On Simulations from the Two-Parameter Poisson-Dirichlet Process and the Normalized Inverse-Gaussian Process

Luai Al Labadi
UAE University, Al Ain, UAE

Mahmoud Zarepour
University of Ottawa, Ottawa, Canada

Abstract

This paper develops simple and efficient procedures for sampling approximations of the two-Parameter Poisson-Dirichlet Process and the normalized inverse-Gaussian process. We compare the efficiency of the new approximations with the corresponding stick-breaking approximations, in which we demonstrate a substantial improvement.

AMS (2000) subject classification. Primary 65C60; Secondary 62F15.
Keywords and phrases. Dirichlet process, nonparametric Bayesian inference, normalized inverse-Gaussian process, simulation, normalized stable law process, stick-breaking representation, two-parameter Poisson-Dirichlet process.

1 Introduction

One of the key components in a Bayesian nonparametric analysis is the elucidation of a prior on the space of probability measures. The Dirichlet process, formally introduced in Ferguson (1973), is one of the most widely used priors for this purpose. Ferguson’s (1973) original definition of the Dirichlet process was based on specifying its finite-dimensional marginals to be Dirichlet distributions. An alternative definition of the Dirichlet process, due to Ferguson (1973), was based on normalizing the gamma process. A different constructive definition of the Dirichlet process was given by Sethuraman (1994) using a “stick-breaking” approach. We refer the reader to the paper of Zarepour and Al Labadi (2012) for more discussion about different representations of the Dirichlet process.

Several alternatives for the Dirichlet process have been proposed in the literature. In this paper, we focus on two such priors, namely the two-parameter Poisson-Dirichlet process (Pitman and Yor, 1997) and the normalized inverse-Gaussian process (Lijoi, Mena and Prünter, 2005). We begin
by introducing the stick-breaking definition of the two-parameter Poisson-
Dirichlet process defined on an arbitrary measurable space \((X, A)\). For more
details, see Pitman and Yor (1997).

**Definition 1.1.** For \(0 \leq \alpha < 1, \theta > -\alpha\), let \((\beta_i)_{i \geq 1}\) be a sequence of
independent random variables, where \(\beta_i\) has the \(\text{beta}(1-\alpha, \theta+i\alpha)\) distribution. Define
\[
p_1' = \beta_1, \ p_i' = \beta_i \prod_{k=1}^{i-1} (1-\beta_k), \ i \geq 2.
\]

Let \(p_1 \geq p_2 \geq \ldots\) be the ranked values of \((p_i')_{i \geq 1}\). Moreover, let \((Y_i)_{i \geq 1}\)
be a sequence of independent and identically distributed (i.i.d.) random
variables with common distribution \(H\), independent of \((\beta_i)_{i \geq 1}\). Then the random probability measure
\[
P_{H,\alpha,\theta}(\cdot) = \sum_{i=1}^{\infty} p_i \delta_{Y_i}(\cdot)
\]
is called a two-parameter Poisson-Dirichlet process on \((X, A)\) with parameters \(\alpha, \theta\) and \(H\), where \(\delta_x\) denotes the dirac measure at \(x\).

It is worth mentioning that Ishwaran and James (2001) referred to the process
\(P'_{H,\alpha,\theta}(\cdot) = \sum_{i=1}^{\infty} p_i' \delta_{Y_i}(\cdot)\) as the Pitman-Yor process, where \((p_i')_{i \geq 1}\)
and \((Y_i)_{i \geq 1}\) are as defined in Definition 1.1. The two-parameter Poisson-Dirichlet process with parameters \(\alpha, \theta\) and \(H\) is denoted by \(PDP(H; \alpha, \theta)\), and we write \(P_{H,\alpha,\theta} \sim PDP(H; \alpha, \theta)\). In the literature, the probability measure \(H\) is called the base measure of \(P_{H,\alpha,\theta}\), while the parameters \(\alpha\) and \(\theta\) are called the discount parameter and the concentration parameter, respectively (Buntine and Hutter, 2010; Teh, 2006). Representation (1.1) clearly shows that any realization of the two-parameter Poisson-Dirichlet process must be a discrete probability measure. Note that, the special case \(PDP(H; 0, \theta)\) represents the Dirichlet process. The law of the weights \((p_1, p_2, \ldots)\) is called the two-parameter Poisson-Dirichlet distribution, denoted by \(PD(\alpha, \theta)\). The two-parameter Poisson-Dirichlet distribution has many applications in different fields such as population genetics, ecology, statistical physics and number theory. See Feng (2010) for more details. On the other hand, the two-parameter Poisson-Dirichlet process has been recently used in applications in Bayesian nonparametric statistics such as computer science (Teh, 2006), species sampling (Jang, Lee and Lee, 2010; Navarrete, Quintana and Müller, 2008) and genomics (Favaro et al., 2009).
The calculations of the moments for the two-parameter Poisson-Dirichlet process are carried out in Carlton (1999). Let $A$ be a measurable subset of $\mathcal{X}$. Then

$$E(P_{H,\alpha,\theta}(A)) = H(A) \quad \text{and} \quad \text{Var}(P_{H,\alpha,\theta}(A)) = H(A)(1 - H(A))\frac{1 - \alpha}{1 + \theta}. \quad (1.2)$$

It follows from (1.2) that the base measure $H$ plays the role of the center of the process, while both $\alpha$ and $\theta$ control the variability of $P_{H,\alpha,\theta}$ around $H$. Observe that, for any fixed set $A \in \mathcal{A}$ and $\epsilon > 0$, by Chebyshev’s inequality we have

$$\Pr \{|P_{H,\alpha,\theta}(A) - H(A)| > \epsilon\} \leq H(A)(1 - H(A)) \frac{1 - \alpha}{(1 + \theta)\epsilon^2}. \quad (1.3)$$

Thus, $P_{H,\alpha,\theta}(A) \overset{p}{\to} H(A)$ as $\theta \to \infty$ (for a fixed $\alpha$) or as $\alpha \to 1$, $\alpha < 1$ (for fixed $\theta$). In this paper, “$\overset{p}{\to}$” and “$\overset{a.s.}{\to}$” denote convergence in probability and almost sure convergence, respectively.

