(1-\alpha)} \rfloor$

`3. for` $i = 1, \ldots, \tau$: `generate` V_i `from` $\text{beta}(1-\alpha, \theta + i\alpha)$.
`set` $p_i = V_i \prod_{j=1}^{i-1}(1 - V_j)$

`4. set` $R_\tau = 1 - \sum_{i=1}^{\tau} p_i = \prod_{i=1}^{\tau}(1 - V_j)$

`5: generate` $\tau + 1$ `random variates` $\xi_0, \xi_1, \ldots, \xi_\tau$ `from` P_0

`6: set` $P_\epsilon(\mathrm{d}x) = \sum_{i=1}^{\tau} p_i \delta_{\xi_i}(\mathrm{d}x) + R\delta_{\xi_0}(\mathrm{d}x)$

ALGORITHM 2 is an *approximate* sampler of the ϵ-PY process (while ALGORITHM 1 is an *exact* one) since it introduces two sources of approximations. First, through the use of the asymptotic distribution of $\tau(\epsilon)$. Second, through `Step 3` since the V_i's are not generated according to the conditional distribution given $\tau(\epsilon)$, rather unconditionally. Finding the conditional distribution of V_i, or an asymptotic approximation thereof, is not an easy task and is object of current research. In terms of the renewal process interpretation in (7)–(8), the problem is to generate the waiting times $Y_i = -\log(1-V_i)$, $i = 1, \ldots, n$, from the conditional distribution of the renewal epochs $(T_1, \ldots, T_n)$ given $N(t) = n$ for $t = -\log 1/\epsilon$.

A typical use of samples from the Pitman–Yor process we have in mind is in infinite mixture models. In fact, the discrete nature of the Pitman–Yor process makes it a suitable prior on the *mixing distribution*. ALGORITHM 1 or ALGORITHM 2 can be then applied to approximate a functional of the posterior distribution of the mixing distribution. In such models, the process components can be seen as latent features exhibited by the data. Let P denote such a process, n denote the sample size and $X_{1:n} = (X_1, \ldots, X_n)$ be an exchangeable sequence from P, that is $X_{1:n}|P \overset{\text{i.i.d.}}{\sim} P$. Variables $X_{1:n}$ are latent variables in a model conditionally on which observed data $Y_{1:n}$ come from: $Y_j|X_j \overset{\text{ind}}{\sim} f(\cdot\,|X_j)$ where f denotes a kernel density. Actually, independence is not necessary here and applications also encompass dependent models such as Markov chain transition density estimation. In order to deal with the infinite dimensionality of the process, a strategy is to marginalize it and to draw posterior inference with a *marginal* sampler. Since draws from a marginal sampler allows to make inference only on posterior expectations of the process, for more general functionals of P, in the form of $\psi(P)$, one typically needs to resort to an additional sampling step. Exploiting the composition rule $\mathcal{L}(\psi(P)|Y_{1:n}) = \mathcal{L}(\psi(P)|X_{1:n}) \times \mathcal{L}(X_{1:n}|Y_{1:n})$ this additional step boils down to sampling P conditional on latent variables $X_{1:n}$. At this stage, recalling the conditional conjugacy of the Pitman–Yor process is useful. Among $X_{1:n}$, there are a number $k \leq n$ of unique values that we denote by $X^*_{1:k}$. Let $n^*_{1:k}$ denote their frequencies. Then the following identity in distribution holds

$$P|X_{1:n} = \sum_{j=1}^{k} q_j \delta_{X^*_j} + q_{k+1} P^*,$$

where, independently, $(q_1, \ldots, q_k, q_{k+1}) \sim \text{Dirichlet}(n_1^* - \alpha, \ldots, n_k^* - \alpha, \theta + \alpha k)$ and P^* is a Pitman–Yor process of parameter $(\alpha, \theta + \alpha k)$, see Corollary 20 of Pitman (1996). Thus sampling from $\mathcal{L}(P|X_{1:n})$, hence from $\mathcal{L}(\psi(P)|X_{1:n})$, requires sampling the infinite dimensional P^*. Cf. Ishwaran and James (2001, Section 4.4). For the sake of comparison, the conjugacy of the Dirichlet process similarly leads to the need of sampling an infinite dimensional process, where $P|X_{1:n}$ takes the form of a Dirichlet process. As already noticed, the truncation of the Dirichlet process is very well understood, both theoretically and practically. The popular R package DPpackage (Jara, 2007; Jara et al., 2011) makes use of the *posterior* truncation point $\tau^*(\epsilon)$, as defined in (5), but here with respect to the posterior distribution of the process. Thus, it satisfies $\tau^*(\epsilon) - 1 \sim \text{Pois}((\theta + n)\log(1/\epsilon))$, where $\theta + n$ is the precision of the posterior Dirichlet process. Adopting here similar lines for the Pitman–Yor process, we replace P^* by the truncated process P_ϵ^*

$$P_\epsilon^*(\mathrm{d}x) = \sum_{i=1}^{\tau^*(\epsilon)} p_i^* \delta_{\xi_i}(\mathrm{d}x) + R_{\tau^*(\epsilon)} \delta_{\xi_0}(\mathrm{d}x),$$

cf. (5). Here $(p_i^*)_{i \geq 1}$ are defined according to (2) with $\theta + \alpha k$ in place of θ, i.e. $V_j \overset{\text{ind}}{\sim} \text{beta}(1 - \alpha, \theta + \alpha(k + j))$. Hence, according to Theorem 2 we have

$$\tau^*(\epsilon) - 1 \sim_{a.s.} (\epsilon T_{\alpha,\theta+\alpha k}/\alpha)^{-\alpha/(1-\alpha)}, \quad \text{as } \epsilon \to 0$$

hence Algorithm 2 can be applied here.

