On Approximations of the Beta Process in Latent Feature Models: Point Processes Approach

Luai Al Labadi
University of Toronto, Toronto, Canada
Mahmoud Zarepour
University of Ottawa, Ottawa, Canada

Abstract

In recent times, the beta process has been widely used as a nonparametric prior for different models in machine learning, including latent feature models. In this paper, we prove the asymptotic consistency of the finite dimensional approximation of the beta process due to Paisley and Carin (2009). In particular, we show that this finite approximation converges in distribution to the Ferguson and Klass representation of the beta process. We implement this approximation to derive asymptotic properties of functionals of the finite dimensional beta process. In addition, we derive an almost sure approximation of the beta process. This new approximation provides a direct method to efficiently simulate the beta process. A simulated example, illustrating the work of the method and comparing its performance to several existing algorithms, is also included.

AMS (2000) subject classification. Primary 62F15, 62G20; Secondary 60G51.

Keywords and phrases. Beta process, Ferguson and Klass representation, Finite dimensional approximation, Latent feature models, Simulation

1 Introduction

The beta process was introduced by Hjort (1990) and later clarified by Kim (1999) in the field of Bayesian survival analysis as a class of prior processes for cumulative hazard function. Most recently, Kim et al. (2012) used the beta process to simulate the beta-Dirichlet process, a nonparametric prior for the cumulative intensity functions of a Markov process. Developments of the beta process in machine learning first appeared in the work of Thibaux and Jordan (2007), where it was shown through the application of document classification that the beta process could be used as a nonparametric prior in latent feature models. They also demonstrated that, when the beta process is marginalized out, one can obtain the Indian buffet process first defined in Griffiths and Ghahramani (2006). Since then, the beta
process has been considered for many other applications in machine learning including factor analysis (Paisley and Carin, 2009), featural representations of multiple time series (Fox et al., 2009), Gene-expression analysis (Chen et al., 2010), linear regression (Chen et al., 2010), dictionary learning for image processing (Zhou et al., 2011), and image interpolation (Zhou et al., 2012).

Deriving a stick breaking representation for the beta process was initiated by Teh et al. (2007), who developed the stick breaking representation of the Indian buffet process. Recently, a stick-breaking construction of the full beta process was derived by Paisley et al. (2010). The derivation relied on a limiting process involving finite matrices, analogous to the limiting process used to derive the Indian buffet process. Broderick et al. (2012) demonstrated that the stick-breaking construction of the beta process could be directly obtained from the characterization of the beta process as a Poisson process. A finite approximation of the beta process was suggested without proof by Paisley and Carin (2009). To date, there is no mathematical proof for this approximation despite its use in several applications, including those previously mentioned. One of the main goals of the paper is to provide a comprehensive proof for the finite approximation of the beta process. In particular, we show that this finite approximation converges in distribution to the Ferguson and Klass representation of the beta process. We implement this approximation to derive asymptotic properties of functionals of the finite dimensional beta process.

Sampling from the beta process plays a central role in applications including latent feature models. For example, in factor analysis models (West, 2003; Paisley and Carin, 2009; Broderick et al., 2012), the data matrix is decomposed into the product of two matrices plus noise. The model takes the form:

\[ X = Z\Phi + E, \]

where \( X \in \mathbb{R}^{N \times P} \) is the data matrix and \( E \in \mathbb{R}^{N \times P} \) is an error matrix. The matrix \( \Phi \in \mathbb{R}^{K \times P} \) is a matrix of factors, and \( Z \in \mathbb{R}^{N \times K} \) is a binary matrix of factor loadings. The dimension \( K \) is infinite, and thus the rows of \( \Phi \) consist of an infinite collection of factors. The matrix \( Z \) is formed via a draw from a beta-Bernoulli process. First a sample from the beta process is drawn. Then applying this draw to a Bernoulli process yields an infinite binary vector of the matrix \( Z \). The previous step is repeated to generate the matrix \( Z \), where each successive draw of the Bernoulli process yields a further row of \( Z \). In other words, the beta process is used to provide an infinite collection of coin-tossing probabilities. Tossing these coins corresponds to a draw from the Bernoulli process, yielding an infinite binary vector that
is considered as a latent feature vector (Broderick et al., 2012). Deriving a simple, yet efficient, way to simulate the beta process is another contribution for this paper. Specifically, we derive a finite sum representation which converges almost surely to the Ferguson and Klass (1972) representation for the beta process. In this paper, since the concentration parameter of the beta process is constant, the representation of Ferguson and Klass (1972) is equivalent to the representation of Wolpert and Ickstadt (1998). See Section 5 of this paper for more details regarding this point. In particular, Wolpert and Ickstadt (1998) described an approximate evaluation of the Lévy measure of the beta process. They clarified their approximation through S-PLUS code. At the same time, the authors did not specify the type of convergence in their approximation. Unlike the proposed S-PLUS code, in this paper, we provide a precise and a formal almost sure approximation of the beta process. We also compare the efficiency of the proposed approximation to several existing algorithms including the finite dimensional approximation of the beta process (Paisley and Carin, 2009), Damien et al.’s (1995) algorithm, Lee and Kim’s (2004) algorithm, Lee’s (2007) algorithm, and Wolpert and Ickstadt’s (1998) algorithm, where a significant improvement is achieved. While the algorithms above converge only in distribution, the new algorithm converges almost surely. Almost sure convergence is not only sounder than convergence in distribution but also provides a proper path-by-path comparison between the suggested approximation and the limit. Thus, the almost sure convergence of the new approximation is an additional striking feature distinguishing it from the above approximations.

The remainder of the paper is organized as follows. In Section 2, we introduce the beta process and its conjugate process, the Bernoulli process. In Section 3, we prove the finite dimensional approximation of the beta process. In Section 4, we implement this approximation to derive asymptotic properties of functionals of the finite dimensional beta process. In Section 5, an efficient and convenient method for simulation of the beta process is proposed. The approach is based on deriving a finite sum-representation which converges almost surely to the Ferguson and Klass representation (1972) of the beta process. An example illustrating the method and its performance to other existing approximations is presented in Section 6. Finally, our findings are briefly summarized in Section 7.

