Bayesian Analysis (2013) **8**, Number 3, pp. 553–568

On Asymptotic Properties and Almost Sure Approximation of the Normalized Inverse-Gaussian Process

Luai Al Labadi * and Mahmoud Zarepour †

Abstract. In this paper, similar to the frequentist asymptotic theory, we present large sample theory for the normalized inverse-Gaussian process and its corresponding quantile process. In particular, when the concentration parameter is large, we establish the functional central limit theorem, the strong law of large numbers and the Glivenko-Cantelli theorem for the normalized inverse-Gaussian process and its related quantile process. We also derive a finite sum representation that converges almost surely to the Ferguson and Klass representation of the normalized inverse-Gaussian process. This almost sure approximation can be used to simulate the normalized inverse-Gaussian process.

Keywords: Brownian bridge, Dirichlet process, Ferguson and Klass representation, Nonparametric Bayesian inference, Normalized inverse-Gaussian process, Quantile process, Weak convergence

1 Introduction

The objective of Bayesian nonparametric inference is to place a prior on the space of probability measures. The Dirichlet process, formally introduced in Ferguson (1973), is considered the first celebrated example on this space. Several alternatives for the Dirichlet process have been proposed in the literature. In this paper, we focus on one such prior, namely the normalized inverse-Gaussian (N-IG) Process introduced by Lijoi et al. (2005b). The authors in the foregoing paper used the normalized inverse-Gaussian process in the context of mixture modeling and showed that this prior exhibits an attractive and useful clustering behavior, quite different from that of the Dirichlet process. We refer the reader to the original paper of Lijoi et al. (2005b) for a more detailed comparison between the two processes. Relevant contributions to the normalized inverse-Gaussian process include, among others, Favaro et al. (2012), Jang et al. (2010), James et al. (2009) and Lijoi et al. (2007).

We begin by recalling the definition of the normalized inverse-Gaussian distribution. The random vector $(Z_1, \ldots, Z_m)$ is said to have the *normalized inverse-Gaussian* distribution with parameters $(\gamma_1, \ldots, \gamma_m)$, where $\gamma_i > 0$ for all i, if it has the joint probability

*Department of Statistics, College of Business and Economics, United Arab Emirates University, P.O. Box 17555, Al Ain, United Arab Emirates, allabadi@uaeu.ac.ae

†Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada, zarepour@uottawa.ca

 DOI:10.1214/13-BA821

density function

$$f(z_1,\ldots,z_m) = \frac{e^{\sum_{i=1}^{\infty}\gamma_i}\prod_{i=1}^m\gamma_i}{2^{m/2-1}\pi^{m/2}} \times K_{-m/2}\left(\sqrt{\sum_{i=1}^m \frac{\gamma_i^2}{z_i}}\right) \times \left(\sum_{i=1}^m \frac{\gamma_i^2}{z_i}\right)^{-m/4} \times \prod_{i=1}^m z_i^{-3/2} \times I_{\mathbb{S}}(z_1,\ldots,z_m), \tag{1}$$

where K is the modified Bessel function of the third type, $\mathbb{S} = \{(z_1,\ldots,z_m) : z_i \geqslant 0, \sum_{i=1}^m z_i = 1\}$, and $I_{\mathbb{S}}$ represents the indicator function of the set $\mathbb{S}$. For more details about the modified Bessel functions consult Abramowitz and Stegun (1972), Chapter 9.

Consider a Polish space $\mathfrak{X}$ with the Borel $\sigma-$algebra $\mathcal{A}$ of subsets of $\mathfrak{X}$. Let H be a fixed probability measure on $(\mathfrak{X},\mathcal{A})$ and a be a positive number. Following Lijoi et al. (2005b), a random probability measure $P_{H,a} = \{P_{H,a}(A)\}_{A\in\mathcal{A}}$ is called a normalized inverse-Gaussian process on $(\mathfrak{X},\mathcal{A})$ with parameters a and H, if for any finite measurable partition $A_1,\ldots, A_m$ of $\mathfrak{X}$, the joint distribution of the vector $(P_{H,a}(A_1),\ldots P_{H,a}(A_m))$ has the normalized inverse-Gaussian distribution with parameter $(aH(A_1),\ldots aH(A_m))$. We assume that if $H(A_i)=0$, then $P_{H,a}(A_i)=0$ with probability one. The normalized inverse-Gaussian process with parameters a and H is denoted by N-IGP(a,H), and we write $P_{H,a} \sim$ N-IGP(a,H).

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

$$E(P_{H,a}(A)) = H(A) \quad \text{and} \quad Var(P_{H,a}(A)) = \frac{H(A)(1-H(A))}{\xi(a)}, \tag{2}$$

where here and throughout this paper

$$\xi(a) = \frac{1}{a^2 e^a \Gamma(-2,a)} \tag{3}$$

and $\Gamma(-2,a) = \int_a^\infty t^{-3}e^{-t}dt$. Furthermore, for any two disjoint sets A_i and $A_j \in \mathcal{A}$,

$$E(P_{H,a}(A_i)P_{H,a}(A_j)) = H(A_i)H(A_j)\frac{\xi(a)-1}{\xi(a)}. \tag{4}$$

Observe that, for large a and any real number r,

$$\Gamma(r,a)a^{-r}e^a \approx \frac{1}{a} + \frac{r-1}{a^2} + \frac{(r-1)(r-2)}{a^3} + \cdots \tag{5}$$

(Abramowitz and Stegun 1972, Formula 6.5.32), where we use the notation $f(a) \approx g(a)$ if $\lim_{a\to\infty} f(a)/g(a) = 1$. Consequently, $\xi(a) \approx a$. It follows from (2) that H plays the role of the center of the process, while a can be viewed as the concentration parameter. The larger the value of a, the more likely the realization of $P_{H,a}$ will be close to H. Specifically, for any fixed set $A \in \mathcal{A}$ and $\epsilon > 0$, by Chebyshev's inequality we have

$$\Pr\{|P_{H,a}(A) - H(A)| > \epsilon\} \leqslant \frac{H(A)(1-H(A))}{\xi(a)\epsilon^2}. \tag{6}$$

That is, $P_{H,a}(A) \xrightarrow{p} H(A)$ as $a \to \infty$.