3 Simulation study

3.1 Stopping time $\tau(\epsilon)$

According to Theorem 2, the asymptotic distribution of $\tau(\epsilon)$ changes with ϵ, α and θ. For illustration, we simulate $\tau(\epsilon)$ from `Steps 1.-2.` in Algorithm 2 using Devroye's sampler, cf. Algorithm 3 in Supplementary Material. In Figure 1 we compare density plots obtained with 10^4 iterations with respect to different combinations of ϵ, α and θ. The plot in the left panel shows how smaller values of ϵ result in larger values of $\tau(\epsilon)$. In fact, as $\epsilon \to 0$, $\tau(\epsilon)$ increases proportional to $1/\epsilon^{\alpha/(1-\alpha)}$. Note also that $(\epsilon T_{\alpha,\theta}/\alpha)^{-\alpha/(1-\alpha)}$ is nonnegative for $T_{\alpha,\epsilon} < \alpha/\epsilon$, which happens with high probability when ϵ is small. As for α, the plot in the central panel shows how $\tau(\epsilon)$ increases as α gets large. In fact, it is easy to see that $(\epsilon T_{\alpha,\theta}/\alpha)^{-\alpha/(1-\alpha)}$ is increasing in α when $T_{\alpha,\epsilon} < \mathrm{e}^{1-\alpha}\alpha/\epsilon$, which also happens with high probability when ϵ is small, so the larger α, the more stick-breaking frequencies are needed in order to account for a prescribed approximation error ϵ. Finally, the plot in the right panel shows that the larger θ, the larger $\tau(\epsilon)$. In fact, by definition, the polynomial tilting makes $T_{\alpha,\theta}$ stochastically decreasing in θ.

In order to illustrate the rate of convergence in Theorem 2, we compare next the exact distribution of $\tau(\epsilon)$ with the asymptotic one. To do so, we repeat the following experiment several times: we simulate $\tau(\epsilon)$ from `Steps 1.-3.` in Algorithm 1, then we compare the empirical distribution of $(\epsilon/\alpha)^\alpha(\tau(\epsilon) - 1)^{1-\alpha}$ with $T_{\alpha,\theta}^{-\alpha}$, the latter corresponding to the α-diversity of the PY process, see Section 4 for a formal definition.

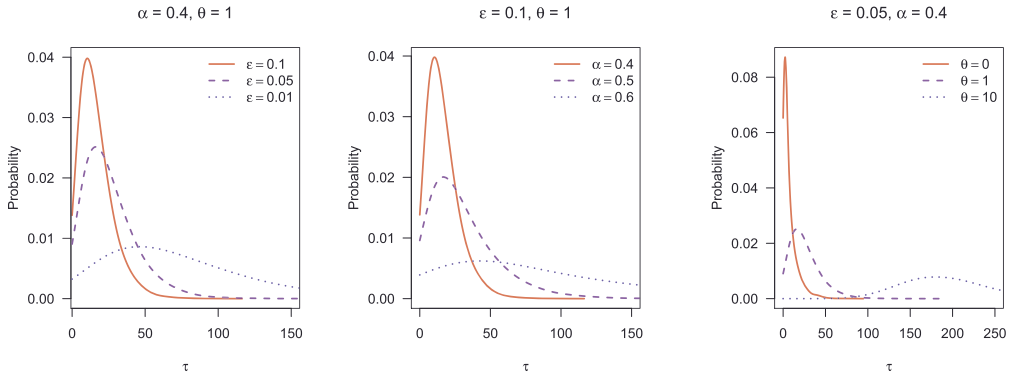

Figure 1: Density plot for the asymptotic approximation of $\tau(\epsilon)$ based on 10^4 values under the following parameter configurations. Left: $\epsilon \in \{0.10, 0.05, 0.01\}$, $\alpha = 0.4$, $\theta = 1$. Center: $\alpha \in \{0.4, 0.5, 0.6\}$, $\theta = 1$, $\epsilon = 0.1$. Right: $\theta \in \{0, 1, 10\}$, $\alpha = 0.25$, $\epsilon = 0.05$.

In Table 1 we report the Kolmogorov distance together with expected value, median, first and third quartiles for both the exact and the asymptotic distribution obtained with 10^4 iterations. This is repeated for $\alpha = 0.5$, $\theta = \{0, 1, 10\}$ and $\epsilon = \{0.10, 0.05, 0.01\}$. As expected, as we decrease ϵ, the Kolmogorov distance gets smaller to somehow different rates according to the parameter choice. The derivation of convergence rates is left for future research.

		d_K	Mean		25%		Median		75%	
θ	ϵ		As	Ex	As	Ex	As	Ex	As	Ex
0	0.10	3.42	1.06	1.05	0.45	0.45	0.89	0.89	1.61	1.55
0	0.05	2.17	1.10	1.08	0.45	0.45	0.95	0.95	1.64	1.58
0	0.01	1.73	1.14	1.11	0.45	0.45	0.97	0.95	1.64	1.60
1	0.10	4.79	2.24	2.14	1.55	1.48	2.14	2.10	2.86	2.76
1	0.05	2.38	2.25	2.20	1.55	1.52	2.17	2.14	2.86	2.79
1	0.01	1.40	2.26	2.25	1.57	1.54	2.19	2.19	2.87	2.85
10	0.10	11.93	6.39	6.07	5.69	5.40	6.34	6.06	7.04	6.72
10	0.05	6.12	6.39	6.24	5.70	5.56	6.34	6.22	7.05	6.88
10	0.01	1.93	6.40	6.37	5.71	5.70	6.34	6.34	7.05	7.00

Table 1: Summary statistics for the asymptotic distribution (As) and exact distribution (Ex) of $\tau(\epsilon)$ at the scale of the α-diversity based on 10^4 values. The Kolmogorov distance (d_K) is between the empirical cumulative distribution function of the sample from the exact distribution and the asymptotic one (multiplied by a factor of 100). The parameter values are $\alpha = 0.5$, $\theta \in \{0, 1, 10\}$ and $\epsilon \in \{0.10, 0.05, 0.01\}$.