Analogous to the Dirichlet process, Lijoi et al. (2005) defined the normalized inverse-Gaussian process $P_{H,\theta} = \{P_{H,\theta}(A)\}_{A \in \mathcal{A}}$ with parameters $\theta$ and $H$ by specifying the distribution of $(P_{H,\theta}(A_1), \ldots, P_{H,\theta}(A_m))$, for a partition $A_1, \ldots, A_m$ of $\mathcal{X}$, as the normalized inverse-Gaussian distribution with parameter $(\theta H(A_1), \ldots, \theta H(A_m))$, where $m \geq 2$. See equation (3) of Lijoi et al. (2005) for the density of the normalized inverse-Gaussian distribution. The normalized inverse-Gaussian process with parameters $\theta$ and $H$ is denoted by $\text{N-IGP}(H; \theta)$, and we write $P_{H,\theta} \sim \text{N-IGP}(H; \theta)$.

One of the basic properties of the normalized inverse-Gaussian process is that for any $A \in \mathcal{A}$,

$$E(P_{H,\theta}(A)) = H(A) \quad \text{and} \quad \text{Var}(P_{H,\theta}(A)) = \frac{H(A)(1 - H(A))}{\xi(\theta)}, \quad (1.4)$$

where $\xi(\theta) = (\theta^2 e^\theta \Gamma(-2, \theta))^{-1}$ and $\Gamma(-2, \theta) = \int_0^\infty t^{-3} e^{-t} dt$. Observe that, for large $\theta$, $\xi(\theta) \approx \theta$ (Abramowitz and Stegun, 1972, Formula 6.5.32, page 263), where we use the notation $f(\theta) \approx g(\theta)$ if $\lim_{\theta \to \infty} f(\theta)/g(\theta) = 1$. Like the two-parameter Poisson-Dirichlet process, it follows from (1.4) that $H$ plays the role of the center of the process, while $\theta$ can be viewed as the concentration parameter. The larger $\theta$ is, the more likely it is that the
realization of $P_{H,\theta}$ is close to $H$. Specifically, for any fixed set $A \in \mathcal{A}$ and $\epsilon > 0$, we have $P_{H,\theta}(A) \xrightarrow{P} H(A)$ as $\theta \to \infty$ since

$$\Pr \{|P_{H,\theta}(A) - H(A)| > \epsilon\} \leq \frac{H(A)(1 - H(A))}{\xi(\theta)\epsilon^2}.$$

(1.5)

We refer the reader to the recent paper of Al Labadi and Zarepour (2013) for several interesting asymptotic properties of the normalized inverse-Gaussian process.

Similar to the Dirichlet process, a series representation of the normalized inverse-Gaussian process can be easily derived from the representation of Ferguson and Klass (1972). Specifically, let $(E_i)_{i \geq 1}$ be a sequence of i.i.d. random variables with an exponential distribution with mean of 1. Define

$$\Gamma_i = E_1 + \cdots + E_i.$$

(1.6)

Let $(Y_i)_{i \geq 1}$ be a sequence of i.i.d. random variables with values in $\mathcal{X}$ and common distribution $H$, independent of $(\Gamma_i)_{i \geq 1}$. Then, by using the representation of Ferguson and Klass (1972), the normalized inverse-Gaussian process with parameter $\theta$ and $H$ takes the form:

$$P_{H,\theta}(:) = \sum_{i=1}^{\infty} \frac{L^{-1}(\Gamma_i)}{\sum_{k=1}^{\infty} L^{-1}(\Gamma_k)} \delta_{Y_i}(:),$$

(1.7)

where

$$L(x) = \frac{\theta}{\sqrt{2\pi}} \int_x^{\infty} e^{-t/2} t^{-3/2} dt, \text{ for } x > 0.$$

(1.8)

See also Nieto-Barajas and Prünster (2009) for further discussion. Observe that, working with (1.7) is relatively difficult in practice because no closed form for the inverse of the Lévy measure (1.8) exists. Moreover, to determine the random weights in (1.7) an infinite sum must be computed. A remedy of such complexity will be given in Section 4, where an almost sure approximation to (1.7) is given.

Using a stick-breaking approach, a radically different constructive definition of the normalized inverse-Gaussian process was recently established by Favaro, Lijoi and Prünster (2012). Let $(Z_i)_{i \geq 1}$ be i.i.d. random variables
where $Z_i$ is $1/2$-stable random variable with scale parameter $1$. Define a sequence of dependent random variables $(V_i)_{i \geq 1}$ as follows

$$V_1 = \frac{X_1}{X_1 + Z_1} \text{ such that } X_1 \sim \text{GIG} \left( \theta^2, 1, -\frac{1}{2} \right),$$

$$(V_1|V_1, \ldots, V_{i-1}) = \frac{X_i}{X_i + Z_i} \text{ such that }$$

$$X_i \sim \text{GIG} \left( \frac{\theta^2}{\prod_{j=1}^{i-1}(1 - V_j)}, 1, -\frac{i}{2} \right), \quad i \geq 2,$$

(1.9)

where the sequences $(X_i)_{i \geq 1}$ and $(Z_i)_{i \geq 1}$ are independent and GIG denotes the generalized inverse-Gaussian distribution (see equation (2) of Favaro et al., 2012). Define

$$p_1 = V_1, \quad p_j = V_j \prod_{i=1}^{j-1}(1 - V_i), \quad j \geq 2.$$  

Moreover, let $(Y_i)_{i \geq 1}$ be a sequence of i.i.d. random variables with common distribution $H$, independent of $(V_i)_{i \geq 1}$. Define

$$P_{H,\theta}(\cdot) = \sum_{i=1}^{\infty} p_i \delta_{Y_i}(\cdot).$$

(1.10)