2 The Beta Process and the Bernoulli Process

The beta process and the Bernoulli process are examples of a general family of random measures known as completely random measures. Consider a space $X$ with a $\sigma$-algebra $\mathcal{B}$ of subsets of $X$. A random measure $\Phi$ is said
to be completely random measure if for any finite collection $A_1, \ldots, A_n$ of disjoint members of $\mathcal{B}$, the random variables $\Phi(A_1), \ldots, \Phi(A_n)$ are independent. For more details about completely random measures, consult Kingman (1967). Let $B_0$ be a fixed continuous (non-atomic) finite measure on $(\mathbb{X}, \mathcal{B})$ and $c$ be a positive number. Following (Thibaux and Jordan, 2007), the beta process $B$, written $B \sim BP(c, B_0)$, is a completely random measure with Lévy measure

$$\nu(d\omega, ds) = cs^{-1}(1-s)^{c-1}dsB_0(\omega), \quad 0 < s < 1, \omega \geq 0. \tag{2.1}$$

For any $S \in \mathbb{X}$, we have Hjort (1990) & Thibaux and Jordan (2007):

$$E[B(S)] = B_0(S) \quad \text{and} \quad \text{Var}[B(S)] = \frac{B_0(S)}{c + 1}. \tag{2.2}$$

As in the Dirichlet process (Ferguson, 1973), $c$ is called the concentration parameter and $B_0$ is called the base measure. Note that, in general, $c$ can be a nonnegative function of $\omega$, but this is not commonly used in latent feature models. The total mass of $B_0$, $\gamma := B_0(\mathbb{X})$, is called the mass parameter. A draw $B \sim BP(c, B_0)$ is described by:

$$B = \sum_{i=1}^{\infty} p_i \delta_{\omega_i}, \tag{2.3}$$

where $(p_1, \omega_1), (p_2, \omega_2), \ldots$ are the set of atoms in a realization of a nonhomogeneous Poisson process with mean measure $\nu$. Here and throughout the present paper, $\delta_X$ denotes the Dirac measure at $X$, i.e. $\delta_X(A) = 1$ if $X \in A$ and 0 otherwise for a set $A \in \mathcal{B}$. As shown in Eq. 2.3, $B$ is a discrete random measure (with probability 1). Note that, $B$ is a finite measure since $E[\sum_{i=1}^{\infty} p_i] = E[B(\mathbb{X})] = B_0(\mathbb{X}) = \gamma$.

The stick-breaking representation of the beta process takes the form: (Paisley et al., 2010; Broderick et al., 2012)

$$B = \sum_{i=1}^{\infty} \sum_{j=1}^{C_i} V_{i,j}^{(l)} \prod_{l=1}^{i-1} (1 - V_{i,j}^{(l)}) \delta_{\omega_{i,j}}, \tag{2.4}$$

where $C_i \stackrel{i.i.d.}{\sim} \text{Poisson}(\gamma)$, $V_{i,j}^{(l)} \stackrel{i.i.d.}{\sim} \text{beta}(1, c)$, and $\omega_{i,j} \stackrel{i.i.d.}{\sim} B_0/\gamma$. When $C_i = 0$, then the corresponding sum is taken to be zero. The key difference between the stick-breaking representation of the Dirichlet process (Sethuraman, 1994) and that of the beta process is the weights (probabilities). The weights in the Dirichlet process depend on each other, while
this is not the case for the beta process. Specifically, the weights that result
from the stick-breaking representation of the Dirichlet process all come from
a single stick (the unit interval). Thus, they add up to one. On the other
hand, in the beta process, the weights all come from different unit intervals,
therefore they need not add up to one. However, as previously stated, their
sum is almost surely (a.s.) finite.

A connection between the beta process and the Poisson process was
established in Paisley et al. (2012) & Broderick et al. (2012). They demon-
strated that the beta process is a Poisson process with the same mean
measure (2.1). In particular, Paisley et al. (2012) showed that the stick-
breaking construction defined in Eq. 2.4 is equivalent to

\[ B \overset{d}{=} \sum_{j=1}^{C_1} V_{1,j} \delta_{\omega_{1,j}} + \sum_{i=2}^{\infty} \sum_{j=1}^{C_i} V_{i,j} e^{-T_{i,j}} \delta_{\omega_{i,j}}, \]

where \( V_{i,j} \sim \text{beta}(1, c) \), \( T_{i,j} \sim \text{gamma}(i - 1, c) \), \( C_i \) and \( \omega_{i,j} \) are as defined
in Eq. 2.4. In this paper, \( \overset{d}{=} \), \( \overset{\text{d}}{\rightarrow} \), \( \overset{\text{v}}{\rightarrow} \) and \( \overset{\text{a.s.}}{\rightarrow} \) denote equal in dis-
tribution, convergence in distribution, vague convergence and almost sure
convergence, respectively. More details about convergence of random me-
asures are given in Appendix A. In addition, we use the same notation for the
probability measure and its corresponding cumulative distribution function,
i.e. \( B(t) = B((\infty, t]) \) for \( t \in \mathbb{R} \). The inverse of a distribution function
(or measure) \( B \) is defined by

\[ B^{-1}(t) = \inf \{ x : B(x) \geq t \}, \quad 0 < t < 1. \]

A direct link to the Poisson process is given in the following proposition.
For an analogous representation of the Dirichlet process, see Ishwaran and
Zarepour (2002).

**Proposition 1.** For the beta process \( B \) defined as above we have

\[ B \overset{d}{=} \sum_{i=1}^{\infty} \sum_{j=1}^{C_i} \left( e^{-\Gamma_{i-1,j}/c} - e^{-\Gamma_{i,j}/c} \right) \delta_{\omega_{i,j}}, \quad (2.5) \]

where \( \Gamma_{i,j} = E_{i,j}^{(1)} + \cdots + E_{i,j}^{(i)} \) and \( (E_{i,j}^{(l)})_{1 \leq l \leq i} \) are i.i.d.
random variables with the exponential distribution with mean 1, independent of \( \omega_{i,j} \). Here we
have \( \Gamma_{0,j} = 0 \).
**Proof.** Since \( e^{-E_{1,j}/c} \overset{d}{=} \text{beta}(c, 1) \overset{d}{=} 1 - \text{beta}(1, c) \) we have

\[
e^{-\Gamma_{i-1,j}/c} - e^{-\Gamma_{i,j}/c} = e^{-E_{i,j}/c} \ldots e^{-E_{i-1,j}/c} \left(1 - e^{-E_{i,j}/c}\right)
\]

\[
\overset{d}{=}(1 - V_{i,j}^{(1)}) \ldots (1 - V_{i,j}^{(i-1)}) V_{i,j}^{(i)}.
\]

The proof now follows from Eq. 2.4.

Expanding the summation in Eq. 2.5 for the first values of \( i \) gives:

\[
B \overset{d}{=} \sum_{j=1}^{C_1} \left(1 - e^{-\Gamma_{1,j}/c}\right) \delta_{\omega_{1,j}}
\]

\[
+ \sum_{j=1}^{C_2} \left(e^{-\Gamma_{1,j}/c} - e^{-\Gamma_{2,j}/c}\right) \delta_{\omega_{2,j}}
\]

\[
+ \sum_{j=1}^{C_3} \left(e^{-\Gamma_{2,j}/c} - e^{-\Gamma_{3,j}/c}\right) \delta_{\omega_{3,j}} + \ldots.
\]

Since, \( \Gamma_{i,j}/i \overset{a.s.}{\rightarrow} 1 \) as \( i \rightarrow \infty \) (by the strong law of large numbers), the weights becomes negligible for a large value of \( i \). This makes representation (2.5) useful for simulation purposes through a truncation approach. See Zarepour and Al Labadi (2012) & Al Labadi and Zarepour (2013a) for further discussion about truncation procedures.