3.2 Functionals of the ϵ-PY process

In the case that P is defined on $\mathcal{X} \subseteq \mathbb{R}$, the total variation bound (6) implies that $|F(x) - F_\epsilon(x)| < \epsilon$ almost surely for any $x \in \mathbb{R}$, where F_ϵ and F are the cumulative

distribution functions of P_ϵ and P. Also, measurable functionals $\psi(P)$ such as the mean $\mu = \int xP(\mathrm{d}x)$ can be approximated in distribution by the corresponding functionals $\psi(P_\epsilon)$. For illustration, we set $\mathcal{X} = [0,1]$ and P_0 the uniform distribution on $[0,1]$. For given α and θ, we then compare the distribution under P with that under the ϵ-PY process P_ϵ for $F(1/2)$, $F(1/3)$ and $\mu = \int xP(\mathrm{d}x)$. As for the distribution of the finite dimensional distributions $F(1/2)$ and $F(1/3)$ under the full process P, we set $\alpha = 0.5$ so to exploit results in James et al. (2010). According to their Proposition 4.7, the finite dimensional distributions of P when $\alpha = 0.5$ are given by

$$f(w_1,\ldots,w_{n-1}) = \frac{(\prod_{i=1}^n p_i)\Gamma(\theta+n/2)}{\pi^{(n-1)/2}\Gamma(\theta+1/2)} \frac{w_1^{-3/2}\cdots w_{n-1}^{-3/2}(1-\sum_{i=1}^{n-1} w_i)^{-3/2}}{\mathcal{A}_n(w_1,\ldots,w_{n-1})^{\theta+n/2}}$$

for any partition $A_1,\ldots,A_n$ of $\mathcal{X}$ with $p_i = P_0(A_i)$ and $\mathcal{A}_n(w_1,\ldots,w_{n-1}) = \sum_{i=1}^{n-1} p_i^2 w_i^{-1} + p_n^2(1-\sum_{i=1}^{n-1} w_i)^{-1}$. Direct calculation shows that $F(1/2)$ has beta distribution with parameters $(\theta+1/2,\theta+1/2)$ while $F(1/3)$ has density

$$f(w) = \frac{2}{\sqrt{\pi}} 9^\theta \frac{\Gamma(\theta+1)}{\Gamma(\theta+1/2)} \frac{(w(1-w))^{\theta-1/2}}{(1+3w)^{\theta+1}}.$$

As for the mean functional $\mu = \int xP(\mathrm{d}x)$, the distribution under the full process P is approximated by simulations by setting a deterministic truncation point sufficiently large. As for the distribution under P_ϵ, we use both Algorithm 1 and Algorithm 2.

In Figure 2 we compare the density plots of $F(1/2)$ for $\epsilon = \{0, 1.0.05, 0.001\}$ and $\theta = \{0, 10\}$ under P_ϵ with the beta density under P so to illustrate that the two distributions get close as ϵ gets small. As for $F(1/3)$ and $\mu = \int xP(\mathrm{d}x)$, in Tables 2 and 3 we report the Kolmogorov distance between P and P_ϵ for the two sampling algorithms, together with expected value, median, first and third quartiles. For each case and each parameter configuration, we have sampled 10^4 trajectories from the ϵ-PY process and 10^4 trajectories from the Pitman–Yor process in the case of $\mu = \int xP(\mathrm{d}x)$. As expected, the Kolmogorov distances are generally larger, still close, when using Algorithm 2 versus Algorithm 1 due to the approximate nature of the former.

3.3 Computation time

In this section, we provide a concrete justification of the computational advantage of using Algorithm 2 versus Algorithm 1. We simulate 10^4 ϵ-PY iterations by using Algorithm 1 and Algorithm 2 for different combinations of the α and θ parameters and of the ϵ error threshold. In Table 4 (resp. Table 5), we report the average computing time[1] *per iteration* (resp. *per support point*) for Algorithm 1 and Algorithm 2. By *iteration*, we mean a full realization of the ϵ-PY process including frequencies and locations, while by *support point*, we mean that we divide the total time by the number of support points $\tau(\epsilon)+1$. In order to account for the computational task required per iteration, the expected stopping time $\mathrm{E}[\tau(\epsilon)]$ is also reported. Both tables illustrate that our proposed approach is faster than Algorithm 1 when the ϵ-PY is composed of about 20 support points or more. The more support points, the faster Algorithm 2

[1]The experiments were conducted on an Intel Core i5 processor (3.1 GHz) computer.

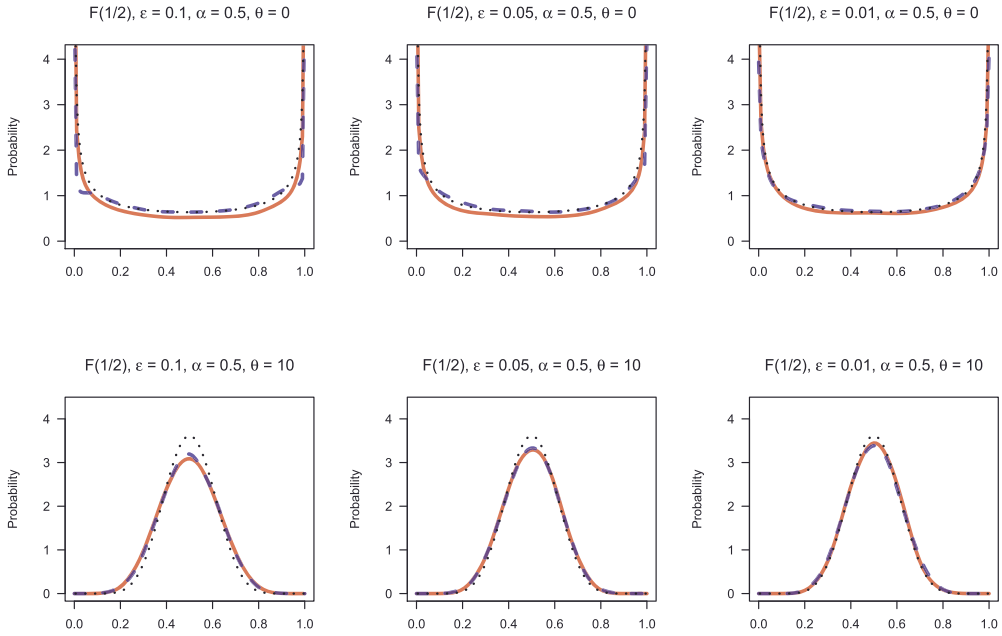

Figure 2: Density plots for the random probability $F(1/2)$ using the ALGORITHM 2 (in red solid curve) and ALGORITHM 1 (in blue dashed curve) to sample from the ϵ-PY process. The density under the Pitman–Yor process is the black dotted curve. The parameter α is fixed equal to 0.5, θ is equal to 0 on the first row and 10 on the second row, while ϵ is respectively equal to $\{0.10, 0.05, 0.01\}$ in the left, center and right columns.