Then $P_{H,\theta}$ is a normalized inverse-Gaussian process with parameters $\theta$ and $H$. Note that, in (1.9), having $1 - V_i$ in the denominator leads to numerical instability. Specifically, the values of the random variables $(V_i)_{i \geq 1}$ in (1.9) become numerically 1 after a few simulation steps (see Table 4). This is due to the fact that the values of the random variables $(X_i)_{i \geq 1}$ become very large compared to the values of the random variables $(Z_i)_{i \geq 1}$. In this case, generating samples for the random variables $(X_i)_{i \geq 1}$ from the generalized inverse-Gaussian distributions is impossible as these distributions become undefined (one of the parameters becomes numerically 0). This problem makes sampling from the normalized inverse-Gaussian process via the stick-breaking representation impossible in most cases. More discussion and an alternative approach to this problem will appear in Sections 4 and 5 of this paper.

The remainder of the paper is organized as follows. In Section 2, we use the stick-breaking representations (1.1) and (1.10) to sample approximations of the two-parameter Poisson-Dirichlet Process and the normalized
inverse-Gaussian process, respectively. We also show in this section that the stick-breaking representations are inefficient for simulation purposes. In Sections 3 and 4, we develop simple, yet efficient, algorithms to simulate approximations of the two-parameter Poisson-Dirichlet Process and the normalized inverse-Gaussian process, respectively. In Section 5, an extensive simulation study is carried out to compare the accuracy of the new methods with the stick-breaking approximations. The simulation results clearly show that the new approximations are more efficient.

2 Simulation from Stick-Breaking Representations in Nonparametric Bayesian Inference

Stick-breaking representations are of special interest in Bayesian nonparametric inference. For example, they are used in modeling Bayesian hierarchical mixture models (Ishwaran and James, 2001; Kottas and Gelfand, 2001). They are also used to compute the moments and some theoretical properties of the related priors (Carlton, 1999). In this section, we focus on stick-breaking representations of the two-parameter Poisson-Dirichlet process and the normalized inverse-Gaussian process from a computational point of view.

The stick-breaking representation of the two-parameter Poisson-Dirichlet process (see Definition 1.1) can be used to simulate the approximated two-parameter Poisson-Dirichlet process using a truncation argument. By truncating the higher order terms in the sum (1.1), we can approximate the stick breaking representation by

\[ P_{n,H,\alpha,\theta}(\cdot) = \sum_{k=1}^{n} p_i \delta_{Y_i}(\cdot), \quad (2.1) \]

where \((\beta_i)_{i \geq 1}, (p_i)_{i \geq 1}\), and \((\alpha_i)_{i \geq 1}\) are as given by Definition 1.1 with \(\beta_n = 1\) (hence \(\beta_n\) does not have a beta distribution). The assumption that \(\beta_n = 1\) is necessary to ensure \(\sum_{i=1}^{n} p_i = 1\), almost surely (Ishwaran and James, 2001). A random stopping rule for choosing \(n = n(\epsilon)\), where \(\epsilon \in (0, 1)\), is:

\[ n = \inf\{i : p_i' = (1 - \beta_1) \ldots (1 - \beta_{i-1})\beta_i < \epsilon\}. \quad (2.2) \]

The random stopping rule in (2.2) is similar to the one proposed by Muliere and Tradella (1998) for the Dirichlet process. The following lemma shows that the weights \((p_i)_{i \geq 1}\) in the stick-breaking representation are not strictly decreasing, almost surely. This makes the truncated stick-breaking representation inefficient for simulation purposes.
Lemma 2.1. Let \((p'_{i})_{i \geq 1}\) be as in Definition 1.1. Then \(\Pr\{p'_{i+1} < p'_i\} = \int_{0}^{1} \int_{0}^{y} f(x,y) \, dx \, dy\), where

\[
f(x,y) = \frac{x^{\alpha_1 - 1}(1+x)^{-\alpha_1 - \beta_1}}{B(\alpha_1, \beta_1)} \times \frac{y^{\alpha_1 - 1}(1-y)^{\beta_2 - 1}}{B(\alpha_1, \beta_2)} I\{x \geq 0\} I\{0 < y < 1\},
\]

\(B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)\), \(\alpha_1 = 1 - \alpha\), \(\beta_1 = \theta + i\alpha\) and \(\beta_2 = \theta + (1+i)\alpha\).

Proof. Since \(p'_i = \beta_i \prod_{k=1}^{i-1} (1 - \beta_k)\), we have

\[
\Pr\{p'_{i+1} < p'_i\} = \Pr\left\{\frac{\beta_{i+1} (1 - \beta_i)}{\beta_i} < 1\right\}.
\]

Since \(\beta_i\) is a random variable with the beta\((1 - \alpha, \theta + i\alpha)\) distribution, it follows that \(\beta_i/(1 - \beta_i)\) has the beta distribution of the second kind with parameters \(\alpha_1 = 1 - \alpha\) and \(\beta_1 = \theta + i\alpha\) (Balakrishnan and Lai, 2009, page 12). That is, \(\beta_i/(1 - \beta_i)\) has the density

\[
f(x) = \frac{x^{\alpha_1 - 1}(1+x)^{-\alpha_1 - \beta_1}}{B(\alpha_1, \beta_1)} I\{x \geq 0\}.
\]

The lemma follows from the fact that \((\beta_i)_{i \geq 1}\) is a sequence of independent random variables, where \(\beta_i\) has the beta\((1 - \alpha, \theta + i\alpha)\) distribution.