As pointed out earlier, the beta process is useful as a parameter for the Bernoulli process. The Bernoulli process can be defined, in general, for any base measure on \( X \). In our case, we consider the base measure to be \( B \), where \( B \sim BP(c, B_0) \). Then a Bernoulli process \( Y \) with base measure \( B \), written \( Y \sim BeP(B) \), is a completely random measure

\[
Y = \sum_i b_i \delta_{\omega_i},
\]

where \( b_i \overset{i.i.d.}{\sim} \text{Bernoulli}(p_i) \) for \( p_i \) given in Eq. 2.3. Observe that, since \( E[Y|B] = B \) and \( B \) is a finite measure, the number of non-zero points in any realization of the Bernoulli process is finite.

The beta process is the conjugate prior for the Bernoulli process (Thibaux and Jordan, 2007; Miller, 2011). This conjugacy extends the conjugacy between the Bernoulli and beta distributions. Specifically, let \( B \sim BP(c, B_0) \), and, for \( i = 1, \ldots, m \), let \( X_i|B \sim BeP(B) \) be \( m \) independent Bernoulli
process draws from $B$. The posterior distribution of $B$ after observing $X_1, \ldots, X_m$ is still a beta process:

$$B_m^* = B|X_1, \ldots, X_m \sim BP \left(c^*, B_{0,m}^* \right),$$

where $c^* = c + m$ and $B_{0,m}^* = \frac{c}{c+m}B_0 + \frac{1}{c+m}\sum_{i=1}^{m} X_i$.

3 Finite Dimensional Approximation of the Beta Process

In this section, we prove convergence of the finite approximation of the beta process, which was originally proposed by Paisley and Carin (2009) without a proof. As mentioned in the Introduction, this approximation plays a crucial role in several applications.

**Theorem 1.** Consider a space $(\mathbb{R}, \mathcal{B})$, where $\mathbb{R}$ denotes the real line and $\mathcal{B}$ is the Borel $\sigma$–algebra of subsets of $\mathbb{R}$. Let $B_0$ be a finite continuous fixed measure on $(\mathbb{R}, \mathcal{B})$ with $B_0(\mathbb{R}) = \gamma$ and $c$ be a positive number. For $n > \gamma$, define the process $B_n$ as follows:

$$B_n = \sum_{i=1}^{n} p_{i,n} \delta_{\omega_i},$$

$$p_{i,n} \text{ i.i.d. } \sim \text{Beta} \left(\frac{c\gamma}{n}, c \left(1 - \frac{\gamma}{n}\right)\right),$$

$$\omega_i \text{ i.i.d. } \sim \frac{B_0}{\gamma},$$

$$(p_{i,n})_{1 \leq i \leq n} \text{ and } (\omega_i)_{1 \leq i \leq n} \text{ are independent.}$$

Then, as $n \to \infty$, $B_n \xrightarrow{d} B$, where $B \sim BP \left(c, B_0 \right)$.

**Proof.** The proof is decomposed in several parts.

**Part I:** We apply Proposition 4 (Appendix A) to show that, as $n \to \infty$,

$$nP \left[p_{1,n} \in (x, 1) \right] = \frac{n\Gamma(c)}{\Gamma(c\gamma/n)\Gamma(c - c\gamma/n)} \int_{x}^{1} s^{c\gamma/n - 1}(1 - s)^{c(1 - \gamma/n) - 1} ds$$

$$\xrightarrow{v} \mu(x) = c\gamma \int_{x}^{1} s^{-1}(1 - s)^{c-1} ds. \quad (3.1)$$
Observe that, for any \( x > 0 \), \( \Gamma(x) = \Gamma(x + 1)/x \). With \( x = c\gamma/n \), we obtain \( n/\Gamma(c\gamma/n) = c\gamma/\Gamma(c\gamma/n + 1) \). Since \( \Gamma(x) \) is a continuous function, as \( n \to \infty \), we get

\[
n \frac{\Gamma(c\gamma/n)}{\Gamma(c - c\gamma/n)} \to c\gamma
\]

and

\[
\frac{\Gamma(c)}{\Gamma(c - c\gamma/n)} \to 1.
\]

It follows that, as \( n \to \infty \),

\[
n \frac{n\Gamma(c)}{\Gamma(c\gamma/n)\Gamma(c - c\gamma/n)} \to c\gamma.
\]

On the other hand, since \( x < s < 1 \), we have \( s^{-1} < x^{-1} \) and \( s^{c/n} < 1 \). Thus, \( s^{c/n-1} < x^{-1} \). Consequently, the integrand in Eq. 3.1 is dominated by \( x^{-1}(1 - s)^{c(1-1/n)-1} \), which is integrable for \( x < s < 1 \). Therefore, by the dominated convergence theorem, Eq. 3.1 holds. It follows immediately from Proposition 4 that, as \( n \to \infty \),

\[
\xi_n = \sum_{i=1}^{n} \delta_{(\mu^{-1}(\Gamma_i) + \omega_i)} \to \xi,
\]

where \( \xi \) is a Poisson random measure with mean \( d\mu \times dB_0/\gamma \) and \( \omega_i \) is as defined in the statement of Theorem 2.

**Part II:** We show that

\[
\xi = \sum_{i=1}^{\infty} \delta_{(\mu^{-1}(\Gamma_i) + \omega_i)};
\]

where \( \Gamma_i = E_1 + \cdots + E_i \) and \( (E_i)_{i \geq 1} \) is a sequence of i.i.d. random variables with the exponential distribution with mean 1. It is clear that \( \sum_{i=1}^{\infty} \delta_{\Gamma_i} \) is a a PRM(\( \lambda \)), where \( \lambda \) is Lebesgue measure (Resnick, 2006, Example 5.1). By the special case of Proposition 3.9 of Resnick (1987), \( \sum_{i=1}^{\infty} \delta_{(\Gamma_i, \omega_i)} \) is a
Approximations of the beta process 67

\[ \text{PRM}(\lambda \times B_0/\gamma). \] Define \( T : [0, \infty) \times (-\infty, \infty) \to [0, \infty) \times (-\infty, \infty) \) via \( T(x, y) = (\mu^{-1}(x), y) \). If \( t > 0 \) and \( a < b \), we have

\[
(\lambda \times B_0/\gamma) \circ T^{-1}([t, \infty) \times (a, b)) = (\lambda \times B_0/\gamma) \{(x, y) : \mu^{-1}(x) \geq t \text{ and } a < y < b\}
\]
\[
= (\lambda \times B_0/\gamma) \{(x, y) : x \leq \mu(t) \text{ and } a < y < b\}
\]
\[
= \lambda([0, \mu(t)]) B_0((a, b))/\gamma
\]
\[
= \mu(t)B_0((a, b))/\gamma.
\]