is compared to ALGORITHM 1. This disadvantage of the former for small numbers of support points comes from the fixed cost of initially generating a random variable with the same distribution as $T_{\alpha,\theta}$. Conversely, as the number of support points increases, this fixed cost is largely counterbalanced by the fast vector-sampling of a prescribed size, which is in contrast with ALGORITHM 1 `while` loop whose cost increases with the number of support points. This can be seen in Table 5 where the actual sampling time per support point is essentially increasing for ALGORITHM 1 and decreasing for ALGORITHM 2. With the parameter configurations tested, ALGORITHM 2 can be up to 90 times faster ALGORITHM 1 for $\alpha = 0.6$, $\theta = 10$ and $\epsilon = 0.01$.

4 Connections with random partition structures

4.1 α-diversity and asymptotic distribution of R_n

The random variable $T_{\alpha,\theta}$ in Theorem 1 plays a key role in the Pitman–Yor process, in particular for its link with the α-*diversity* of the process. The α-diversity is defined

		d_K		Mean			25%			Median			75%		
θ	ϵ	AL1	AL2	AL1	AL2	PY	AL1	AL2	PY	AL1	AL2	PY	AL1	AL2	PY
0	0.10	16.29	16.48	0.33	0.33	0.33	0.04	0.01	0.04	0.20	0.16	0.20	0.60	0.64	0.59
0	0.05	11.53	12.52	0.33	0.33	0.33	0.05	0.01	0.04	0.20	0.17	0.20	0.58	0.63	0.59
0	0.01	5.49	5.60	0.34	0.33	0.33	0.04	0.03	0.04	0.21	0.19	0.20	0.59	0.61	0.59
1	0.10	3.08	5.65	0.33	0.33	0.33	0.14	0.12	0.14	0.29	0.28	0.28	0.49	0.50	0.49
1	0.05	1.34	3.11	0.33	0.33	0.33	0.14	0.13	0.14	0.28	0.28	0.28	0.48	0.50	0.49
1	0.01	0.56	0.89	0.33	0.34	0.33	0.14	0.14	0.14	0.28	0.29	0.28	0.49	0.49	0.49
10	0.10	3.10	3.81	0.33	0.33	0.33	0.25	0.25	0.26	0.32	0.32	0.32	0.40	0.41	0.40
10	0.05	1.41	1.38	0.33	0.33	0.33	0.26	0.26	0.26	0.32	0.32	0.32	0.40	0.40	0.40
10	0.01	0.75	0.65	0.33	0.33	0.33	0.26	0.26	0.26	0.33	0.32	0.32	0.40	0.40	0.40

Table 2: Simulation study on $F(1/3)$.

		d_K		Mean			25%			Median			75%		
θ	ϵ	AL1	AL2	AL1	AL2	PY	AL1	AL2	PY	AL1	AL2	PY	AL1	AL2	PY
0	0.10	1.60	3.57	0.50	0.50	0.50	0.36	0.34	0.36	0.50	0.50	0.50	0.64	0.67	0.65
0	0.05	0.94	2.72	0.50	0.50	0.50	0.35	0.34	0.36	0.50	0.50	0.50	0.64	0.66	0.65
0	0.01	1.18	2.10	0.50	0.50	0.50	0.36	0.35	0.36	0.50	0.50	0.50	0.64	0.65	0.65
1	0.10	1.61	3.18	0.50	0.50	0.50	0.40	0.39	0.40	0.50	0.50	0.50	0.60	0.61	0.60
1	0.05	1.28	2.32	0.50	0.50	0.50	0.40	0.40	0.40	0.50	0.50	0.50	0.59	0.60	0.60
1	0.01	1.12	0.57	0.50	0.50	0.50	0.40	0.41	0.40	0.50	0.50	0.50	0.60	0.60	0.60
10	0.10	2.81	4.18	0.50	0.50	0.50	0.45	0.46	0.46	0.50	0.50	0.50	0.55	0.55	0.54
10	0.05	1.78	1.28	0.50	0.50	0.50	0.46	0.46	0.46	0.50	0.50	0.50	0.54	0.54	0.54
10	0.01	2.01	1.09	0.50	0.50	0.50	0.46	0.46	0.46	0.50	0.50	0.50	0.54	0.54	0.54

Table 3: Simulation study on $\mu = \int xP(\mathrm{d}x)$.

Summary statistics for $F(1/3)$ (Table 2) and $\mu = \int xP(\mathrm{d}x)$ (Table 3) using ALGORITHM 1 (AL1) and ALGORITHM 2 (AL2) to sample from the ϵ-PY process. The Kolmogorov distance (d_K) is between the cumulative distribution functions with respect to the Pitman–Yor (PY) process (multiplied by a factor of 100). The parameter values are $\alpha = 0.5$, $\theta \in \{0, 1, 10\}$ and $\epsilon \in \{0.10, 0.05, 0.01\}$.

as the almost sure limit of $n^{-\alpha}K_n$ where K_n denotes the (random) number of unique values in the first n terms of an exchangeable sequence from P in (1). According to Theorem 3.8 in Pitman (2006), $n^{-\alpha}K_n \sim_{a.s.} (T_{\alpha,\theta})^{-\alpha}$, in particular, for $\theta = 0$, $T_\alpha^{-\alpha}$ has a Mittag-Leffler distribution with p-th moment $\Gamma(p+1)/\Gamma(p\alpha+1)$, $p > -1$. According to Pitman (2006, Lemma 3.11, eqn (3.36)), the asymptotic distribution of the truncation error R_n can be derived from that of K_n to get $R_n \sim_{a.s.} \alpha(T_{\alpha,\theta})^{-1}\, n^{1-1/\alpha}$ as $n \to \infty$. The proof relies on Kingman's representation of random partitions (Kingman, 1978) together with techniques set forth by Gnedin et al. (2007). In the proof of Theorem 1 the asymptotic distribution of $T_n = -\log R_n$ is a direct consequence of the above by an application of the continuous mapping theorem.