Note that, the probability \(\Pr\{p'_{i+1} < p'_i\}\) given in Lemma 2.1 depends on \(i\), \(\alpha\) and \(\theta\). Table 1 depicts some values for this probability.

It follows from Table 1 that in the random stopping rule (2.1) we typically have \(p'_{k} = (1 - \beta_1) \ldots (1 - \beta_{k-1})\beta_k > \epsilon\), for some \(k > n\). This makes the stopping rule (2.1) to fail in most cases. To mitigate the effect of non-decreasing weights of the stick-breaking representation, one may apply the following stopping rule:

\[
n = \inf \left\{ i : 1 - \sum_{j=1}^{i-1} p'_j < \epsilon \right\}. \tag{2.3}
\]

Note that, the stopping rule (2.3) yields a large \(n\). In addition, we still need to assume that \(\beta_n = 1\), which changes the structure of the weights in Definition 1.1.

Similar to the two-parameter Poisson-Dirichlet process, the stick-breaking representation of the normalized inverse-Gaussian process can be used to approximately simulate the normalized inverse-Gaussian process using a
truncation argument. By truncating the higher order terms in the sum (1.10), we can approximate the stick breaking representation by

\[ P_{n,H,\theta}(\cdot) = \sum_{k=1}^{n} p_i \delta_{Y_i}(\cdot), \quad (2.4) \]

where \((V_i)_{i \geq 1}, (p_i)_{i \geq 1}\) are as given by Definition 1.1 and we set \(V_n|V_1, \ldots, V_{n-1} = 1\) to ensure that \(P_{n,H,\theta}\) is well defined. A random stopping rule for choosing \(n = n(\epsilon)\), where \(\epsilon \in (0,1)\), is similar to (2.2) or (2.3) with \(\beta_i\) replaced by \(V_i\). Note that, since the random variables \((V_i)_{i \geq 1}\) are not independent, finding the joint density of \((V_i, V_{i+1})\) is relatively complex. Thus, establishing a lemma similar to Lemma 2.1 for the normalized inverse-Gaussian process is not straightforward and may require an excursion into copula theory, which would take us beyond the scope of the paper. Alternatively, approximate values of \(\Pr\{p_{i+1} < p_i\}\) can be obtained by simulating 500 values of each weight with \(\theta = 1\) and \(n = 50\) in (2.4). Some of these values are given in Table 2. It follows from Table 2 that the weights in the stick-breaking representation (1.10) are not strictly decreasing.

| \(i\) | \(\alpha\) | \(\theta\) | \(\Pr\{p_{i+1}^\prime < p_i^\prime\}\) |
|------|------|------|----------------|
| 1    | 0.1  | 1    | 0.672          |
| 10   | 0.1  | 1    | 0.607          |
| 100  | 0.1  | 1    | 0.521          |
| 1    | 0.5  | 1    | 0.598          |
| 10   | 0.5  | 1    | 0.526          |
| 100  | 0.5  | 1    | 0.503          |
| 1    | 0.9  | 1    | 0.523          |
| 10   | 0.9  | 1    | 0.504          |
| 100  | 0.9  | 1    | 0.500          |
| 1    | 0.1  | 10   | 0.523          |
| 10   | 0.1  | 10   | 0.521          |
| 100  | 0.1  | 10   | 0.511          |
| 1    | 0.5  | 10   | 0.515          |
| 10   | 0.5  | 10   | 0.511          |
| 100  | 0.5  | 10   | 0.503          |
| 1    | 0.9  | 10   | 0.504          |
| 10   | 0.9  | 10   | 0.502          |
| 100  | 0.9  | 10   | 0.500          |
Table 2: Some values of $\Pr\{p_{i+1} < p_i\}$.

| $i$ | $\theta$ | $\Pr\{p_{i+1} < p_i\}$ |
|-----|----------|-------------------------|
| 1   | 1        | 0.536                   |
| 10  | 1        | 0.538                   |
| 20  | 1        | 0.540                   |
| 30  | 1        | 0.548                   |
| 40  | 1        | 0.664                   |

3 Simulating an Approximation of the Two-Parameter Poisson-Dirichlet Process

The following proposition provides an interesting approach to construct the two-parameter Poisson-Dirichlet process. This proposition is due to Pitman and Yor (1997).

**Proposition 3.1** (Pitman and Yor, 1997). For $0 < \alpha < 1$ and $\theta > 0$, suppose that $(p_1(0, \theta), p_2(0, \theta), \ldots)$ and $(p_1(\alpha, 0), p_2(\alpha, 0), \ldots)$ have respective distributions $PD(0, \theta)$ and $PD(\alpha, 0)$. Independent of $(p_1(0, \theta), p_2(0, \theta), \ldots)$, let $(p_1^i(\alpha, 0), p_2^i(\alpha, 0), \ldots), i = 1, 2, \ldots$, be a sequence of independent copies of $(p_1(\alpha, 0), p_2(\alpha, 0), \ldots)$. Let $(p_i)_{i \geq 1}$ be the descending order statistics of $\{p_i(0, \theta)p_j^i(\alpha, 0), i, j = 1, 2, \ldots\}$. Then $(p_1, p_2, \ldots)$ has a $PD(\alpha, \theta)$ distribution.

It follows from Proposition 3.1 that the weights in the two-parameter Poisson-Dirichlet process can be constructed based on two boundary selections of the parameters. The first selection is when $\alpha = 0$. This choice of parameters corresponds to the Dirichlet process. The other selection of parameters is when $\theta = 0$, which yields a measure whose random weights are based on a normalized stable law with index $0 < \alpha < 1$. Therefore, the Dirichlet process $P_{H, 0, \theta}$ and the normalized stable law process $P_{H, \alpha, 0}$ are two essential processes in simulating the two-parameter Poisson-Dirichlet process. First we consider simulating these two key processes.