Thus, by Proposition 3.7 of Resnick (1987), \( \xi = \sum_{i=1}^{\infty} \delta_{(\mu^{-1}(\Gamma_i), \omega_i)} \) is a PRM\((d\mu \times dB_0/\gamma)\). Therefore, as \( n \to \infty \),

\[
\xi_n = \sum_{i=1}^{n} \delta_{(p_{i,n}, \omega_i)} \overset{d}{\longrightarrow} \xi = \sum_{i=1}^{\infty} \delta_{(\mu^{-1}(\Gamma_i), \omega_i)}. \quad (3.2)
\]

**Part III:** For \( h > 0 \), the map

\[
T_h \left( \sum_{i=1}^{n} \delta_{(p_{i,n}, \omega_i)} \right) = \sum_{i=1}^{n} p_{i,n}I(p_{i,n} > h)\delta_{\omega_i}
\]

defined on the set of point processes is continuous with respect to vague topology for random measures (there are finite number of terms in the summation). Therefore, as \( n \to \infty \) and for \( h > 0 \), applying \( T_h \) to Eq. 3.2 and the continuous mapping theorem (Resnick, 1987, p. 152) implies

\[
B_{n,h} = \sum_{i=1}^{n} p_{i,n}I(p_{i,n} > h)\delta_{\omega_i} \overset{d}{\longrightarrow} B_h = \sum_{i=1}^{\infty} \mu^{-1}(\Gamma_i)I(\mu^{-1}(\Gamma_i) > h)\delta_{\omega_i}.
\]

Note that, as \( h \to 0 \), \( B_h \overset{d}{\to} B \), where \( B = \sum_{i=1}^{\infty} \mu^{-1}(\Gamma_i)\delta_{\omega_i} \). To complete the proof, by Theorem 3.2 of Billingsley (1999), it remains to show that for any Borel set \( A \),

\[
D_{n,\epsilon}^{h,\epsilon}(A) = P \{|B_n(A) - B_{n,h}(A)| \geq \epsilon\}
\]
\[
= P \left\{ \left| \sum_{i=1}^{n} p_{i,n}\delta_{\omega_i}(A) - \sum_{i=1}^{n} p_{i,n}I(p_{i,n} \geq h)\delta_{\omega_i}(A) \right| \geq \epsilon \right\} \to 0,
\]
as $n \to \infty$ and $h \to 0$. We have
\[
D_{n}^{h,\epsilon}(A) = P\left\{ \sum_{i=1}^{n} p_{i,n} I(p_{i,n} \leq h) \delta_{\omega_i}(A) \geq \epsilon \right\}
\]
\[
\leq \epsilon^{-1} E\left[ \sum_{i=1}^{n} p_{i,n} I(p_{i,n} \leq h) \delta_{\omega_i}(A) \right]
\]
\[
\leq \epsilon^{-1} \sum_{i=1}^{n} E[p_{i,n}I(p_{i,n} \leq h)] E[\delta_{\omega_i}(A)]
\]
\[
= \epsilon^{-1} n E[p_{1,n}I(p_{1,n} \leq h)] E[\delta_{\omega_1}(A)]
\]
\[
= c \epsilon^{-1} E[\delta_{\omega_1}(A)] \int_{0}^{h} x n P\{p_{1,n} \in dx\}.
\]

By Eq. 3.1, as $n \to \infty$, we get
\[
D_{n}^{h,\epsilon}(A) \to c \epsilon^{-1} E[\delta_{\omega_1}(A)] \int_{0}^{h} (1-x)^{c-1} dx. \tag{3.3}
\]

Observe that, the integral in Eq. 3.3 goes to zero as $h \downarrow 0$. Therefore, by Theorem 3.2 of Billingsley (1999), $B_{n} \overset{d}{\to} B$, where
\[
B = \sum_{i} \mu^{-1}(\Gamma_{i}) \delta_{\omega_i} \tag{3.4}
\]
is the Ferguson and Klass (1972) representation of the beta process. This completes the proof of the theorem.

4 Asymptotic Properties of Functionals of the Finite Dimensional Beta Process

The finite dimensional approximation of the Beta process is a very convenient tool to derive various interesting properties of the beta process. In this section, asymptotic properties of some functionals of the finite dimensional beta process are derived. The method can be used to derive variety of interesting theoretical results about the beta process. Specifically, as discussed in the proof of Theorem 1, we have
\[
\sum_{i=1}^{n} \delta_{p_{i,n}} \overset{d}{\to} \sum_{i=1}^{\infty} \delta_{\mu^{-1}(\Gamma_{i})}.
\]
Thus for any, $f \in C_{R}^{+}(\mathbb{E})$ (see Appended A), we have
\[
\int f d \sum_{i=1}^{n} \delta_{p_{i,n}} \overset{d}{\to} \int f d \sum_{i=1}^{\infty} \delta_{\mu^{-1}(\Gamma_{i})}.
\]
It follows that

\[
\sum_{i=1}^{n} f(p_{i,n}) \xrightarrow{d} \sum_{i=1}^{\infty} f(\mu^{-1}(\Gamma_i)).
\] (4.1)

An important instance in Eq. 4.1 is when \(f(p_{i,n}) = -p_{i,n} \log p_{i,n}\), with the convention \(0 \log 0 = 0\). The interest on this choice of \(f\) arises from its association with Shannon entropy in information theory. Note that, despite the fact that \(\sum_{i=1}^{n} p_{i,n}\) is not necessarily 1, for simplicity, we refer to \(H(B_n) = -\sum_{i=1}^{n} p_{i,n} \log p_{i,n}\) as the entropy of \(B_n\). In the next proposition, we derive the expected value of the entropy of the beta process.

**Proposition 2.** Let \(B \sim BP(c,B_0)\) and \(B_n\) be the finite dimensional approximation of the beta process as defined in Theorem 1. If \(H(B) = -\sum_{i=1}^{\infty} \mu^{-1}(\Gamma_i) \log \mu^{-1}(\Gamma_i)\) denotes the entropy of \(B\), then

\[
E(H(B)) = \lim_{n \to \infty} E(H(B_n)) = \gamma [\psi(c + 1) - \psi(1)],
\]

where \(\psi(x) = \frac{d}{dx} \log \Gamma(x)\) is the digamma function.