When $\theta = 0$ it is possible to give an interpretation of the asymptotic distribution of R_n in terms of the jumps of a stable subordinator. In this case the weights of P can

		$\alpha = 0.4$			$\alpha = 0.5$			$\alpha = 0.6$		
θ	ϵ	AL1	AL2	n	AL1	AL2	n	AL1	AL2	n
0	0.10	0.01	0.20	5	0.02	0.04	11	0.11	0.05	38
0	0.05	0.01	0.04	8	0.04	0.04	21	0.20	0.06	105
0	0.01	0.04	0.04	20	0.36	0.05	101	15.10	0.26	1163
1	0.10	0.03	0.19	17	0.07	0.05	31	0.23	0.07	92
1	0.05	0.06	0.06	26	0.13	0.06	61	0.80	0.12	258
1	0.01	0.18	0.09	73	0.80	0.12	301	27.75	0.57	2877
10	0.10	0.22	0.15	121	0.61	0.10	211	2.11	0.18	567
10	0.05	0.45	0.10	191	1.52	0.15	421	9.22	0.37	1603
10	0.01	1.93	0.20	558	13.24	0.48	2101	760.68	4.01	17911

Table 4: Computing time (ms) per **iteration**.

		$\alpha = 0.4$			$\alpha = 0.5$			$\alpha = 0.6$		
θ	ϵ	AL1	AL2	n	AL1	AL2	n	AL1	AL2	n
0	0.10	1.92	38.46	5	1.82	3.64	11	2.91	1.32	38
0	0.05	1.30	5.22	8	1.90	1.90	21	1.91	0.57	105
0	0.01	1.95	1.95	20	3.56	0.50	101	12.98	0.22	1163
1	0.10	1.81	11.45	17	2.26	1.61	31	2.50	0.76	92
1	0.05	2.33	2.33	26	2.13	0.98	61	3.10	0.46	258
1	0.01	2.45	1.23	73	2.66	0.40	301	9.65	0.20	2877
10	0.10	1.82	1.24	121	2.89	0.47	211	3.72	0.32	567
10	0.05	2.35	0.52	191	3.61	0.36	421	5.75	0.23	1603
10	0.01	3.46	0.36	558	6.30	0.23	2101	42.47	0.22	17911

Table 5: Computing time (μs) per **support point**.

Average computing time per iteration (in *milli*second in Table 4) and per support point (in *micro*second in Table 5) for ALGORITHM 1 (AL1) and ALGORITHM 2 (AL2) based on 10^4 iterations, and expected stopping time $n = \mathrm{E}[\tau(\epsilon)]$. The parameter values are $\alpha \in \{0.4, 0.5, 0.6\}$, $\theta \in \{0, 1, 10\}$ and $\epsilon \in \{0.10, 0.05, 0.01\}$.

be represented as the renormalized jumps of a stable subordinator, with T_α denoting the total mass. Denote the (unormalized) jumps as $(J_i)_{i\geq 1}$ in decreasing order and as $(\tilde{J}_i)_{i\geq 1}$ when in size-biased order,

$$T_\alpha = \sum_{i\geq 1} J_i = \sum_{i\geq 1} \tilde{J}_i, \quad \text{and } T_\alpha R_n = \sum_{i>n} \tilde{J}_i.$$

By the asymptotic distribution of R_n, $n^{1/\alpha - 1} \sum_{i>n} \tilde{J}_i \to_{a.s.} \alpha$ as $n \to \infty$. That is, once properly scaled, the small jumps of the stable subordinator (in size-biased random order), interpreted as the "dust", converge to the "proportion" α. This is reminiscent to the number of singletons which is asymptotically ($n \to \infty$) a α proportion of the number of groups in a sample of size n, see Lemma 3.11, eqn (3.39), of Pitman (2006).

4.2 Regenerative random compositions and Anscombe's theorem

We review next the connections of the counting renewal process $N(t)$ defined in (7)–(8) and the theory of regenerative random compositions. The reader is referred to the survey of Gnedin (2010) for a review. Recall that, when $\alpha = 0$ (Dirichlet process case), $V_i \overset{\text{i.i.d.}}{\sim} \text{beta}(1,\theta)$ in the stick-breaking representation (2), and in turns $Y_i = -\log(1-V_i) \overset{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$ and $T_n = -\log R_n \sim \text{Gamma}(n,\theta)$. By direct calculus, $N(t) \sim \text{Pois}(\theta t)$ so that $\tau(\epsilon) - 1 = N(\log 1/\epsilon)$ has $\text{Pois}(\theta \log 1/\epsilon)$ distribution. More generally, the stick-breaking frequencies $(p_i)_{i\geq 1}$ correspond to the gaps in $[0,1]$ identified by the *multiplicative regenerative* set $\mathcal{R} \subset (0,1)$ consisting of the random partial sums $1 - R_k = \sum_{i\leq k} p_i$. The complement open set $\mathcal{R}^c = (0,1)/\mathcal{R}$ can be represented as a disjoint union of countably many open intervals or gaps, $\mathcal{R}^c = \bigcup_{k=0}^{\infty}(1-R_k, 1-R_{k+1})$, $R_0 = 1$. A random composition of the integer n into an ordered sequence $\varkappa_n = (n_1, n_2, \ldots, n_k)$ of positive integers with $\sum_j n_j = n$ can be generated as follows: independently of $\mathcal{R}$, sample $U_1, U_2, \ldots$ from the uniform distribution on $[0,1]$ and group them in clusters by the rule: U_i, U_j belong to the same cluster if they hit the same interval. The random composition of $\varkappa_n$ corresponds then to the record of positive counts in the left-to-right order of the intervals. The composition structure $(\varkappa_n)$ is called regenerative since for all $n > m \geq 1$, conditionally given the first part of $\varkappa_n$ is m, if the part is deleted then the remaining composition of $n-m$ is distributed like $\varkappa_{n-m}$. The regenerative set $\mathcal{R}$ corresponds to the closed range of the multiplicative subordinator $\{1 - \exp(-S_t), t \geq 0\}$, where S_t is the compound Poisson process with Lévy intensity $\tilde{\nu}(\mathrm{d}y) = \theta \mathrm{e}^{-\theta y}\mathrm{d}y$. Since the range of S_t is a homogeneous Poisson point process on $\mathbb{R}_+$ with rate θ, $\mathcal{R}$ is an inhomogeneous Poisson point process $\mathcal{N}(\mathrm{d}x)$ on $[0,1]$ with Lévy intensity $\nu(\mathrm{d}x) = \theta/(1-x)\mathrm{d}x$ so that, for $t = \log 1/\epsilon$,