Similar to the Dirichlet process, a series representation of the normalized inverse-Gaussian process can be easily derived from the Ferguson and Klass (1972). See also Nieto-Barajas and Prünster (2009). Specifically, let $(E_i)_{i\geqslant 1}$ be a sequence of independent and identically distributed (i.i.d.) random variables with an exponential distribution with mean of 1. Define

$$\Gamma_i = E_1 + \cdots + E_i. \tag{7}$$

Let $(\theta_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables with values in $\mathfrak{X}$ and common distribution H, independent of $(\Gamma_i)_{i\geqslant 1}$. Then the normalized inverse-Gaussian process with parameters a and H has the series representation

$$P_{H,a}(\cdot) = \sum_{i=1}^{\infty} \frac{L^{-1}(\Gamma_i)}{\sum_{i=1}^{\infty} L^{-1}(\Gamma_i)} \delta_{\theta_i}(\cdot), \tag{8}$$

where

$$L(x) = \frac{a}{\sqrt{2\pi}} \int_x^{\infty} e^{-t/2} t^{-3/2}\, dt, \text{ for } x > 0, \tag{9}$$

and δ_X denotes the Dirac measure at X (i.e. $\delta_X(B) = 1$ if $X \in B$ and 0 otherwise). Note that, working with (4) is relatively difficult in practice because no closed form for the inverse of the Lévy measure (6) exists. Moreover, to determine the random weights in (4), an infinite sum must be computed. A remedy of such complexity will appear in Theorem 4 of Section 5 in this paper, where an almost sure approximation to (4) is given based on a similar result in Zarepour and Al Labadi (2012).

The main objective of the present paper is to study the weak convergence of the centered and scaled process

$$D_{H,a}(\cdot) = \sqrt{a}\left(P_{H,a}(\cdot) - H(\cdot)\right), \tag{10}$$

as $a \to \infty$. Nonparametric Bayesian procedures can be viewed as functionals on priors. Therefore, like frequentists' empirical process (Shorack and Wellner 1986), large sample theory of many important functionals of the N-IGP(a, H) will simply follow from this result. For example, it will pave the way for studying the Bayesian bootstrap based on the normalized inverse-Gaussian process. We point out that, the weak convergence of the centered and scaled Dirichlet process was studied by Lo (1987) to establish asymptotic validity of the Bayesian bootstrap. See also Ishwaran et al. (2009). An interesting generalization of Lo (1987) to the two-parameter Poisson-Dirichlet process was obtained by James (2008). We would like to emphasize that the result of James for the two-parameter Poisson-Dirichlet process holds for any discount parameter $\alpha \in [0, 1]$. In the two foregoing papers, the proofs of the results are based on constructing certain distributional identities which conclude summability. We refer the reader to Proposition 4.1 of James (2008) for more details. Since constructing an analogous distributional identity for the normalized inverse-Gaussian process does not seem to be trivial, the approach of James (2008) is not followed in this paper.

This paper is organized as follows. In Section 2, as $a \to \infty$, we show that the limiting process for the process (10) is the Brownian bridge. In Section 3, we derive the limiting process for the quantile process

$$Q_{H,a}(\cdot) = \sqrt{a}\left(P_{H,a}^{-1}(\cdot) - H^{-1}(\cdot)\right), \tag{11}$$

as $a \to \infty$, where, in general, the inverse of a distribution function F is defined by

$$F^{-1}(t) = \inf\{x : F(x) \geqslant t\}, \quad 0 < t < 1.$$

The strong law of large numbers and the Glivenko-Cantelli theorem for the normalized inverse-Gaussian process are discussed in Section 4. In Section 5, we derive a finite sum-representation which converges almost surely to the Ferguson and Klass representation of the normalized inverse-Gaussian process. Section 6 contains some concluding remarks.

2 Asymptotic Properties of the N-IG Process

In this section, we study the weak convergence of the process $D_{H,a}$ defined in (10) for large values of a. Let $\mathscr{S}$ be a collection of Borel sets in $\mathbb{R}$ and H be a probability measure on $\mathbb{R}$. We recall the definition of a Brownian bridge indexed by $\mathscr{S}$. A Gaussian process $\{B_H(S) : S \in \mathscr{S}\}$ is called a *Brownian bridge with parameter measure* H if $E[B_H(S)] = 0$ for any $S \in \mathscr{S}$ and

$$Cov\,(B_H(S_i), B_H(S_j)) = H(S_i \cap S_j) - H(S_i)H(S_j)$$

for any $S_i, S_j \in \mathscr{S}$ (Kim and Bickel 2003).

The next lemma gives the limiting distribution of the process (10) for any finite Borel set $S_1, \ldots, S_m \in \mathscr{S}$, as $a \to \infty$. The proof of the lemma for $m = 2$ is given in the appendix and can be generalized easily to the case of arbitrary m.

Lemma 1. *Let $D_{H,a}$ be defined by (10). For any fixed sets $S_1, \ldots, S_m$ in $\mathscr{S}$ we have*

$$(D_{H,a}(S_1), D_{H,a}(S_2), \ldots, D_{H,a}(S_m)) \xrightarrow{d} (B_H(S_1), B_H(S_2), \ldots, B_H(S_m)),$$

as $a \to \infty$, where B_H is the Brownian bridge with parameter H.

In this paper, $D[-\infty, \infty]$ denotes the space of càdlàg functions (right continuous with left limits) on $[-\infty, \infty]$ which is equipped with the Skorokhod topology. The process $D_{H,a}$ in (10) can be defined on $[-\infty, \infty]$ in which its sample path is in $D[-\infty, \infty]$ such that $D_{H,a}(t) = D_{H,a}((-\infty, t])$. Right continuity at $-\infty$ can be achieved by setting $D_{H,a}(-\infty) = 0$; the left limit at $+\infty$ also equals zero, the natural value of $D_{H,a}(+\infty)$. For more details, consult Pollard (1984), Chapter 5. The following theorem shows that the process $D_{H,a}$ converges to the process B_H in $D[-\infty, \infty]$. If X and $(X_a)_{a>0}$ are random variables with values in a metric space M, we say that $(X_a)_a$ converges in distribution to X as $a \to \infty$ (and we write $X_a \xrightarrow{d} X$) if for any sequence $(a_n)_n$ converging to ∞, X_{a_n} converges in distribution to X. The proof of the theorem is given in the appendix.