**Proof.** Set \(u = c\gamma/n\) and \(v = c(1 - \gamma/n)\). Then,

\[
E(H(B_n)) = -\int_{0}^{1} \sum_{i=1}^{n} p_{i,n} \log p_{i,n} \frac{\Gamma(u + v)}{\Gamma(u)\Gamma(v)} p_{i,n}^{u-1} (1 - p_{i,n})^{v-1} dp_{i,n}
\]

\[
= -\sum_{i=1}^{n} \frac{\Gamma(u + v)}{\Gamma(u)\Gamma(v)} \int_{0}^{1} \log p_{i,n} \ p_{i,n}^{u} (1 - p_{i,n})^{v-1} dp_{i,n}
\]

\[
= -\sum_{i=1}^{n} \frac{\Gamma(u + v)}{\Gamma(u)\Gamma(v)} \frac{\partial}{\partial u} \left[p_{i,n}^{u} (1 - p_{i,n})^{v-1}\right] dp_{i,n}
\]

\[
= -\sum_{i=1}^{n} \frac{\Gamma(u + v)}{\Gamma(u)\Gamma(v)} \frac{\partial}{\partial u} \left[\frac{\Gamma(u + 1)\Gamma(v)}{\Gamma(u + v)}\right]
\]

\[
= -\sum_{i=1}^{n} \frac{u}{u + v} \frac{\Gamma(u + v + 1)}{\Gamma(u + v)\Gamma(u + 1)} \log \left[\frac{\Gamma(u + 1)\Gamma(v)}{\Gamma(u + v + 1)}\right]
\]

\[
= -\sum_{i=1}^{n} \frac{u}{u + v} [\psi(u + 1) - \psi(u + v + 1)].
\]
It follows that,

\[
E(H(B_n)) = - \frac{\gamma}{n} \sum_{i=1}^{n} \frac{\psi(c\gamma/n + 1) - \psi(c + 1)}{

As \( n \to \infty \), by the monotone convergence theorem, the result follows.

It is worth noting that, when \( \gamma = 1 \), the expected value of the entropy based on probabilities from the beta process coincides with the entropy based on probabilities from the Dirichlet process (Paisley, 2010), which is unanticipated. Also, it follows from relation 6.3.18 of Abramowitz and Stegun (1972) that the expected value of the entropy of the beta process increases as \( c \) increases.

A weak convergence result for the partial sum process based on the process \( B_n \) is presented in the next proposition. The proof follows from Eq. 3.1 and Proposition 3.21 of Resnick (1987). See also Ishwaran and Zarepour (2009) for an analogous result of the Dirichlet process.

**Proposition 3.** Let \( B_n \) be the finite dimensional approximation of the beta process as defined in Theorem 1. If \( (U_i)_{i \geq n} \) is a sequence of i.i.d. uniform[0, 1] variables, independent of \( (\Gamma_i)_{i \geq n} \), then

\[
\sum_{i=1}^{n} \delta(p_{i,n,i/n}) \overset{d}{\to} \sum_{i=1}^{\infty} \delta(\mu^{-1}(\Gamma_i) U_i),
\]

and

\[
\sum_{i=1}^{\lfloor nt \rfloor} p_{i,n} \overset{d}{\to} \sum_{i=1}^{\infty} \mu^{-1}(\Gamma_i) I(U_i \leq t), \quad 0 \leq t \leq 1.
\]

Here weak convergence occurs in \( D[0, 1] \) with respect to the Skorohod topology.

5 A New Algorithm to Generate the Beta Process

There are two general techniques to writing a series representation for any Lévy process having no Gaussian component. The first one is due to Ferguson and Klass (1972). The second technique is due to Wolpert and Ickstadt (1998). A major difference between the two representations is whether we generate first the jump sizes \((p_i)_{i \geq 1}\) or the jump locations \((\omega_i)_{i \geq 1}\) in Eq. 2.3. We refer the reader to Al Labadi and Zarepour (2013b, 2014a, b) who compared these two representations from the computational point of view and clarified why the representation of Wolpert and Ickstadt is more appropriate.
for nonhomogeneous processes (i.e., when the concentration parameter is not constant but a nonnegative function). On the other hand for homogeneous processes (i.e., when the concentration parameter is constant), the approach of Wolpert and Ickstadt (1998) and the approach of Ferguson and Klass (1972) are equivalent.

Let \( B \sim BP(c, B_0) \) with a continuous \( B_0 \). Then the Ferguson and Klass (1972) representation of \( B \) takes the form given in Eq. 3.4. Since no closed form for the inverse of the Lévy measure (2.1) exists, working with Eq. 3.4 is relatively difficult in practice. The next theorem outlines a remedy to this problem, where an almost sure approximation to Eq. 3.4 is developed based on a similar result in Zarepour and Al Labadi (2012) for the Dirichlet process. Convergence of random measures is taken with respect to the vague topology on the space of point measures. Consult Appendix A for a background on convergence of random measures. As pointed out in the introduction, Wolpert and Ickstadt (1998) provided S-PLUS code to simulate the (homogeneous) beta process. It is worth mentioning that the proposed representation can be easily generalized to cover the nonhomogeneous case. See Al Labadi and Zarepour (2013b) for the details, where a finite sum representation is derived for the (nonhomogeneous) beta-Stacy process (Walker and Muliere, 1997).

**Theorem 2.** Let \( B_0 \) be a fixed continuous finite measure on \( \mathbb{R} \) with \( B_0(\mathbb{R}) = \gamma \). Let \( (\omega_i)_{i \geq 1} \) be i.i.d. random variables with common distribution \( B_0/\gamma \) and \( \Gamma_i = E_1 + \cdots + E_i \), where \( (E_i)_{i \geq 1} \) are i.i.d. random variables with the exponential distribution with mean 1, independent of \( (\omega_i)_{i \geq 1} \). For \( n > \gamma \), define

\[
\mu_n(x) = \frac{\Gamma(c)}{\Gamma(c/n) \Gamma(c - c\gamma/n)} \int_x^1 s^{c\gamma/n - 1} (1 - s)^{c(1 - \gamma/n) - 1} ds.
\]

Then, as \( n \to \infty \),

\[
B_n = \sum_{i=1}^{n} \mu_n^{-1} \left( \frac{\Gamma_i}{\Gamma_{n+1}} \right) \delta_{\omega_i} \overset{a.s.}{\to} B = \sum_{i=1}^{\infty} \mu^{-1} (\Gamma_i) \delta_{\omega_i}.
\]

**Proof.** It follows from the proof of Theorem 2 that, for any \( x > 0 \),

\[
n\mu_n(x) \to \mu(x).
\] (5.1)

Notice that, the left hand side of Eq. 5.1 is a sequence of continuous monotone functions converging to a monotone function for every \( x > 0 \). This is
equivalent to the convergence of their inverse function to the inverse function of the right hand side (Resnick, 1987, Proposition 0.1). Thus,

$$\mu_n^{-1}\left(\frac{x}{n}\right) \to \mu^{-1}(x).$$ \hspace{1cm} (5.2)

Now, taking $x = \Gamma_i$ in Eq. 5.2 and the fact that $\Gamma_{n+1}/n \to 1$ as $n \to \infty$ (by the strong law of large numbers) we get

$$\mu_n^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right) \to \mu^{-1}(\Gamma_i).$$ \hspace{1cm} (5.3)

To prove Theorem 2, by Lemma 1 of Al Labadi and Zarepour (2013b), we show that, for all $k$ fixed,

$$\sum_{i=1}^{k} \mu_n^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right) \delta_{\omega(i)} \xrightarrow{a.s.} \sum_{i=1}^{k} \mu^{-1}(\Gamma_i) \delta_{\omega(i)}$$

as $n \to \infty$, where $\omega(1) \leq \cdots \leq \omega(n)$ represent the corresponding order statistics of $\omega_1, \ldots, \omega_n$. This directly follows by Eq. 5.3.