A simple, yet efficient, procedure for approximating the Dirichlet process was recently developed by Zarepour and Al Labadi (2012). Specifically, let $X_n$ be a random variable with distribution Gamma($\theta/n, 1$). Define

$$G_n(x) = \Pr(X_n > x) = \int_x^\infty \frac{1}{\Gamma(\theta/n)} e^{-t \theta/n - 1} dt,$$

and

$$G_n^{-1}(y) = \inf\{x : G_n(x) \geq y\}.$$

Let $(Y_i)_{i \geq 1}$ be a sequence of i.i.d. random variables with values in $X$ and common distribution $H$, independent of $(\Gamma_i)_{i \geq 1}$. Let $\Gamma_i = E_1 + \cdots + E_i$. 


where \((E_i)_{i \geq 1}\) are i.i.d. random variables with exponential distribution of mean 1, independent of \((Y_i)_{i \geq 1}\). Then as \(n \to \infty\),

\[
P_{n,H,0,\theta}^{\text{new}}(\cdot) = \sum_{i=1}^{n} \frac{G^{-1}_n \left( \frac{\Gamma_i}{\Gamma_{n+1}} \right)}{\sum_{k=1}^{n} G^{-1}_n \left( \frac{\Gamma_k}{\Gamma_{n+1}} \right)} \delta_{Y_i}(\cdot) \quad \text{a.s.} \to \sum_{i=1}^{\infty} \frac{N^{-1}(\Gamma_i)}{N^{-1}(\Gamma_k)} \delta_{Y_i}(\cdot),
\]

(3.1)

where \(N(x) = \theta \int_{x}^{\infty} t^{-1} e^{-t} dt, x > 0\). The right-hand side of (3.1) represents Ferguson’s representation (1973) of the Dirichlet process. See also Ishwaran and Zarepour (2009), Theorem 3 and Section 4.1 for further discussion. Zarepour and Al Labadi (2012) showed that the weights \(\frac{G^{-1}_n \left( \frac{\Gamma_i}{\Gamma_{n+1}} \right)}{\sum_{k=1}^{n} G^{-1}_n \left( \frac{\Gamma_k}{\Gamma_{n+1}} \right)}\) in (3.1) decrease monotonically for any fixed positive integer \(n\). They also provided empirical evidence that their new representation yields a highly accurate approximation of the Dirichlet process.

On the other hand, for the normalized stable law process, Pitman and Yor (1997) proved that

\[
P_{H,\alpha,0}(\cdot) = \sum_{i=1}^{\infty} \frac{\Gamma^{-1/\alpha}_i}{\sum_{k=1}^{\infty} \Gamma^{-1/\alpha}_k} \delta_{Y_i}(\cdot),
\]

where \(\Gamma_i = E_1 + \cdots + E_i\) and \((E_i)_{i \geq 1}\) is a sequence of i.i.d. random variables with an exponential distribution with mean of 1. Therefore,

\[
P_{n,H,\alpha,0}(\cdot) = \sum_{i=1}^{n} \frac{\Gamma^{-1/\alpha}_i}{\sum_{k=1}^{n} \Gamma^{-1/\alpha}_k} \delta_{Y_i}(\cdot)
\]

(3.2)

is an almost sure approximation of the normalized stable law process \(P_{H,\alpha,0}\). It is easy to see that the weights \(\frac{\Gamma^{-1/\alpha}_i}{\sum_{k=1}^{n} \Gamma^{-1/\alpha}_k}\) are strictly decreasing. Thus, simulating the normalized stable law process through representation (3.2) is very efficient.

Now we present an efficient algorithm for simulating the two-parameter Poisson-Dirichlet process \(P_{H,\alpha,0}\). This algorithm is based on Proposition 3.1, representation (3.1), and representation (3.2).

**Algorithm A: Simulating an Approximation of the Two-Parameter Poisson-Dirichlet Process.**

1. Generate \(n\) approximate weights of the Dirichlet process. We apply the following steps:
   
   a. For \(i = 1, \ldots, n+1\), generate \(E_i\) from an exponential distribution with mean 1 and let \(\Gamma_i = E_1 + \cdots + E_i\).
(b) For \( i = 1, \ldots, n \), compute \( p_i(0, \theta) = G_{n-1}^{-1}(\Gamma_i/\Gamma_{n+1}) \), which is simply the quantile function of the \( \text{Gamma}(\theta/n, 1) \) distribution evaluated at \( 1 - \Gamma_i/\Gamma_{n+1} \).

(2) Generate \( m \) approximate weights of the normalized stable law process. We apply the following steps:

(c) For \( j = 1, \ldots, m+1 \), generate \( E_j \) from an exponential distribution with mean 1 and let \( \Gamma_j = E_1 + \cdots + E_j \).

(d) For each \( j = 1, \ldots, m \), compute the weights \( p_j(\alpha, 0) = \Gamma_j^{-1/\alpha} / \sum_{j=1}^{\alpha} \Gamma_j^{-1/\alpha} \).

(3) Repeat step (2) to obtain \( n \) i.i.d. copies of the weights \( (p_1(\alpha, 0), \ldots, p_m(\alpha, 0)) \). Denote these copies by \( (p_1^1(\alpha, 0), \ldots, p_m^1(\alpha, 0)), \ldots, (p_1^n(\alpha, 0), \ldots, p_m^n(\alpha, 0)) \).

(4) Use the weights generated in step (1) and step (3) to produce the weights \( (p_1(0, \theta)p_1^1(\alpha, 0), \ldots, p_1(0, \theta)p_m^1(\alpha, 0), \ldots, p_n(0, \theta)p_1^n(\alpha, 0), \ldots, p_n(0, \theta)p_m^n(\alpha, 0)) \).