Theorem 1. *We have, as $a \to \infty$,*

$$D_{H,a}(\cdot) = \sqrt{a}\,(P_{H,a}(\cdot) - H(\cdot)) \xrightarrow{d} B_H(\cdot)$$

in $D[-\infty, \infty]$ with respect to Skorokhod topology, where B_H is the Brownian bridge with parameter measure H.

3 Asymptotic Properties of the N-IG Quantile Process

Ferguson (1973) considered different estimation problems including estimation of the median and quantiles using different loss functions under the Dirichlet prior. In this section, similar to the frequentist asymptotic theory of quantile process, we establish large sample theory for the normalized inverse-Gaussian quantile process. The following corollary derives the weak limit of the inverse-Gaussian quantile process defined in (11) when the concentration parameter a is large.

Corollary 1. *Let $0 < p < q < 1$, and H be a continuous function with positive derivative h on the interval $\left[H^{-1}(p) - \epsilon, H^{-1}(q) + \epsilon\right]$ for some $\epsilon > 0$. Let λ be the Lebesgue measure on $[0, 1]$. Let $Q_{H,a}$ be the normalized inverse-Gaussian process defined in (11). As $a \to \infty$, we have*

$$Q_{H,a}(\cdot) \xrightarrow{d} -\frac{B_\lambda(\cdot)}{h(H^{-1}(\cdot))} = Q(\cdot),$$

in $D[p, q]$. That is, the limiting process is a Gaussian process with zero-mean and covariance function

$$Cov\,(Q(s), Q(t)) = \frac{\lambda(s \wedge t) - \lambda(s)\lambda(t)}{h(H^{-1}(s))h(H^{-1}(t))}, \quad s, t \in \mathbb{R},$$

where $s \wedge t$ denotes the minimum of s and t.

Proof. By Theorem 1, the process $\sqrt{a}\,(P_{H,a}(\cdot) - H(\cdot))$ converges in distribution to the process $B_H = B_\lambda \circ H$. Notice that, almost all sample paths of the limiting process are continuous on the interval $\left[H^{-1}(p) - \epsilon, H^{-1}(q) + \epsilon\right]$. By Lemma 3.9.23 of van der Vaart and Wellner (1996), the inverse map $H \mapsto H^{-1}$ is Hadamard tangentially differentiable at H to the subspace of functions that are continuous on this interval. By the functional delta method (van der Vaart and Wellner 1996, Theorem 3.9.4) we have

$$Q_{H,a}(\cdot) \xrightarrow{d} -\frac{B_\lambda \circ H \circ H^{-1}(\cdot)}{h(H^{-1}(\cdot))} = -\frac{B_\lambda(\cdot)}{h(H^{-1}(\cdot))}$$

in $D[p, q]$. This completes the proof of the corollary. □

Remark 1. *Parallel to Remark 1 of Bickel and Freedman (1981), if $H^{-1}(0+) > -\infty$ and $H^{-1}(1) < \infty$ and h is continuous on $[H^{-1}(0+),\, H^{-1}(1)]$, the conclusion of Corollary 1 holds in $D\left[H^{-1}(0+), H^{-1}(1)\right]$. For example, if H is a uniform distribution on $[0, 1]$, then the convergence holds in $D[0, 1]$.*

The following simple example shows how a practitioner can apply Corollary 1 to special cases such as median and interquartile range.

Example 1. In this example we derive the asymptotic distribution for the median and the interquartile range for the normalized inverse-Gaussian process. Let $Q^1_{H,a}$, $Q^2_{H,a}$ and $Q^3_{H,a}$ be the first, the second (median) and the third quartiles of $P_{H,a}$ (i.e. $P^{-1}_{H,a}(0.25) = Q^1_{H,a}$, $P^{-1}_{H,a}(0.5) = Q^2_{H,a}$ and $P^{-1}_{H,a}(0.75) = Q^3_{H,a}$). Let q_1, q_2 and q_3 be the first, the second (median) and the third quartiles of H. From Corollary 1, after some simple calculations, the asymptotic distribution of the median and the interquartile range are given, respectively, by

$$\sqrt{a}\left(Q^2_{H,a} - q_2\right) \xrightarrow{d} N\left(0, \frac{1}{4h^2(q_2)}\right)$$

and

$$\sqrt{a}\left(IQR - (q_3 - q_1)\right) \xrightarrow{d} N\left(0, \frac{3}{h^2(q_3)} + \frac{3}{h^2(q_1)} - \frac{2}{h(q_1)h(q_3)}\right),$$

where $h = H'$ and $IQR = Q^3_{H,a} - Q^1_{H,a}$. Note that, the asymptotic distributions of the median and the interquartile range for the normalized inverse-Gaussian process coincide with the asymptotic distribution of the sample median and the sample interquartile range (DasGupta 2008, page 93).

4 Glivenko-Cantelli Theorem for the N-IG Process

In this section, we show that a similar form of the empirical strong law of large numbers and the empirical Glivenko-Cantelli theorem continue to hold for the normalized inverse-Gaussian process.

Theorem 2. *Let $P_{H,a} \sim$ N-IGP(a, H). Assume that $a = n^2c$, for a fixed positive number c. Then, as $n \to \infty$,*

$$P_{H,n^2c}(A) \xrightarrow{a.s.} H(A)$$

for any measurable subset A of $\mathfrak{X}$.