The next algorithm is used to generate samples from an approximation of the beta process with parameters $c$ and $B_0$, where $B_0$ continuous. The steps of the algorithm are:

1. Fix a relatively large positive integer $n$.
2. Generate $\omega_i \overset{\text{i.i.d.}}{\sim} B_0/\gamma$ for $i = 1, \ldots, n$.
3. Generate $E_i \overset{\text{i.i.d.}}{\sim} \text{exponential}(1)$ for $i = 1, \ldots, n+1$ such that $(E_i)_{1 \leq i \leq n+1}$ and $(\omega_i)_{1 \leq i \leq n}$ are independent where $\Gamma_i = E_1 + \cdots + E_i$.
4. For $i = 1, \ldots, n+1$, compute $\Gamma_i = E_1 + \cdots + E_i$.
5. For $i = 1, \ldots, n$, compute $(\mu_n^{-1}(\Gamma_i/(\Gamma_{n+1})))$, which is simply the quantile function of the beta $(c\gamma/n, c(1-\gamma/n))$ distribution evaluated at $1 - \Gamma_i/\Gamma_{n+1}$.
6. Set $B_n$ as in Theorem 2.

### 6 Empirical Results: Comparison to Other Methods

Several algorithms to sample from the beta process exist in the literature. In this section, we compare the new approximation of the beta process with the finite dimensional approximation of the beta process (Paisley and
Carin (2009), Damien et al.’s (1995) algorithm, Lee and Kim’s (2004) algorithm, Lee’s (2007) algorithm, and Wolpert and Ickstadt’s (1998) algorithm. For the algorithm of Wolpert and Ickstadt (1998), we used the S-PLUS code given in their paper. A summary of these algorithms is given in Appendix B.

In order to make comparisons between the algorithms, we use equivalent settings for the parameters characterizing these algorithms (see Table 1). We refer the reader to the original papers for the details of the algorithms. We consider the beta process with $c = 2$ and $B_0(x) = x$ (i.e., a uniform distribution on $[0, 1]$). We compute the absolute maximum difference between an approximate sample mean and the exact mean. See also Lee and Kim (2004) and Lee (2007) for similar comparisons. The exact mean is $x$; see Eq. 2.2. We refer to this statistic by the maximum mean error. Specifically,

$$
\text{maximum mean error} = \max_x |E[B_n(x)] - E[B(x)]| = \max_x |E[B_n(x)] - x|,
$$

where $x = 0.1, 0.2, \ldots, 0.9, 1.0$, $B_n$ is an approximation of the beta process, and $B \sim BP(c = 2, B_0(x) = x)$. Note that $E[B_n(x)]$ is approximated by obtaining the mean at $x = 0.1, 0.2, \ldots, 0.9, 1.0$ of 3000 i.i.d. sample paths from the approximated process $B_n$. Similarly, we compute the maximum standard deviation error between an approximate sample standard deviation (s.d.) and the exact standard deviation. The exact standard deviation is $\sqrt{x/3}$; see Eq. 2.2. Thus,

$$
\text{maximum s.d. error} = \max_x |s.d.[B_n(x)] - s.d.[B(x)]| = \max_x \left| s.d.[B_n(x)] - \sqrt{x/3} \right|.
$$

Table 1: This table reports the maximum mean error, the maximum standard deviation error, and the corresponding computation time

| Algorithm | Parameters | max. mean error | max. s.d. error | Time |
|-----------|------------|----------------|----------------|------|
| DSL       | $m = n = 200$ | 0.0103         | 0.0104         | 54.51|
| LK        | $\epsilon = 0.01$ | 0.0136         | 0.0133         | 0.31 |
| Lee       | $n = 200, \epsilon = 0.05$ | 0.0109         | 0.0548         | 0.92 |
| PC        | $n = 200$ | 0.0087         | 0.0200         | 0.28 |
| WI        | $M = 200, \epsilon = 0.005$ | 0.0138         | 0.0109         | 21.26|
| New       | $n = 200$ | 0.0086         | 0.0069         | 9.67 |

Here, DSL, KL, PC, and WI stand for Damien et al.’s algorithm (1995), Lee and Kim’s algorithm (2004), Paisley and Carin’s (2009) algorithm, and Wolpert and Ickstadt (1998) algorithm, respectively.
Table 1 depicts values of the maximum mean error, the maximum standard deviation error, and the corresponding computational time. The computational time is computed by applying the code “System.Time” available in R. As seen in Table 1, the new algorithm has the smallest mean and standard deviation errors. In addition, it has a reasonable computation time.

7 Conclusions

In this paper, we have proved the finite dimensional approximation of the beta process (Paisley and Carin, 2009). This approximation has been used in several machine learning models. We employed this approximation to derive asymptotic properties of functionals of the finite dimensional beta process. We have also derived an almost surely approximation of the beta process. This new approximation provides a simple, yet efficient, way to simulate the beta process.

References

Abramowitz, M. and Stegun, I. A. (1972). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 10th. Dover, New York.

Al Labadi, L. and Zarepour, M. (2013a). On asymptotic properties and almost sure approximation of the norMalized inverse-Gaussian process. Bayesian Anal. 8, 3, 553–568.

Al Labadi, L. and Zarepour, M. (2013b). A Bayesian nonparametric goodness of fit test for right censored data based on approximate samples from the beta-Stacy process. Canad. J. Statist. 41, 3, 466–487.

Al Labadi, L. and Zarepour, M. (2014a). On Simulations from the Two-Parameter Poisson-Dirichlet Process and the NorMalized Inverse-Gaussian Process. Sankhya A 76, 158–176.

Al Labadi, L. and Zarepour, M. (2014b). Goodness of fit tests based on the distance between the Dirichlet process and its base measure. J. Nonparametr. Stat. 26, 341–357.

Billingsley, P. (1999) Convergence of Probability Measures, third edition. John Wiley and Sons, Inc.

Broderick, T., Jordan, M. I. and Pitman, J. (2012). Beta processes, stick-breaking, and power laws. Bayesian Anal. 7, 439–476.