(5) The weights of the two-parameter Poisson-Dirichlet process are those weights obtained in step (4) written in descending order. Denote these weight by \( (p_i)_{1 \leq i \leq nm} \).

(6) Generate \( Y_i \) i.i.d. \( \sim H \) for \( i = 1, \ldots, nm \).

(7) The approximated two-parameter Poisson-Dirichlet process is given by the representation (2.1) with \( n \) in the summation replaced by \( nm \).

4 Monotonically Decreasing Approximation to the Normalized Inverse-Gaussian Process

Mimicking Theorem 1 of Zarepour and Al Labadi (2012), we can construct a similar approximation for the normalized inverse-Gaussian process. Specifically, let \( X_n \) be a random variable with distribution \( \text{IG}(\theta/n, 1) \). Define

\[
Q_n(x) = \Pr(X_n > x) = \int_x^{\infty} \frac{\theta}{n\sqrt{2\pi}} t^{-3/2} \exp\left\{ -\frac{1}{2} \left( \frac{\theta^2}{n^2 t} + t \right) + \frac{\theta}{n} \right\} dt.
\]

(4.1)
Let \((Y_i)_{i \geq 1}\) be a sequence of i.i.d. random variables with values in \(\mathcal{X}\) and common distribution \(H\), independent of \((\Gamma_i)_{i \geq 1}\). Then as \(n \to \infty\),

\[
P_{\text{new}}^n, H, 0, \theta (\cdot) = \sum_{i=1}^{n} \frac{Q_n^{-1} \left( \frac{\Gamma_i}{\Gamma_{i+1}} \right)}{\sum_{k=1}^{n} Q_n^{-1} \left( \frac{\Gamma_k}{\Gamma_{n+1}} \right)} \delta_{Y_i}(\cdot) \xrightarrow{a.s.} \sum_{i=1}^{\infty} \frac{L^{-1}(\Gamma_i)}{\sum_{k=1}^{\infty} L^{-1}(\Gamma_k)} \delta_{Y_i}(\cdot).
\]

Here \(\Gamma_i, L(x)\), and \(Q_n(x)\), are defined in (1.6), (1.8), and (4.1), respectively.

Note that, for any \(1 \leq i \leq n\), \(\Gamma_i/\Gamma_{n+1} < \Gamma_{i+1}/\Gamma_{n+1}\) almost surely. Since \(Q_n^{-1}\) is a decreasing function, we have \(Q_n^{-1}(\Gamma_i/\Gamma_{n+1}) > Q_n^{-1}(\Gamma_{i+1}/\Gamma_{n+1})\) almost surely. That is, the weights of the new representation (4.2) decrease monotonically for any fixed positive integer \(n\). As demonstrated in Zarepour and Al Labadi (2012) for the Dirichlet process, we also anticipate that this new representation will yield highly accurate approximations to the normalized inverse-Gaussian process.

Algorithm B: Simulating an Approximation of the Normalized Inverse-Gaussian Process.

1. Fix a relatively large positive integer \(n\).
2. Generate \(Y_i \overset{\text{i.i.d.}}{\sim} H\) for \(i = 1, \ldots, n\).
3. For \(i = 1, \ldots, n+1\), generate \(E_i\) from an exponential distribution with mean 1, independent of \((Y_i)_{1 \leq i \leq n}\) and let \(\Gamma_i = E_1 + \cdots + E_i\).

Table 3: This table reports the Kolmogorov distance \(d(P_{n,H,\alpha,\theta}, H)\) and the computational time in seconds per one simulation, where \(H\) is a uniform distribution on \([0, 1]\).

| \(\alpha\) | \(\theta\) | \(\text{New}\) | \(\text{Time}\) | \(\text{Stick-breaking}\) | \(\text{Time}\) |
|---|---|---|---|---|---|
| 0.1 | 1 | 0.4956 | 2.27 | 0.5059 | 2.34 |
| 0.5 | 1 | 0.3659 | 2.26 | 0.3714 | 2.32 |
| 0.9 | 1 | 0.2295 | 2.25 | 0.3431 | 2.37 |
| 0.1 | 10 | 0.2338 | 2.25 | 0.2328 | 2.34 |
| 0.5 | 10 | 0.1767 | 2.30 | 0.1707 | 2.42 |
| 0.9 | 10 | 0.1097 | 2.31 | 0.3713 | 2.39 |
| 0.1 | 50 | 0.1138 | 2.28 | 0.1123 | 2.4 |
| 0.5 | 50 | 0.0816 | 2.34 | 0.0824 | 2.38 |
| 0.9 | 50 | 0.0635 | 2.34 | 0.4223 | 2.47 |
(4) For $i = 1, \ldots, n$, compute $Q_{n}^{-1}(\Gamma_{i}/\Gamma_{n+1})$, which is simply the quantile function of the inverse-Gaussian distribution with parameter $a/n$ and $1$ evaluated at $1 - \Gamma_{i}/\Gamma_{n+1}$. Computing such values is straightforward in R. For example, one may use the package “GeneralizedHyperbolic”.