Proof. For any $\epsilon > 0$, by (6), we have

$$\Pr\left\{\left|P_{H,n^2c}(A) - H(A)\right| > \epsilon\right\} \leqslant \frac{H(A)(1 - H(A))}{\xi(n^2c)\epsilon^2},$$

where $\xi(n^2c)$ is defined by (3). Note that

$$\lim_{n\to\infty} \frac{\xi(n^2c)}{n^2c} = 1$$

(Abramowitz and Stegun 1972, Formula 6.5.32). Since the series $\sum_{n=1}^{\infty} 1/n^2$ converges, it follows by the limit comparison test that the series $\sum_{n=1}^{\infty} \xi(n^2c)$ is also convergent. Thus,

$$\sum_{n=1}^{\infty} \Pr\left\{|P_{H,n^2c}(A) - H(A)| > \epsilon\right\} < \infty.$$

Therefore, by the first Borel-Cantelli lemma, the proof follows. □

The next theorem establishes the Glivenko-Cantelli theorem for the normalized inverse-Gaussian process. The proof of the theorem follows by arguments similar to that given in the proof of the Glivenko-Cantelli theorem for the empirical process. See, for example, Billingsley (1995), Theorem 20.6 and for the exchangeable sequences see Berti and Rigo (1997).

Theorem 3. *Let $P_{H,a} \sim$ N-IGP(a, H). Assume that $a = n^2c$, for a fixed positive number c. Then*

$$\sup_{x\in\mathbb{R}} \left|P_{H,n^2c}(x) - H(x)\right| \xrightarrow{a.s.} 0,$$

as $n \to \infty$, where H and $P_{H,a}$ live in $(\mathfrak{X}, \mathcal{A})$.

Remark 2. *Results similar to Corollary 1 and Theorem 3 can also be established for the two-parameter Poisson-Dirichlet process. The proof will based on Theorem 4.1 and Theorem 4.2 of James (2008) and arguments analogous to that used in the present paper.*

5 Monotonically Decreasing Approximation to the N-IG Process

In next theorem, we mimic the approach of Zarepour and Al Labadi (2012) for the Dirichlet process to derive a finite sum representation which converges almost surely to the Ferguson and Klass sum representation of the normalized inverse-Gaussian process. The proof of the theorem is similar to the proof of Lemma 2 in Zarepour and Al Labadi (2012) and a simple application of the continuous mapping theorem. Hence, it is omitted.

Theorem 4. *Let $(\theta_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables with values in $\mathfrak{X}$ and common distribution H, independent of $(\Gamma_i)_{i\geqslant 1}$, then as $n \to \infty$*

$$P_{n,H,a}^{new} = \sum_{i=1}^{n} \frac{F_n^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right)}{\sum_{i=1}^{n} F_n^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right)} \delta_{\theta_i} \xrightarrow{a.s.} P_{H,a} = \sum_{i=1}^{\infty} \frac{L^{-1}(\Gamma_i)}{\sum_{i=1}^{\infty} L^{-1}(\Gamma_i)} \delta_{\theta_i}, \qquad (12)$$

where

$$F_n(x) = \int_x^{\infty} \frac{a}{n\sqrt{2\pi}} t^{-3/2} \exp\left\{-\frac{1}{2}\left(\frac{a^2}{n^2 t} + t\right) + \frac{a}{n}\right\} dt.$$

Here Γ_i and $L(x)$ are defined in (3) and (6), respectively.

Note that, for any $1 \leqslant i \leqslant n$, $\Gamma_i/\Gamma_{n+1} < \Gamma_{i+1}/\Gamma_{n+1}$ almost surely. Since F_n^{-1} is a decreasing function, we have $F_n^{-1}(\Gamma_i/\Gamma_{n+1}) > F_n^{-1}(\Gamma_{i+1}/\Gamma_{n+1})$ almost surely. That is, the weights of the new representation given in Theorem 4 decrease monotonically for any fixed positive integer n. The suggested approximation in Theorem 4 and the stick-breaking representation of the normalized inverse-Gaussian process (Favaro et al. 2012) have been studied and compared in Al Labadi and Zarepour (2012). We refer the reader to the previous paper for more details.

Remark 3. *For $P^{new}_{n,H,a}$ in Theorem 4, we can write*

$$P^{new}_{n,H,a} \stackrel{d}{=} \sum_{i=1}^{n} p_{i,n}\delta_{\theta_i}, \tag{13}$$

where $(p_{1,n}, \ldots, p_{n,n}) \sim$ N-IG$(a/n, \ldots, a/n)$, $\stackrel{d}{=}$ denotes equality in distribution and N-IG is the normalized inverse-Gaussian distribution with probability density function given in (1). Therefore, a similar result to Theorem 2 of Ishwaran and Zarepour (2002) for the normalized inverse-Gaussian process follows immediately. Similar to $P^{new}_{n,H,a}$ in (12), representation (13) can be used for simulation purposes.

6 Concluding Remarks

The approach used in this paper can be applied to some similar processes with tractable finite dimensional distributions. An interesting example is the class of the generalized Dirichlet process introduced by Lijoi et al. (2005a) and further investigated in Favaro et al. (2011). Another class of interest is the class of normalized generalized gamma processes (Lijoi et al. 2007), which contains the normalized inverse-Gaussian process as a special case ($\alpha = 1/2$). Unfortunately, the finite dimensional distributions of the normalized generalized gamma processes are unknown. Therefore, a proper modification of the approach discussed in this paper is necessary.

The results obtained in this paper can be applied to derive asymptotic properties of any Hadamard-differentiable functional of the N-IGP(a, H) as $a \to \infty$. For different applications in statistics, we refer the reader to van der Vaart and Wellner (1996), Section 3.9 and Lo (1987). Moreover, the obtained results for the univariate cumulative distribution functions can be generalized to multivariate cumulative distribution functions. To achieve these generalizations, the results of Bickel and Wichura (1971) can be employed in the proof.