Chen, B., Paisley, J. and Carin, L. (2010) Sparse linear regression with beta process priors. International Conference on Acoustics, Speech and Signal Processing. Dallas, TX.

Chen, B., Chen, M., Paisley, J., Zaas, A., Woods, C., Ginsburg, G. S., Hero, A., Lucas, J., Dunson, D. and Carin, L. (2010). Bayesian inference of the number of factors in gene-expression analysis: Application to human virus challenge studies. BMC Bioinformatics 11, 552.

Damien, P., Laud, P. and Smith, A. F. M. (1995). Approximate random variate generation from infinitely divisible distributions with applications to Bayesian inference. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 547–563.
FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* 1, 209–230.

FERGUSON, T. S. and KLAAS, M. J. (1972). A representation of independent increment processes without Gaussian components. *Annals of Mathematical Statistics* 1, 209–230.

FOX, E., SUDDERTH, E., JORDAN, M. and WILLSKY, A. (2009). Sharing features among dynamic systems with beta processes. *Adv. Neural Inf. Process. Syst.* 22, 549–557.

GRIFFITHS, T. and GHAHRAMANI, Z. (2006). Infinite latent feature models and the Indian buffet process. *Adv. Neural Inf. Process. Syst.* 18, 475–482.

HJORT, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *Ann. Statist.* 18, 1259–1294.

ISHWARAN, H. and ZAREPOUR, M. (2009). Series Representations for Multivariate Generalized Gamma Processes via a Scale Invariance Principle. *Statist. Sinica* 19, 1665–1682.

ISHWARAN, H. and ZAREPOUR, M. (2002). Exact and approximate sum representations for the Dirichlet process. *Canad. J. Statist.* 30, 269–283.

KALLENBERG, O. (1983) Random Measures, third edition. Akademie-Verlag, Berlin.

KIM, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* 27, 562–588.

KIM, Y., JAMES, L. and WEIBBACH, R. (2012). Bayesian analysis of multistate event history data: Beta-dirichlet process prior. *Biometrika* 99, 127–140.

KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math.* 21, 59–78.

LEE, J. (2007). Sampling methods for neutral to the right processes. *J. Comput. Graph. Statist.* 16, 656–671.

LEE, J. and KIM, Y. (2004). A new algorithm to generate beta processes. *Comput. Statist. Data Anal.* 25, 401–405.

MILLER, K. T. (2011) Bayesian Nonparametric Latent Feature Models. PhD thesis, University of California, Berkeley.

PAISLEY, J. (2010) Machine Learning with Dirichlet and Beta Process Priors: Theory and Applications. PhD Thesis, Duke University.

PAISLEY, J., BLEI, D. and JORDAN, M. (2012). *Stick-breaking beta processes and the Poisson process*. Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS).

PAISLEY, J. and CARIN, L. (2009). *Nonparametric factor analysis with beta process priors*. Proceedings of the International Conference on Machine learning (ICML).

PAISLEY, J., ZAAS, A., WOODS, C. W., GINSBURG, G. S. and CARIN, L. (2010). *A stick-breaking construction of the beta process*. Proceedings of the International Conference on Machine learning (ICML).

RESNICK, S. I. (1987). *Extreme Values Regular Variation and Point Processes*. Springer, New York.

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

TEH, Y. W., GÖRÜR, D. and GHAHRAMANI, Z. (2007). *Stick-breaking construction for the Indian buffet process*. Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS).

THIBAUX, R. and JORDAN, M. I. (2007). *Hierarchical beta processes and the Indian buffet process*. Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS).
walker, S. and muliere, P. (1997). Beta-stacy processes and a generalisation of the polya-urn scheme. Ann. Statist. 25, 1762–1780.

west, M. (2003). Bayesian factor regression models in the “large p, small n” paradigm. Bayesian Anal. 7, 723–732.

wolpert, R. L. and ickstadt, K. (1998). Simulation of Lévy random fields. Springer, New York. Practical Nonparametric and Semiparametric Bayesian Statistics, Day, D., nuller, P. and sinha, D. (eds.), p. 237–242.

zarepour, M. and al labadi, L. (2012). On a rapid simulation of the Dirichlet process. Statist. Probab. Lett. 82, 916–924.

zhou, M., yang, H., sapiro, G., dunson, D. and carin, L. (2011) Dependent hierarchical beta process for image interpolation and denoising. In proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS).

zhou, M., chen, H., paisley, J., ren, L., li, L., xing, Z., dunson, D., sapiro, G. and carin, L. (2012). Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images. IEEE Trans. Image Process 21, 130–144.

Appendix

A Vague Convergence

The main objective of this appendix is to give a brief introduction about convergence of random measures. Let \((E, E')\) be a state space as before. Let \(C^+_K(E)\) be the set of continuous functions \(f : E \to [0, \infty)\) with compact support. A measure \(\mu\) is called Radon if \(\mu(K) < \infty\) for any compact set \(K\) in \(E\). Let \(M_+(E)\) be the space of Radon measures in \(E\). Let \(\mathcal{M}_+(E)\) be the smallest \(\sigma\)–algebra of subsets of \(M_+(E)\) making the maps \(\mu \to \mu(f) = \int f(x) d\mu(x)\) from \(M_+(E)\) to \(\mathbb{R}\) measurable for all functions \(f \in C^+_K(E)\). Note that, \(\mathcal{M}_+(E)\) is the Borel \(\sigma\)–algebra generated by the topology of vague convergence. If \(\mu_n, \mu \in M_+(E)\), we say that \((\mu_n)_n\) converges vaguely to \(\mu\) (and we write \(\mu_n \overset{v}{\to} \mu\)) if \(\mu_n(f) \overset{v}{\to} \mu(f)\) for any \(f \in C^+_K(E)\).

A random measure on \(E\) is any measurable map \(\xi\) defined on a probability space \((\Omega, \mathcal{A}, P)\) with values in \((M_+(E), \mathcal{M}_+(E))\). If \(\xi_n, \xi\) are random measures on \(E\), we say that \((\xi_n)_n\) converges in distribution to \(\xi\) (and we write \(\xi_n \overset{d}{\to} \xi\)) if \(\{P \circ \xi_n^{-1}\}_n\) converges weakly to \(P \circ \xi^{-1}\). By Theorem 4.2 of Kallenberg (1983), \(\xi_n \overset{d}{\to} \xi\) if and only if \(\xi_n(f) \to \xi(f)\), i.e.

\[
\int_E f(x) \xi_n(dx) \to \int_E f(x) \xi(dx), \forall f \in C^+_K(E).
\]
We say that \((\xi_n)_n\) converges vaguely almost surely to \(\xi\) (and write \(\xi_n \xrightarrow{a.s.} \xi\)) if there exists a set \(\tilde{\Omega} \in \mathcal{A}\) with \(P(\tilde{\Omega}) = 1\) such that \(\forall \omega \in \tilde{\Omega}, \xi_n(\omega, \cdot) \xrightarrow{v} \xi(\omega, \cdot)\), i.e.

\[
\int_E f(x)\xi_n(\omega, dx) \to \int_E f(x)\xi(\omega, dx), \quad \forall f \in C_K^+(\mathbb{E}).
\]

The space \(M_+^{bp}(\mathbb{E})\) endowed with the vague topology is a complete separable metric space (Resnick, 1987, Proposition 3.17). See also Kallenberg (1983).