5 Empirical Results: A Comparison with the Stick-Breaking Approximation

In this section, we compare the new approximation of the two-parameter Poisson-Dirichlet process (Algorithm A) and the new approximation of the

Figure 1: Sample paths of the two-parameter Poisson-Dirichlet process $P_{H,\alpha,\theta}$, where $H$ is the uniform distribution on $[0,1]$, $\theta = 10$ and $\alpha = 0.1, 0.5$. The solid line denotes the cumulative distribution function of $H$. 
normalized inverse-Gaussian process (Algorithm B) with the corresponding stick-breaking approximations given in (2.1) and (2.4). First we consider the two-parameter Poisson-Dirichlet process. In the simulation, we set \( n = 100, m = 100 \) in Algorithm A and \( n = 100 \times 100 = 10000 \) in (2.1). Throughout this section, we take \( H \) to be the uniform distribution on \([0, 1]\). We compute the Kolmogorov distance between the two-parameter Poisson-Dirichlet process and \( H \) for different values of \( \alpha \) and \( \theta \). The Kolmogorov distance between \( P_{n,H,\alpha,\theta} \) and \( H \), denoted by \( d(P_{n,H,\alpha,\theta}, H) \), is defined by

\[
d(P_{n,H,\alpha,\theta}, H) = \sup_{x \in \mathbb{R}} |P_{n,H,\alpha,\theta}(-\infty, x] - H(-\infty, x]|
\]

Figure 2: Sample paths of a two-parameter Poisson-Dirichlet process \( P_{H,\alpha,\theta} \), where \( H \) is the uniform distribution on \([0, 1]\), \( \theta = 10 \) and \( \alpha = 0.8, 0.99 \). The solid line denotes the cumulative distribution function of \( H \).
\[ := \sup_{x \in \mathbb{R}} |P_{n,H,\alpha,\theta}(x) - H(x)|. \]

For each different value of \( \alpha \) and \( \theta \), we have obtained 500 values of \( d(P_{n,H,\alpha,\theta}, H) \) and reported the average of these values in Table 3. We also listed the corresponding computational time per one simulation, where we have applied the code “system.time” available in R to estimate the time. For instance, for \( \alpha = 0.9 \) and \( \theta = 10 \), the Kolmogorov distance is 0.1097 and the computational time is 2.31 seconds in the new approach, while they are 0.3713 and 2.39 seconds in the stick-breaking approximation. For cases when \( \alpha \) is close to 1 (\( \alpha < 1 \)), inequality (1.3) suggests that distance should

![Figure 3: Sample paths of a two-parameter Poisson-Dirichlet process \( P_{H,\alpha,\theta} \), where \( H \) is the uniform distribution on \([0, 1]\), \( \theta = 100, 1 \) and \( \alpha = 0.99 \). The solid line denotes the cumulative distribution function of \( H \).](image-url)
be close to 0. It follows from the simulation results in Table 3 that the stick-breaking approximation performs very poorly when $\alpha$ is close to 1 ($\alpha < 1$). On the other hand, the new approach performs very well in all cases.

Figures 1, 2 and 3 show sample paths for the approximate two-parameter Poisson-Dirichlet process with different concentration and discount parameters. Clearly, the new approximation performs very well in all cases. As noted earlier, a clear disadvantage of the stick-breaking approximation appears when $\alpha$ is close to 1 ($\alpha < 1$). In this case, as seen in Figures 2 and 3, the two-parameter Poisson-Dirichlet process is not in the proximity of the base measure. This behavior contradicts inequality (1.3).

Next, we compare the new approximation of the normalized inverse-Gaussian process (Algorithm B) with the stick-breaking approximation (2.4). First we empirically illustrate the reason for the numerical instability associated with applying (2.4). We have generated some values of the random

| $n$ | $\theta$ | $i$ | $X_i$ | $Z_i$ | $V_i$ |
|-----|----------|-----|-------|-------|-------|
| 50  | 1        | 1   | 0.6036| $9.4390 \times 10^{-01}$ | $3.9003 \times 10^{-01}$ |
|     |          | 2   | 0.8446| 3.3531 | 2.0111 $\times 10^{-01}$ |
|     |          | 14  | 0.5864| $2.3959 \times 10^{02}$  | $2.4417 \times 10^{-03}$ |
|     |          | 15  | 0.2678| 1.5133 | 1.5035 $\times 10^{-01}$ |
|     |          | 49  | 2.2317| 3.7780 | 3.7134 $\times 10^{-01}$ |
| 50  | 10       | 1   | 5.4321| $1.8579 \times 10^{01}$  | 0.2262 |
|     |          | 2   | 7.3499| $3.9419 \times 10^{01}$  | 0.1571 |
|     |          | 14  | $3.8721 \times 10^{09}$ | $5.1070 \times 10^{01}$ | 0.9999 |
|     |          | 15  | $3.3716 \times 10^{13}$ | $1.8384 \times 10^{02}$ | 1.0000 |
|     |          | 49  | $1.4439 \times 10^{85}$ | $2.5119 \times 10^{02}$ | 1.0000 |
| 100 | 1        | 1   | $4.8603 \times 10^{-01}$ | 3.6576 | 0.1173 |
|     |          | 2   | $2.9847 \times 10^{-01}$ | 5.5739 | 0.0508 |
|     |          | 66  | $1.9751 \times 10^{10}$ | $8.2765 \times 10^{01}$ | 0.9999 |
|     |          | 67  | $3.0512 \times 10^{14}$ | $1.1631 \times 10^{01}$ | 1.0000 |
|     |          | 99  | $1.5636 \times 10^{83}$ | $9.6699 \times 10^{01}$ | 1.0000 |
| 100 | 10       | 1   | 1.9382| 2.2471 | 0.4630 |
|     |          | 2   | 3.8588| $8.6290 \times 10^{02}$ | 0.0044 |
|     |          | 66  | $2.3599 \times 10^{112}$ | 1.5467 | 1.0000 |
|     |          | 67  | $2.3599 \times 10^{114}$ | $7.8622 \times 10^{-01}$ | 1.0000 |
|     |          | 86  | $2.3599 \times 10^{152}$ | 1.4752 | 1.0000 |
variables \(X_i, Z_i\) and \(V_i\) as defined in (1.9) for different values of \(n\) and \(\theta\) and reported the results in Table 4.