References

Abramowitz, M. and Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Mineola, New York: Dover Publications. 554, 559, 564

Al Labadi, L. and Zarepour, M. (2012). "On simulation from the two-parameter Poisson-Dirichlet process and the normalized inverse-Gaussian process." *Submitted*. 560

Berti, P. and Rigo, P. (1997). "A Glivenko-Cantelli theorem for exchangeable random variables." *Statistics and Probability Letters*, 32: 385–391. 559

Bickel, P. J. and Freedman, D. A. (1981). "Some asymptotic theory for the bootstrap." *The Annals of Statistics*, 9: 1196–1217. 557

Bickel, P. J. and Wichura, M. J. (1971). "Convergence criteria for multiparameter stochastic processes and some applications." *The Annals of Mathematical Statistics*, 42: 1656–1670. 560

Billingsley, P. (1995). *Probability and Measure*. New York: John Wiley & Sons, second edition. 559

— (1999). *Convergence of Probability Measures*. New York: John Wiley & Sons, third edition. 563, 565, 568

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. New York: Springer. 558

Favaro, S., Lijoi, A., and Prünster, I. (2012). "On the stick-breaking representation of normalized inverse Gaussian priors." *Biometrika*, 99: 663–774. 553, 560

Favaro, S., Prünster, I., and Walker, S. G. (2011). "On a class of random probability measures with general predictive structure." *Scandinavian Journal of Statistics*, 38: 359–376. 560

Ferguson, T. S. (1973). "A Bayesian Analysis of Some Nonparametric Problems." *The Annals of Statistics*, 1: 209–230. 553, 557

Ferguson, T. S. and Klass, M. J. (1972). "A Representation of independent increment processes without Gaussian components." *The Annals of Mathematical Statistics*, 1: 209–230. 555

Ishwaran, H., James, L. F., and Zarepour, M. (2009). "An alternative to m out of n bootstrap." *Journal of Statistical Planning and Inference*, 39: 788–801. 555

Ishwaran, H. and Zarepour, M. (2002). "Exact and approximate sum representations for the Dirichlet process." *Canadian Journal of Statistics*, 30: 269–283. 560

James, L. F. (2008). "Large sample asymptotics for the two-parameter Poisson-Dirichlet process." In Clarke, B. and Ghosal, S. (eds.), *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, volume 3, 187–199. Cambridge, MA: MIT Press. 555, 559

James, L. F., Lijoi, A., and Prünster, I. (2006). "Conjugacy as a distinctive feature of the Dirichlet process." *Scandinavian Journal of Statistics*, 33: 105–120. 565, 568

— (2009). "Posterior analysis for normalized random measures with independent increments." *Scandinavian Journal of Statistics*, 36: 76–97. 553

Jang, G. H., Lee, J., and Lee, S. (2010). "Posterior consistency of species sampling priors." *Statistica Sinica*, 20: 581–593. 553

Kim, N. and Bickel, P. (2003). "The limit distribution of a test statistic for bivariate normality." *Statistica Sinica*, 13: 327–349. 556

Lijoi, A., Mena, R. H., and Prünster, I. (2005a). "Bayesian nonparametric analysis for a generalized Dirichlet process prior." *Statistical Inference for Stochastic Processes*, 8: 283–309. 560

— (2005b). "Hierarchical mixture modelling with normalized inverse-Gaussian priors." *Journal of the American Statistical Association*, 100: 1278–1291. 553, 554

— (2007). "Controlling the reinforcement in Bayesian nonparametric mixture models." *Journal of the Royal Statistical Society: Series B*, 69: 715–740. 553, 560

Lo, A. Y. (1987). "A large sample study of the Bayesian bootstrap." *The Annals of Statistics*, 15: 360–275. 555, 560

Nieto-Barajas, L. E. and Prünster, I. (2009). "A sensitivity analysis for Bayesian nonparametric density estimators." *Statistica Sinica*, 19: 685–705. 555

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag. 556, 568

Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley. 555

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer. 557, 560

Zarepour, M. and Al Labadi, L. (2012). "On a Rapid Simulation of the Dirichlet Process." *Statistics and Probability Letters*, 82: 916–924. 555, 559

Appendix

Proof of Lemma 1 for m = 2: Let S_1 and S_2 be any two intervals in $\mathbb{R}$. Without loss of generality, we assume that $S_1 \cap S_2 = \emptyset$. The general case when S_1 and S_2 are not necessarily disjoint follows from the continuous mapping theorem.

Note that

$$\begin{aligned}(P_{H,a}(S_1), P_{H,a}(S_2), 1 - P_{H,a}(S_1) - P_{H,a}(S_2)) \quad \sim \quad & \text{N-IG}\big(aH(S_1), aH(S_2), \\ & a(1 - H(S_1) - H(S_2))\big),\end{aligned}$$

where N-IG denotes the normalized inverse-Gaussian distribution with probability density function given in (1). For notational simplicity, set $X_i = P_{H,a}(S_i)$, $l_i = H(S_i)$ and

$D_i = \sqrt{a}\,(X_i - l_i)$ for $i = 1, 2$. Thus, the joint density function of X_1 and X_2 is

$$\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &= \frac{e^a a^3 l_1 l_2 (1-l_1-l_2)}{2^{1/2}\pi^{3/2}} \times x_1^{-3/2} x_2^{-3/2} (1-x_1-x_2)^{-3/2} \\
&\quad \times K_{-3/2}\left(a\sqrt{\frac{l_1^2}{x_1} + \frac{l_2^2}{x_2} + \frac{(1-l_1-l_2)^2}{1-x_1-x_2}} \right) \\
&\quad \times a^{-3/2} \left(\frac{l_1^2}{x_1} + \frac{l_2^2}{x_2} + \frac{(1-l_1-l_2)^2}{1-x_1-x_2} \right)^{-3/4} \\
&= \frac{a^{3/2} e^a l_1 l_2 (1-l_1-l_2)}{2^{1/2}\pi^{3/2}} \times x_1^{-3/2} x_2^{-3/2} (1-x_1-x_2)^{-3/2} \\
&\quad \times K_{-3/2}\left(a\sqrt{\frac{l_1^2}{x_1} + \frac{l_2^2}{x_2} + \frac{(1-l_1-l_2)^2}{1-x_1-x_2}} \right) \\
&\quad \times \left(\frac{l_1^2}{x_1} + \frac{l_2^2}{x_2} + \frac{(1-l_1-l_2)^2}{1-x_1-x_2} \right)^{-3/4}.
\end{aligned}$$