$$N(\log 1/\epsilon) = \mathcal{N}[0, 1-\epsilon] \sim \text{Pois}(\lambda), \quad \lambda = \int_0^{1-\epsilon} \frac{\theta}{1-x}\mathrm{d}x = \theta \log 1/\epsilon$$

as expected. Suppose now that $(V_i)_{i\geq 1}$ are independent copies of some random variable V on $[0,1]$, not necessarily $\text{beta}(1,\theta)$ distributed. The corresponding random composition structure has been studied in Gnedin (2004); Gnedin et al. (2009) as the outcome of a *Bernoulli sieve* procedure. We recall here the relevant asymptotic analysis. Let $\mu = \mathrm{E}(-\log(1-V))$ and $\sigma^2 = \mathrm{Var}(-\log(1-V))$, equal respectively to $1/\theta$ and $1/\theta^2$ in the DP case, respectively. If those moments are finite, by the CLT,

$$\frac{T_n - n\mu}{\sqrt{n}\sigma} \to_d Z, \quad \text{as } n \to \infty,$$

where $Z \sim \mathrm{N}(0,1)$, and, by means of Anscobe's Theorem, one obtains that

$$\frac{N(t) - t/\mu}{\sqrt{\sigma^2 t/\mu^3}} \to_d Z, \quad \text{as } t \to \infty.$$

It turns out that the normal limit of $N(\log n)$ corresponds to the normal limit of K_n,

$$\frac{K_n - \log n/\mu}{\sqrt{\sigma^2 \log n/\mu^3}} \to_d Z, \quad \text{as } n \to \infty$$

provided that $\mathrm{E}(-\log V) < \infty$. To see why, consider iid random variables $X_1, X_2, \ldots$ with values in $\mathbb{N}$ such that $\{X_i = k\} = \{U_i \in (1-R_{k-1}, 1-R_k)\}$. Hence $\mathrm{P}(X_1 = k|\mathcal{R}) = p_k$. We then have that $K_n = \#\{k : X_i = k \text{ for at least one } i \text{ among } 1, \ldots, n\}$. Define $M_n = \max\{X_1, \ldots, X_n\}$. For $U_{1,n} \leq U_{2,n} \leq \ldots \leq U_{n,n}$ denoting the order statistics corresponding to the uniform variates $U_1, \ldots, U_n$, we have $M_n = \min\{j : \ 1 - R_j \geq U_{n,n}\} = \min\{j : \ T_j \geq E_{n,n}\}$ upon transformation $x \to -\log(1-x)$, where $E_{n,n}$ is the maximum of an iid sample of size n from the standard exponential distribution. Since $N(t) = \max\{n : \ T_n \leq t\} = \min\{n : \ T_n \geq t\} - 1$ we have $M_n - 1 = N(E_{n,n})$. Gnedin et al. (2009) proves the equivalence

$$\frac{M_n - b_n}{a_n} \to_d X \quad \Longleftrightarrow \quad \frac{N(\log n) - b_n}{a_n} \to_d X,$$

where X is a random variable with a proper and non degenerate distribution with $a_n > 0$, $a_n \to \infty$ and $b_n \in \mathbb{R}$. A key fact exploited in the proof is that, from extreme-value theory, $E_{n,n} - \log n$ has an asymptotic distribution of Gumbel type. That M_n can be replaced by K_n in the equivalence relation above follows from the fact that $M_n - K_n$, the number of integers $k < M_n$ not appearing in the sample $X_1, \ldots, X_n$, is bounded in probability when $\mathrm{E}(-\log V) < \infty$, see Proposition 5.1 in Gnedin et al. (2009).

Back to the Pitman–Yor process case, by Theorem 1 we have $n^{-\alpha/(1-\alpha)} N(\log n) \to_d (T_{\alpha,\theta}/\alpha)^{-\alpha/(1-\alpha)}$ while $n^{-\alpha} K_n \to_{a.s.} (T_{\alpha,\theta})^{-\alpha}$. So we see that $N(\log n)$ and K_n do not have the same asymptotic behavior as in the $\alpha = 0$ case. By using the fact that

$$\mathrm{P}(X_1 > n|(p_i)) = R_n, \quad R_n \sim_{a.s.} \alpha n^{-(1-\alpha)/\alpha} T_{\alpha,\theta}^{-1},$$

and the fact that, conditional on $(p_i)_{i \geq 1}$, M_n belongs to the domain of attraction of Fréchet distribution, Pitman and Yakubovich (2017, Theorem 6.1) establishes that

$$\mathrm{P}(M_n \leq x n^{\alpha/(1-\alpha)}) \to \mathrm{E}\big[\exp\big(-\alpha T_{\alpha,\theta}^{-1} x^{-(1-\alpha)/\alpha}\big)\big]$$

so we see that $N(\log n)$ and M_n do not have the same asymptotic behavior as in the $\alpha = 0$ case, although they share the same growth rate $n^{\alpha/(1-\alpha)}$. Finally, the non correspondence of the asymptotic distribution of M_n and K_n suggests that the behavior of $M_n - K_n$ is radically different with respect to the $\alpha = 0$ case.