The next proposition is fundamental in studying the weak convergence of Poisson processes. It gives necessary and sufficient conditions for empirical measures to converge to a Poisson random measure. The proof follows by mimicking the proof of Proposition 3.21 of Resnick (1987) with \(j \geq 1\) and \(j/n\) are replaced by \(1 \leq j \leq n\) and \(\omega_j\), respectively. See also Ishwaran and Zarepour (2009) for an analogous result.

**Proposition 4.** Suppose for each \(n \geq 1\), we have \((X_{j,n})_{1 \leq j \leq n}\) are i.i.d. random variable random elements of \((\mathbb{E}, \mathcal{E})\) and \(\mu\) is a Radon measure on \((\mathbb{E}, \mathcal{E})\). Let \(\omega_i \sim \text{i.i.d.} F\) such that \((\omega_j)_{1 \leq j \leq n}\) and \((X_{j,n})_{1 \leq j \leq n}\) are independent. Define \(\xi_n = \sum_{j=1}^n \delta_{(\omega_j, X_{j,n})}\) and suppose \(\xi\) is a Poisson random measure on \(\mathbb{R} \times \mathbb{E}\) with mean \(dF \times d\mu\). Then

\[
\xi_n \xrightarrow{d} \xi \text{ in } M_p(\mathbb{R} \times \mathbb{E}) \text{ if and only if } nP[X_{1,n} \in \cdot] \xrightarrow{v} \mu \text{ on } \mathbb{E}.
\]

**B Other Sampling Algorithms**

Several algorithms are suggested to sample from the beta process \(B \sim BP(c, B_0)\) with a continuous \(B_0\). We consider the algorithm Damien et al. (1995), the algorithm of Lee and Kim (2004) and the algorithm Lee (2007). Below is a brief discussion of these algorithms. We refer the reader to the original papers for more details.

**• Damien-Laud-Smith Algorithm:** Using the fact that the distributions of the increments of a nondecreasing Lévy process are infinitely divisible, Damien et al. (1995) derived an algorithm to generate approximations for infinitely divisible random variables and used it to generate the beta process. Let \(p_i\) denotes the increment of the process \(B\) in the interval \(\Delta_i = (\omega_{i-1}, \omega_i]\), i.e. \(p_i = B(\omega_i) - B(\omega_{i-1})\). The steps of the Damien-Laud-Smith algorithm for simulating an approximation for the jump \(p_i\) are:

1. Fix a relatively large positive integer \(n\).
(2) Generate independent values $z_{ij}$ from the probability density function $dB_0(t)/B_0(\Delta_i)$, for $j = 1, \ldots, n$.

(3) Generate $x_{ij} \sim \text{beta}(1, c)$, for $j = 1, \ldots, n$.

(4) Generate $y_{ij}$: $y_{ij} \mid x_{ij} \sim \text{Poisson}(\lambda_i n^{-1} x_{ij}^{-1})$, for $j = 1, \ldots, n$, where

$$\lambda_i = B_0(\omega_i) - B_0(\omega_{i-1}).$$

(5) Set $p_{i,n} = \sum_{j=1}^{n} x_{ij} y_{ij}$. For large $n$, $p_{i,n}$ is an approximation of $p_i$.

Damien et al. (1995) showed that $p_{i,n} \xrightarrow{d} p_i$, as $n \to \infty$. That is, $p_{i,n}$ is an approximate sample from the $i$th increment of $B$. Note that, the Damien-Laud-Smith algorithm generates only the increments of the process and not the entire process. To obtain the whole process, we set

$$B_{m,n} = \sum_{i=1}^{m} p_{i,n} \delta_{\omega_i}.$$  

For large $m$ and $n$, $B_{m,n}$ is an approximation of $B$.

• **Lee-Kim Algorithm:** The Kim and Lee algorithm for the beta process with parameters $c$ and $B_0$ with $B_0$ continuous can be described as follows. First the Lévy measure $\nu$ of the beta process given by Eq. 2.1 is approximated by:

$$\nu_\epsilon(d\omega, ds) = \frac{c}{\epsilon} b(s : \epsilon, c) dB_0(\omega) ds,$$

where

$$b(x : a, b) = \frac{\Gamma(a + b)}{\Gamma(a) \Gamma(b)} x^{a-1} (1 - x)^{b-1}, \quad \text{for } 0 < x < 1, a > 0, b > 0. \quad (7.1)$$

The steps of the Lee-Kim algorithm for the beta process $B$ are:

(1) Fix a relatively small positive number $\epsilon$.

(2) Generate the total number of jumps $n \sim \text{Poisson}(c \gamma / \epsilon)$.

(3) Generate i.i.d. random variables $\omega_1, \ldots, \omega_n$ from the probability density function $dB_0/\gamma$.

(4) Let $\omega_{(1)} \leq \ldots \leq \omega_{(n)}$ be the corresponding order statistics of $\omega_1, \ldots, \omega_n$.

(5) Generate the jump sizes $p_1, \ldots, p_n: p_i \mid \omega_{(i)} \sim \text{Beta}(\epsilon, c)$.

(6) Set $B_\epsilon = \sum_{i=1}^{n} p_i \delta_{\omega_{(i)}}$.

Lee and Kim (2004) showed that $B_\epsilon \xrightarrow{d} B$, as $\epsilon \to 0$. 
• **Lee Algorithm:** The steps of the Lee algorithm are:

1. Fix a relatively large positive integer $n$.
2. Generate i.i.d. random variables $\omega_1, \ldots, \omega_n$ from the probability density function $dB_0/\gamma$.
3. For $i = 1, \ldots, n$, generate $x_i \sim b(s : \epsilon, c)$, where $b(s : \epsilon, c)$ is defined in Eq. 7.1
4. For $i = 1, \ldots, n$, generate $y_i \sim \text{Poisson} \left( \gamma b(x_i : 1, c)/nx_i b(x_i : \epsilon, c) \right)$.
5. Set $B_n = \sum_{i=1}^{n} x_i y_i \delta_{\omega_i}$.

Lee (2007) proved that, as $n \to \infty$, $B_n \overset{d}{\to} B$.

Luai Al Labadi  
Department of Statistical Sciences,  
University of Toronto, Toronto, ON  
M5S 3G3, Canada  
E-mail: luai.allabadi@utoronto.ca

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

Paper received: 9 November 2015.