It follows that, the joint probability density function of $D_1 = \sqrt{a}\,(X_1 - l_1)$ and $D_2 = \sqrt{a}(X_2 - l_2)$ is

$$\begin{aligned}
f_{D_1,D_2}(y_1,y_2) &= \frac{a^{1/2} e^a l_1 l_2 (1-l_1-l_2)}{2^{1/2}\pi^{3/2}} \\
&\quad \times \left(y_1/\sqrt{a} + l_1\right)^{-3/2} \left(y_2/\sqrt{a} + l_2\right)^{-3/2} \\
&\quad \times \left(1 - y_1/\sqrt{a} - l_1 - y_2/\sqrt{a} - l_2\right)^{-3/2} \\
&\quad \times K_{-3/2}\Bigg(a \Bigg(\frac{l_1^2}{y_1/\sqrt{a} + l_1} + \frac{l_2^2}{y_2/\sqrt{a} + l_2} \\
&\quad + \frac{(1-l_1-l_2)^2}{1 - y_1/\sqrt{a} - l_1 - y_2/\sqrt{a} - l_2} \Bigg)^{1/2} \Bigg) \\
&\quad \times \Bigg(\frac{l_1^2}{y_1/\sqrt{a} + l_1} + \frac{l_2^2}{y_2/\sqrt{a} + l_2} \\
&\quad + \frac{(1-l_1-l_2)^2}{1 - y_1/\sqrt{a} - l_1 - y_2/\sqrt{a} - l_2} \Bigg)^{-3/4}.
\end{aligned}$$

By Scheffé's theorem (Billingsley 1999, page 29), it is enough to show that

$$f_{D_1,D_2}(y_1,y_2) \to f(y_1,y_2) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-(y_1\ y_2)\Sigma^{-1}(y_1\ y_2)^T/2\right\},$$

where $\Sigma = \begin{bmatrix} l_1(1-l_1) & -l_1 l_2 \\ -l_1 l_2 & l_2(1-l_2) \end{bmatrix}$.

Since, for large z and fixed ν, $K_\nu(z) \approx \sqrt{\pi/2}z^{-1/2}e^{-z}$ (Abramowitz and Stegun 1972, Formula 9.7.2) we get

$$\lim_{a\to\infty} f_{D_1,D_2}(y_1,y_2) = \lim_{a\to\infty}\Bigg[\frac{l_1 l_2(1-l_1-l_2)}{2\pi}$$
$$\times\left(\frac{l_1^2}{y_1/\sqrt{a}+l_1}+\frac{l_2^2}{y_2/\sqrt{a}+l_2}+\frac{(1-l_1-l_2)^2}{1-y_1/\sqrt{a}-l_1-y_2/\sqrt{a}-l_2}\right)^{-1}$$
$$\times\left(y_1/\sqrt{a}+l_1\right)^{-3/2}\left(y_2/\sqrt{a}+l_2\right)^{-3/2}$$
$$\times\left(1-y_1/\sqrt{a}-l_1-y_2/\sqrt{a}-l_2\right)^{-3/2}$$
$$\times\exp\Bigg(a\Bigg(1-\Bigg(\frac{l_1^2}{y_1/\sqrt{a}+l_1}+\frac{l_2^2}{y_2/\sqrt{a}+l_2}$$
$$+\frac{(1-l_1-l_2)^2}{1-y_1/\sqrt{a}-l_1-y_2/\sqrt{a}-l_2}\Bigg)^{1/2}\Bigg)\Bigg)\Bigg].$$

Notice that,

$$\frac{l_1 l_2(1-l_1-l_2)}{2\pi}$$
$$\times\left(\frac{l_1^2}{y_1/\sqrt{a}+l_1}+\frac{l_2^2}{y_2/\sqrt{a}+l_2}+\frac{(1-l_1-l_2)^2}{1-y_1/\sqrt{a}-l_1-y_2/\sqrt{a}-l_2}\right)^{-1}$$
$$\times\left(y_1/\sqrt{a}+l_1\right)^{-3/2}\left(y_2/\sqrt{a}+l_2\right)^{-3/2}\left(1-y_1/\sqrt{a}-l_1-y_2/\sqrt{a}-l_2\right)^{-3/2}$$

converges to $1/\left(2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}\right)$, where

$$\sigma_{11}=l_1(1-l_1),\quad \sigma_{22}=l_2(1-l_2),\quad \rho_{12}=-\sqrt{\frac{l_1 l_2}{(1-l_1)(1-l_2)}}.$$

To prove the lemma, it remains to show that

$$a\left(1-\left(\frac{l_1^2}{y_1/\sqrt{a}+l_1}+\frac{l_2^2}{y_2/\sqrt{a}+l_2}+\frac{(1-l_1-l_2)^2}{1-y_1/\sqrt{a}-l_1-y_2/\sqrt{a}-l_2}\right)^{1/2}\right)$$

converges to

$$-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{y_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{y_2}{\sqrt{\sigma_{22}}}\right)^2-2\rho_{12}\left(\frac{y_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{y_1}{\sqrt{\sigma_{11}}}\right)\right].$$

The last argument follows straightforwardly from L'Hospital's rule. □

Proof of Theorem 1: Let (a_n) be an arbitrary sequence such that $a_n \to \infty$. To simplify the notations, in the arguments below, we omit writing the index n of a_n. Assume first that $H(t) = \lambda(t) = t$ (i.e. λ is the Lebesgue measure on $[0,1]$). Thus the process (10) reduces to

$$D_{\lambda,a}(t) = \sqrt{a}\,(P_{\lambda,a}(t) - t)\,.$$

To prove the theorem, we use Lemma 1 and Theorem 13.5 of Billingsley (1999). Therefore, we only need to show that for any $0 \leqslant t_1 \leqslant t \leqslant t_2 \leqslant 1$,

$$E\left[|D_{\lambda,a}(t) - D_{\lambda,a}(t_1)|^{2\beta}\,|D_{\lambda,a}(t_2) - D_{\lambda,a}(t)|^{2\beta}\right] \leqslant |F(t_2) - F(t_1)|^{2\gamma}\,,$$

for some $\beta \geqslant 0$, $\gamma > 1/2$, and a nondecreasing continuous function F on $[0,1]$. Take $\beta = \gamma = 1$ and $F(t) = 2t$. Relying on the technique of James et al. (2006) for calculating moments of random probability measures, we show that

$$E\left[(D_{\lambda,a}(t) - D_{\lambda,a}(t_1))^2\,(D_{\lambda,a}(t_2) - D_{\lambda,a}(t))^2\right] \leqslant 4\,(t_2 - t_1)^2\,. \tag{14}$$