5 Discussion

In this paper we have studied stochastic approximations of the Pitman–Yor process consisting in the truncation of the sequence of stick-breaking frequencies at a random stopping time $\tau(\epsilon)$ that controls the accuracy of the approximation in the total variation distance by ϵ. We name this finite dimensional approximation the ϵ-Pitman–Yor process. We have derived the asymptotic distribution of $\tau(\epsilon)$ as ϵ goes to zero and we have advanced its use to devise a sampling scheme that generates the stopping time first, and then the frequencies up to that point. The simulations in Section 3 show that the proposed sampler proves computationally very efficient in the moderate to large stopping time regime (for approximately $\tau(\epsilon) \geq 20$). The asymptotic distribution illustrates how large the stopping time is as the approximation error gets small in terms of the prior parameters θ and α. In particular, it shows that the distribution of $\tau(\epsilon)$ in the Dirichlet process case is not recovered in the limit $\alpha \to 0$ in Theorem 2. In fact, in the Dirichlet

process case $\tau(\epsilon)$ grows at a logarithmic rate in $1/\epsilon$ while in Pitman–Yor case it grows at the polynomial rate $\epsilon^{\alpha/(1-\alpha)}$ and the first regime is not recovered by letting α approach 0 in the second regime. We have also drawn important connections with the theory of random partition structures developed by Jim Pitman and coauthors which highlight the relationship of the stopping time $\tau(\epsilon)$ with the number K_n of unique values in a sample of size n from the Pitman–Yor process.

We have left as open problem for future research the study of the conditional distribution of the stick-breaking frequencies given the stopping time. In the Dirichlet process case one can exploit the renewal process interpretation to generate exactly from this conditional distribution. In fact, when $\alpha = 0$, the sequence $(-\log R_n)_{n\geq 1}$ corresponds to the jump times of a Poisson process and the conditional distribution of the jumps given the number of jumps at time t can be described in terms of the ordered statistics of i.i.d. uniform random variates on $(0, t)$. The case $\alpha > 0$ does not seem to be easily tractable, as it would be if the counting process associated to $\tau(\epsilon)$ were a mixed sample process or, equivalently, a Cox process, cf. Grandell (1997, Section 6.3).

It would be also interesting to compare the accuracy of our finite dimensional approximation of the Pitman–Yor process to the one proposed in Al Labadi and Zarepour (2014). The latter is based on a representation of the frequencies in decreasing order, cf. Pitman and Yor (1997, Proposition 22). Al Labadi and Zarepour (2014) compare the accuracy of their approximation scheme to a stick-breaking truncation at a number n of stick-breaking frequencies that matches the number of frequencies used in their scheme. Not surprisingly, their approximation is superior since it generates weights in decreasing order, specially when α is large. In contrast, Theorem 2 describes precisely how large the truncation threshold n should be as α gets large for a given approximation level ϵ, cf. the center panel of Figure 1. It also underlines that the approximation deteriorates for fixed n and increasing α, which is coherent with the findings in Al Labadi and Zarepour (2014). A fair comparison with their approach can only be done for a given nominal approximation error, but unfortunately the authors did not provide a precise assessment of it. The number of stick-breaking frequencies needed to match the approximation accuracy of Al Labadi and Zarepour (2014) would be *de facto* larger due to the non monotonicity. However, since the stopping rule (4) adapts to the size of α, we do not expect the accuracy of our approximation scheme to deteriorate for α large. As for computation time, the techniques used by Al Labadi and Zarepour (2014) in order to obtain decreasing frequencies are computational heavy. Their average computing time for $\alpha = 0.5$ is about 2.30 seconds/iteration with 10^4 locations. This amounts to 0.23 *milli*seconds/support point, which is 1000 times slower than the computing time for our Algorithm 2 in the parameter configuration $\alpha = 0.5$, $\theta = 10$ and $\epsilon = 0.01$, equal to 0.23 *micro*second/support point. It would be interesting to investigate what are the consequences in terms of computation time per iteration for a given approximation error.

Supplementary Material

Supplementary Material of "Stochastic Approximations to the Pitman–Yor Process" (DOI: 10.1214/18-BA1127SUPP; .pdf). ALGORITHM 3 for generating from a polynomially tilted positive stable random variable (in a separate document).

References

Al Labadi, L. and Zarepour, M. (2014). "On simulations from the two-parameter Poisson-Dirichlet process and the normalized Inverse-Gaussian process." *Sankhya*, 76-A: 158–176. MR3167777. doi: https://doi.org/10.1007/s13171-013-0033-0. 1203, 1216

Arbel, J., De Blasi, P., and Prünster, I. (2018). "Supplementary Material of "Stochastic Approximations to the Pitman–Yor Process"." *Bayesian Analysis*. doi: https://doi.org/10.1214/18-BA1127SUPP. 1203

Bassetti, F., Casarin, R., and Leisen, F. (2014). "Beta-product dependent Pitman–Yor processes for Bayesian inference." *Journal of Econometrics*, 180(1): 49–72. MR3188911. doi: https://doi.org/10.1016/j.jeconom.2014.01.007. 1201

Canale, A., Lijoi, A., Nipoti, B., and Prünster, I. (2017). "On the Pitman-Yor process with spike and slab base measure." *Biometrika*, 104: 681–697. MR3694590. doi: https://doi.org/10.1093/biomet/asx041. 1201

Caron, F., Neiswanger, W., Wood, F., Doucet, A., and Davy, M. (2017). "Generalized Pólya Urn for Time-Varying Pitman-Yor Processes." *Journal of Machine Learning Research*, 18(27): 1–32. MR3634894. 1201

Devroye, L. (2009). "Random variate generation for exponentially and polynomially tilted stable distributions." *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 19(4): Article No. 18. 1203, 1206

Favaro, S., Lijoi, A., Mena, R., and Prünster, I. (2009). "Bayesian non-parametric inference for species variety with a two-parameter Poisson–Dirichlet process prior." *Journal of the Royal Statistical Society. Series B*, 71: 993–1008. MR2750254. doi: https://doi.org/10.1111/j.1467-9868.2009.00717.x. 1201

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *Annals of Statistics*, 1: 209–230. MR0350949. 1202

Gelfand, A. and Kottas, A. (2002). "A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models." *Journal of Computational and Graphical Statistics*, 11: 289–305. MR1938136. doi: https://doi.org/10.1198/106186002760180518. 1202

Ghosal, S. and van der Vaart, A. W. (2017). *Foundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. MR3587782. doi: https://doi.org/10.1017/9781139029834. 1202