It is clear from Table 4 that as \(i\) increases, the values of the random variable \(X_i\) become very large compared with the values of the random variable \(Z_i\). Thus, the random variable \(V_i = X_i/(X_i + Z_i)\) becomes numerically 1. This makes generating new values of the random variable \(X_i\) from the generalized inverse-Gaussian distribution impossible as these distributions become undefined (one of the parameters is numerically zero). In other words, all future weights in the stick breaking representations will vanish. As high-

![Sample paths of the normalized inverse-Gaussian process](image)

**Figure 4:** Sample paths of the normalized inverse-Gaussian process \(P_{H,\alpha,\theta}\), where \(H\) is the uniform distribution on \([0, 1]\), \(\theta = 1, 50\). The solid line denotes the cumulative distribution function of \(H\).
lighted earlier, this causes sampling from the normalized inverse-Gaussian process via the stick-breaking representation to fail in all cases.

Similar to the two-parameter Poisson-Dirichlet process, we calculate the Kolmogorov distance between the normalized inverse-Gaussian process and $H$. We consider only the case when $n = 50$ and $\theta = 1, 10$ (other cases using the stick-breaking representation fail). We have obtained 500 values of the Kolmogorov distance. The average value of the Kolmogorov distance are $0.2908$ ($\theta = 1$) and $0.2481$ ($\theta = 10$) in the new approach, while they are $0.3210$ and $0.4540$ in the stick-breaking approximation. In addition, Figure 4 shows sample paths for the approximate normalized inverse-Gaussian process with different concentration parameters. As seen in this figure, the new approximation performs very well in all case. On the other hand, when $\theta$ is large, contrary to inequality (1.5), the normalized inverse-Gaussian process is not in the proximity of the base measure.

6 Concluding Remarks

Simple and efficient procedures for sampling approximations of the two-Parameter Poisson-Dirichlet Process and the normalized inverse-Gaussian process are proposed. The new approximations are compared with the corresponding stick-breaking approximations. Unlike the stick-breaking approximations, which are limited to certain cases only, the new approximations are very general and can be implemented efficiently in all cases.

Acknowledgement. We thank the Editor, Associate Editor and an anonymous referees for their careful review of the paper and their extremely helpful comments. Thanks are also for Professor Raluca Balan for her suggestions and comments. Research of the authors is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

ABRAMOWITZ, M. and STEGUN, I.A. (1972). Handbook of mathematical functions with formulas, graphs, and mathematical tables. New York, Dover.

AL LABADI, L. and ZAREPOUR, M. (2013). On asymptotic properties and almost sure approximation of the normalized inverse-Gaussian process. Bayesian Anal., 8, 1–16.

BALAKRISHNAN, N. and LAI, C. (2009). Continuous Bivariate Distributions, 2nd edition. Springer-Verlag, New York.

BUNTINE, W. and HUTTER, M. (2010). A Bayesian review of the Poisson-Dirichlet process. arXiv:1007.0296.

CARLTON, M.A. (1999). Applications of the two-parameter Poisson-Dirichlet distribution. Unpublished Ph.D. thesis, Department of Statistics, University of California, Los Angeles.
FAVARO, S., LIJOI, A., MENA, R. and PRÜNSTER, I. (2009). Bayesian nonparametric inference for species variety with a two parameter Poisson-Dirichlet process prior. J. R. Stat. Soc. Ser. B, 71, 993–1008.

FAVARO, S., LIJOI, A. and PRÜNSTER, I. (2012). On the stick-breaking representation of normalized inverse Gaussian priors. Biometrika, 99, 663–774.

FENG, S. (2010). The Poisson-Dirichlet Distribution and Related Topics. Springer-Verlag, New York.

FERGUSON, T.S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Statist., 1, 209–230.

FERGUSON, T.S. and KLASS, M.J. (1972). A representation of independent increment processes without gaussian components. Ann. Math. Statist., 1, 209–230.

ISHWARAN, H. and JAMES, L.F. (2001). Gibbs sampling methods for stick-breaking priors. J. Amer. Statist. Assoc., 96, 161–173.

ISHWARAN, H. and ZAREPOUR, M. (2009). Series representations for multivariate generalized gamma processes via a scale invariance principle. Statist. Sinica, 19, 1665–1682.

JANG, G.H., LEE, J. and LEE, S. (2010). Posterior consistency of species sampling priors. Statist. Sinica, 20, 581–593.

KOTTAS, A. and GELFAND, A.E. (2001). Bayesian semiparametric median regression modeling. J. Amer. Statist. Assoc., 96, 1458–1468.

LIJOI, A., MENA, R.H. and PRÜNSTER, I. (2005). Hierarchical mixture modelling with normalized inverse Gaussian priors. J. Amer. Statist. Assoc., 100, 1278–1291.

MULIERE, P. and TARDELLA, L. (1998). Approximating distributions of random functionals of Ferguson-Dirichlet prior. Canad. J. Statist., 26, 283–297.

NAVARRETE, C., QUINTANA, F.A. and MÜLLER, P. (2008). Some issues on nonparametric Bayesian modeling using species sampling models. Stat. Model., 8, 3–21.

NIETO-BARAJAS, L.E. and PRÜNSTER (2009). A sensitivity analysis for Bayesian nonparametric density estimators. Statist. Sinica, 19, 685–705.

PITMAN, J. and YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Ann. Probab., 2, 855–900.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. Statist. Sinica, 4, 639–650.

TEH, Y.W. (2006). A hierarchical Bayesian language model based on Pitman-Yor processes. In proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp. 985–992.

ZAREPOUR, M. and AL LABADI, L. (2012). On a rapid simulation of the Dirichlet process. Statist. Probab. Lett., 82, 916–924.

Luai Al Labadi
Department of Statistics
Al Ain, Abu Dhabi
P.O. Box 15551, UAE
E-mail: allabadi@uaeu.ac.ae

Mahmoud Zarepour
Department of Mathematics and Statistics
University of Ottawa, Ottawa, ON
K1N 6N5, Canada
E-mail: zarepour@uottawa.ca

Paper received: 23 July 2012; revised: 18 April 2013.