Observe that

$$D_{\lambda,a}(t) - D_{\lambda,a}(t_1) = D_{\lambda,a}((t_1,t]) \text{ and } D_{\lambda,a}(t_2) - D_{\lambda,a}(t) = D_{\lambda,a}((t,t_2]).$$

Thus, the expectation in the left hand side of (14) is equal to

$$a^2 E\left[\{P_{\lambda,a}((t_1,t]) - \lambda((t_1,t])\}^2\,\{P_{\lambda,a}((t,t_2]) - \lambda((t,t_2])\}^2\right], \tag{15}$$

where $\lambda((t,t_2]) = t_2 - t$ and $\lambda((t_1,t]) = t - t_1$. Expanding the expression

$$\{P_{\lambda,a}((t_1,t]) - \lambda((t_1,t])\}^2\,\{P_{\lambda,a}((t,t_2]) - \lambda((t,t_2])\}^2$$

gives

$$\begin{aligned}
&P^2_{\lambda,a}((t_1,t])P^2_{\lambda,a}((t,t_2]) - 2\lambda((t,t_2])P^2_{\lambda,a}((t_1,t])P_{\lambda,a}((t,t_2])\\
&+ \lambda^2((t,t_2])P^2_{\lambda,a}((t_1,t]) - 2\lambda((t_1,t])P_{\lambda,a}((t_1,t])P^2_{\lambda,a}((t,t_2])\\
&+ 4\lambda((t_1,t])\lambda((t,t_2])P_{\lambda,a}((t_1,t])P_{\lambda,a}((t,t_2]) - 2\lambda((t_1,t])\lambda^2((t,t_2])P_{\lambda,a}((t_1,t])\\
&+ \lambda^2((t_1,t])P^2_{\lambda,a}((t,t_2]) - 2\lambda^2((t_1,t])\lambda((t,t_2])P_{\lambda,a}((t,t_2]) + \lambda^2((t_1,t])\lambda^2((t,t_2]).
\end{aligned}$$

Applying the general technique of James et al. (2006) to calculate the moments yields

$$\begin{aligned}
E\left[P^2_{\lambda,a}((t_1,t])P^2_{\lambda,a}((t,t_2])\right] &= \frac{1}{48}\Big[\Big(3\Gamma(0,a)a^4e^a - \Gamma(-2,a)a^6e^a - 2a^3 + 6a + 6\Big)\\
&\times\lambda^2((t_1,t])\lambda^2((t,t_2]) + \Big(3\Gamma(-1,a)a^4e^a\\
&-\Gamma(-3,a)a^6e^a - 2a^2 + 2a + 2\Big)\lambda^2((t_1,t])\lambda((t,t_2])\\
&+\Big(3\Gamma(-1,a)a^4e^a - \Gamma(-3,a)a^6e^a - 2a^2 + 2a + 2\Big)\\
&\times\lambda((t_1,t])\lambda^2((t,t_2]) + \Big(- \Gamma(-4,a)a^6e^a\\
&+3\Gamma(-2,a)a^4e^a - 3\Gamma(0,a)a^2e^a + a + 1\Big)\\
&\times\lambda((t_1,t])\lambda((t,t_2])\Big].
\end{aligned}$$

Using the approximation (5) gives

$$\Gamma(0,a)a^4e^a \approx a^3 - a^2 + 2a - 6 + \frac{24}{a} - \frac{120}{a^2} + \cdots$$

$$\Gamma(-2,a)a^6e^a \approx a^3 - 3a^2 + 12a - 60 + \frac{360}{a} - \frac{2520}{a^2} + \cdots$$

$$\Gamma(-1,a)a^4e^a \approx a^2 - 2a + 6 - \frac{24}{a} + \frac{120}{a^2} - \cdots$$

$$\Gamma(-3,a)a^6e^a \approx a^2 - 4a + 20 - \frac{120}{a} + \frac{840}{a^2} - \cdots$$

$$\Gamma(-4,a)a^6e^a \approx a - 5 + \frac{30}{a} - \frac{210}{a^2} + \cdots$$

$$\Gamma(-2,a)ae^a \approx a - 3 + \frac{12}{a} - \frac{60}{a^2} + \cdots$$

$$\Gamma(0,a)a^2e^a \approx a - 1 + \frac{2}{a} - \frac{6}{a^2} + \cdots .$$

Therefore,

$$\begin{aligned} E\left[P^2_{\lambda,a}((t_1,t])P^2_{\lambda,a}((t,t_2])\right] \quad &\approx \quad \Big(1 - \frac{6}{a} + \frac{45}{a^2}\Big)\lambda^2((t_1,t])\lambda^2((t,t_2]) \\ &\quad + \Big(\frac{1}{a} - \frac{10}{a^2}\Big)\lambda^2((t_1,t])\lambda((t,t_2]) \\ &\quad + \Big(\frac{1}{a} - \frac{10}{a^2}\Big)\lambda((t_1,t])\lambda^2((t,t_2]) \\ &\quad + \frac{1}{a^2}\lambda((t_1,t])\lambda((t,t_2]). \end{aligned} \tag{16}$$