Gnedin, A. (2004). "The Bernoulli sieve." *Bernoulli*, 10: 79–96. MR2044594. doi: https://doi.org/10.3150/bj/1077544604. 1214

Gnedin, A. (2010). "Regeneration in random combinatorial structures." *Probability Surveys*, 7: 105–156. MR2684164. doi: https://doi.org/10.1214/10-PS163. 1214

Gnedin, A., Hansen, B., and Pitman, J. (2007). "Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws." *Probability Surveys*, 4: 146–171. MR2318403. doi: https://doi.org/10.1214/07-PS092. 1212

Gnedin, A., Iksanov, A. M., Pavlo, N., and Uwe, R. (2009). "The Bernoulli sieve revisited." *The Annals of Applied Probability*, 19: 1634–1655. MR2538083. doi: https://doi.org/10.1214/08-AAP592. 1214, 1215

Grandell, J. (1997). *Mixed Poisson Processes*. Monographs on Statistics and Applied Probability. Springer US. MR1463943. doi: https://doi.org/10.1007/978-1-4899-3117-7. 1216

Gut, A. (2013). *Probability: a graduate course*. Springer texts in statistics. Springer, 2nd ed edition. MR2977961. doi: https://doi.org/10.1007/978-1-4614-4708-5. 1204, 1205

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96: 161–173. MR1952729. doi: https://doi.org/10.1198/016214501750332758. 1201, 1202, 1208

Ishwaran, H. and Zarepour, M. (2002). "Exact and approximate sum representations for the Dirichlet Process." *Canadian Journal of Statistics*, 30: 269–283. MR1926065. doi: https://doi.org/10.2307/3315951. 1202

James, L. F., Lijoi, A., and Prünster, I. (2010). "On the posterior distribution of classes of random means." *Bernoulli*, 16(1): 155–180. MR2648753. doi: https://doi.org/10.3150/09-BEJ200. 1210

Jara, A. (2007). "Applied Bayesian non- and semi-parametric inference using DPpackage." *Rnews*, 7: 17–26. 1203, 1208

Jara, A., Hanson, T. E., Quintana, F. A., Müller, P., and Rosner, G. L. (2011). "DPpackage: Bayesian Semi- and Nonparametric Modeling in R." *Journal of Statistical Software*, 40(5): 1–30. MR3309338. doi: https://doi.org/10.1007/978-3-319-18968-0. 1203, 1208

Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. (2010). "Bayesian semiparametric inference for multivariate doubly-interval-censored data." *Annals of Applied Statistics*, 4(4): 2126–2149. MR2829950. doi: https://doi.org/10.1214/10-AOAS368. 1201

Kingman, J. F. C. (1978). "The representation of partition structures." *Journal of the London Mathematical Society*, 18: 374–380. MR0509954. doi: https://doi.org/10.1112/jlms/s2-18.2.374. 1212

Lijoi, A., Mena, R., and Prünster, I. (2007). "Bayesian nonparametric estimation of the probability of discovering a new species." *Biometrika*, 94: 769–786. MR2416792. doi: https://doi.org/10.1093/biomet/asm061. 1201

Muliere, P. and Tardella, L. (1998). "Approximating distributions of random functionals of Ferguson-Dirichlet priors." *The Canadian Journal of Statistics*, 26(2): 283–297. MR1648431. doi: https://doi.org/10.2307/3315511. 1203

Navarrete, C., Quintana, F. A., and Mueller, P. (2008). "Some issues in nonparametric Bayesian modeling using species sampling models." *Statistical Modelling*, 8(1): 3–21. MR2750628. doi: https://doi.org/10.1177/1471082X0700800102. 1201

Ni, Y., Müller, P., Zhu, Y., and Ji, Y. (2018). "Heterogeneous reciprocal graphical models." *Biometrics*, 74(2): 606–615. MR3825347. 1201

Perman, M., Pitman, J., and Yor, M. (1992). "Size-biased sampling of Poisson point processes and excursions." *Probability Theory and Related Fields*, 92(1): 21–39. MR1156448. doi: https://doi.org/10.1007/BF01205234. 1201

Pitman, J. (1995). "Exchangeable and partially exchangeable random partitions." *Probability Theory and Related Fields*, 102(2): 145–158. MR1337249. doi: https://doi.org/10.1007/BF01213386. 1201

Pitman, J. (1996). "Some developments of the Blackwell-MacQueen urn scheme." In *Statistics, probability and game theory*, volume 30 of *IMS Lecture Notes Monogr. Ser.*, 245–267. Inst. Math. Statist., Hayward, CA. MR1481784. doi: https://doi.org/10.1214/lnms/1215453576. 1208

Pitman, J. (2006). *Combinatorial stochastic processes*. Ecole d'Eté de Probabilités de Saint-Flour XXXII. Lecture Notes in Mathematics N. 1875. Springer, New York. 1203, 1205, 1212, 1213

Pitman, J. and Yakubovich, Y. (2017). "Extremes and gaps in sampling from a GEM random discrete distribution." *Electronic Journal of Probability*, 22. MR3646070. doi: https://doi.org/10.1214/17-EJP59. 1215

Pitman, J. and Yor, M. (1997). "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator." *The Annals of Probability*, 25(2): 855–900. MR1434129. doi: https://doi.org/10.1214/aop/1024404422. 1201, 1202, 1216

Sudderth, E. B. and Jordan, M. I. (2009). "Shared segmentation of natural scenes using dependent Pitman-Yor processes." In *Advances in Neural Information Processing Systems 21*, 1585–1592. Curran Associates, Inc. 1201

Teh, Y. W. (2006). "A hierarchical Bayesian language model based on Pitman-Yor processes." In *Proc. Coling/ACL*, 985–992. Stroudsburg, PA, USA. 1201

Tricomi, F. G. and Erdélyi, A. (1951). "The asymptotic expansion of a ratio of gamma functions." *Pacific Journal of Mathematics*, 1(1): 133–142. MR0043948. 1205

Acknowledgments

The authors are grateful to an associate editor and a referee for helpful comments and suggestions. P. De Blasi and I. Prünster are supported by MIUR, PRIN Project 2015SNS29B.