Likewise,

$$\begin{aligned} E\left[P^2_{\lambda,a}((t_1,t])P_{\lambda,a}((t,t_2])\right] \quad &= \quad \frac{1}{8}\Big[\Big(\Gamma(-1,a)a^4e^a - a^2 + 2a + 2\Big)\lambda^2((t_1,t])\lambda((t,t_2]) \\ &\quad + \Big(\Gamma(-2,a)a^4e^a - 2\Gamma(0,a)a^2e^a + a + 1\Big) \\ &\quad \times \lambda((t_1,t])\lambda((t,t_2])\Big] \\ &\approx \quad \Big(1 - \frac{3}{a} + \frac{15}{a^2}\Big)\lambda^2((t_1,t])\lambda((t,t_2]) \\ &\quad + \Big(\frac{1}{a} - \frac{6}{a^2}\Big)\lambda((t_1,t])\lambda((t,t_2]) \end{aligned} \tag{17}$$

and

$$E\left[P_{\lambda,a}((t_1,t])P^2_{\lambda,a}((t,t_2])\right] = \frac{1}{8}\Big[\Big(\Gamma(-1,a)a^4e^a - a^2 + 2a + 2\Big)\lambda((t_1,t])\lambda^2((t,t_2]) + \Big(\Gamma(-2,a)a^4e^a - 2\Gamma(0,a)a^2e^a + a + 1\Big)\lambda((t_1,t])\lambda((t,t_2])\Big]$$
$$\approx \Big(1 - \frac{3}{a} + \frac{15}{a^2}\Big)\lambda((t_1,t])\lambda^2((t,t_2]) + \Big(\frac{1}{a} - \frac{6}{a^2}\Big)\lambda((t_1,t])\lambda((t,t_2]). \tag{18}$$

On the other hand, by (2) and (4), the expectation of the expression

$$\lambda^2((t,t_2])P^2_{\lambda,a}((t_1,t]) + 4\lambda((t_1,t])\lambda((t,t_2])P_{\lambda,a}((t_1,t])P_{\lambda,a}((t,t_2]) - 2\lambda((t_1,t])\lambda^2((t,t_2])P_{\lambda,a}((t_1,t]) + \lambda^2((t_1,t])P^2_{\lambda,a}((t,t_2]) - 2\lambda^2((t_1,t])\lambda((t,t_2])P_{\lambda,a}((t,t_2]) + \lambda^2((t_1,t])\lambda^2((t,t_2])$$

simplifies to

$$3\lambda^2((t_1,t])\lambda^2((t,t_2]) - 6\Gamma(-2,a)a^2e^a\lambda^2((t_1,t])\lambda^2((t,t_2]) + \Gamma(-2,a)a^2e^a\lambda^2((t_1,t])\lambda((t,t_2]) + \Gamma(-2,a)a^2e^a\lambda((t_1,t])\lambda^2((t,t_2])$$
$$\approx 3\lambda^2((t_1,t])\lambda^2((t,t_2]) - 6\left(\frac{1}{a} - \frac{3}{a^2}\right)\lambda^2((t_1,t])\lambda^2((t,t_2]) + \left(\frac{1}{a} - \frac{3}{a^2}\right)\lambda^2((t_1,t])\lambda((t,t_2]) + \left(\frac{1}{a} - \frac{3}{a^2}\right)\lambda((t_1,t])\lambda^2((t,t_2]). \tag{19}$$

By (16), (17), (18) and (19), we have

$$E\left[\{P_{\lambda,a}((t_1,t]) - \lambda((t_1,t])\}^2\{P_{\lambda,a}((t,t_2]) - \lambda((t,t_2])\}^2\right] \approx \frac{3}{a^2}\lambda^2((t_1,t])\lambda^2((t,t_2]) - \frac{1}{a^2}\lambda^2((t_1,t])\lambda((t,t_2]) - \frac{1}{a^2}\lambda((t_1,t])\lambda^2((t,t_2]) + \frac{1}{a^2}\lambda((t_1,t])\lambda((t,t_2])$$
$$\leqslant \frac{4}{a^2}\lambda((t_1,t])\lambda((t,t_2]). \tag{20}$$

Thus, by (15) and (20), we have

$$E\left[\left(D_{\lambda,a}(t) - D_{\lambda,a}(t_1)\right)^2\left(D_{\lambda,a}(t_2) - D_{\lambda,a}(t)\right)^2\right] \leqslant 4\lambda(t_1,t]\lambda(t,t_2] = 4\left(t - t_1\right)\left(t_2 - t\right) \leqslant 4\left(t_2 - t_1\right)^2,$$

for $0 \leqslant t_1 \leqslant t \leqslant t_2 \leqslant 1$. This proves the theorem in the case when $H(t) = t$, i.e. H is the uniform distribution. Observe that, the quantile function $H^{-1}(s) = \inf\{t : H(t) \geqslant s\}$ has the property: $H^{-1}(s) \leqslant t$ if and only if $s \leqslant H(t)$. If U_i is uniformly distributed over $[0,1]$, then $H^{-1}(U_i)$ has distribution H. Thus, we can use the representation

$a_i = H^{-1}(U_i)$, where $(U_i)_{i\geqslant 1}$ is a sequence of i.i.d. random variables with uniform distribution on $[0,1]$, to have

$$P_{H,a}(t) = P_{\lambda,a}(H(t)) \quad \text{and} \quad D_{H,a}(t) = D_{\lambda,a}(H(t)) = D_{\lambda,a} \circ H(t), \quad t \in \mathbb{R},$$

where $P_{\lambda,a}$ is the normalized inverse-Gaussian process with concentration parameter a and Lebesgue base measure λ on $[0,1]$. From the uniform case, which was already treated, we have $D_{\lambda,a}(\cdot) = \sqrt{a}\,(P_{\lambda,a}(\cdot) - \lambda(\cdot)) \xrightarrow{d} B_{\lambda}(\cdot)$. Define $\Psi : D[0,1] \to D[-\infty,\infty]$ by $(\Psi x)(t) = x(H(t))$. Since the function Ψ is uniformly continuous (Billingsley 1999, page 150; Pollard 1984, page 97), it follows, from the continuous mapping theorem and the fact that $D_{\lambda,a} \xrightarrow{d} B_{\lambda}$, that $D_{H,a} = \Psi(D_{\lambda,a}) \xrightarrow{d} \Psi(B_{\lambda}) = B_H$. This completes the proof of the theorem. □

Acknowledgments

The authors thank the Editor, the Associate Editor and a referee for their important and constructive comments that led to significant improvement of the paper. In particular, their remarks about applying the technique of James et al. (2006) for calculating the moments of the normalized inverse-Gaussian process is highly appreciated. We also thank 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)**.