ISSN 1066-5307, Mathematical Methods of Statistics, 2017, Vol. 26, No. 3, pp. 212–225. © Allerton Press, Inc., 2017.

Two-Sample Kolmogorov-Smirnov Test Using a Bayesian Nonparametric Approach

L. Al-Labadi[1*] and M. Zarepour[2]**

[1]*Dept. Math. & Comput. Sci., Univ. of Toronto Mississauga, Canada*

[2]*Dept. Math. and Statist., Univ. of Ottawa, Ottawa, Canada*

Received November 24, 2016; in final form, May 22, 2017

Abstract—In this paper, a Bayesian nonparametric approach to the two-sample problem is proposed. Given two samples $\mathbf{X} = X_1, \ldots, X_{m_1} \overset{i.i.d.}{\sim} F$ and $\mathbf{Y} = Y_1, \ldots, Y_{m_2} \overset{i.i.d.}{\sim} G$, with F and G being unknown continuous cumulative distribution functions, we wish to test the null hypothesis $\mathcal{H}_0\colon F = G$. The method is based on computing the Kolmogorov distance between two posterior Dirichlet processes and comparing the results with a reference distance. The parameters of the Dirichlet processes are selected so that any discrepancy between the posterior distance and the reference distance is related to the difference between the two samples. Relevant theoretical properties of the procedure are also developed. Through simulated examples, the approach is compared to the frequentist Kolmogorov–Smirnov test and a Bayesian nonparametric test in which it demonstrates excellent performance.

Keywords: Dirichlet process, goodness-of-fit tests, Kolmogorov distance, two-sample problem.

2000 Mathematics Subject Classification: 62F15, 62N03.

DOI: 10.3103/S1066530717030048

1. INTRODUCTION

Two-sample comparison is a common problem in statistics. Namely, given two samples $\mathbf{X} = X_1, \ldots, X_{m_1} \overset{i.i.d.}{\sim} F$ and $\mathbf{Y} = Y_1, \ldots, Y_{m_2} \overset{i.i.d.}{\sim} G$, with F and G being unknown continuous cumulative distribution functions (cdf's), the problem is to decide whether $F = G$. Although there have been many procedures developed for the two-sample problem, the approach considered in this paper is Bayesian in nature. First, two Dirichlet processes $DP(a_1, H_1)$ and $DP(a_2, H_2)$ are considered as priors for F and G. Then the distance between the two processes is compared with a reference distance. The parameters of the Dirichlet processes are chosen so that any disagreement between the posterior distance and the reference distance is a consequence of the difference between the two samples.

Recently, there has been considerable interest in developing Bayesian nonparametric techniques for hypothesis testing. Most of these include goodness-of-fit tests for one-sample problems. Two standard nonparametric Bayesian approaches for one-sample goodness-of-fit tests can be found in the literature. The first approach consists in embedding the proposed model in the null hypothesis into a larger family of models (the alternative family). Following this step, a prior is placed on the alternative family. Then, the Bayes factor of the null hypothesis to the alternative is computed. For example, see Carota and Parmigiani [12] and Florens, Richard, and Rolin [17] used a Dirichlet process prior for the alternative distribution. McVinish, Rousseau, and Mengersen [29] considered mixtures of triangular distributions. Another form of the prior, the Pólya tree process [28], was suggested by Berger and Guglielmi [7]. The second approach for one-sample goodness-of-fit tests is based on placing a prior on the true distribution generating the data. For this test, the distance between the posterior distribution and the proposed one is measured. Muliere and Tardella [30], Swartz [33], Al-Labadi and Zarepour [3, 6] considered the Dirichlet process and used the Kolmogorov distance to derive a goodness-of-fit test for continuous distributions.

*E-mail: luai.allabadi@utoronto.ca

**E-mail: zarepour@uottawa.ca

Viele [35] used the Dirichlet process and the Kullback-Leibler distance to test only discrete distributions. Explicit expressions for calculating the different types of distance between the Dirichlet process and its base measure were derived in [6]. On the other hand, Hsieh [22] used the Pólya tree prior and the Kullback–Leibler distance to test continuous distributions. As for two-sample tests, Holmes, Caron, Griffin, and Stephens [23] developed a way to compute the Bayes factor for testing the null hypothesis through the marginal likelihood of the data with Pólya tree priors centered either subjectively or using an empirical procedure. Under the null hypothesis, they modeled the two samples to come from a single random measure distributed as a Pólya tree, whereas under the alternative hypothesis the two samples come from two separate Pólya tree random measures. Ma and Wong [28] allowed the two distributions to be generated jointly through optional coupling of a Pólya tree prior. Borgwardt and Ghahramani [11] discussed two-sample tests based on Dirichlet process mixture models and derived a formula to compute the Bayes factor in this case. Generalizations of the Bayes factor approach based on Pólya tree priors to censored and multivariate data were proposed in [13]. Huang and Ghosh [24] considered the two-sample hypothesis testing problems under Pólya tree priors and Lehmann alternatives. Recently, Shang and Reilly [32] introduced a class of tests, which use the connection between the Dirichlet process prior and the Wilcoxon rank sum test. They also extend their idea using the Dirichlet process mixture prior and developed a Bayesian counterpart to the Wilcoxon rank sum statistic and the weighted log rank statistic for right and interval censored data.

Note that, the two-sample Bayesian nonparametric tests based on the distance approach are not found in the literature. The method proposed in this paper is considered the first endeavor in this direction.

This paper is structured as follows. In Section 2, the Dirichlet process prior $DP(a, H)$ is briefly reviewed. In Section 3, the Kolmogorov distance between two Dirichlet processes is considered and several of its theoretical properties are derived. The proposed approach is developed in Section 4. It also addresses how to choose parameters of the Dirichlet processes. In Section 5, illustrative examples and simulation results are included. Some additional properties of the proposed approach are discussed in Section 6. Finally, some concluding remarks are made in Section 7.

2. THE DIRICHLET PROCESS

Before proceeding to describe the Bayesian nonparametric approach to the two-sample problem, it is necessary to provide a short introduction of the Dirihclet process. The Dirichlet process, formally introduced in [16], is the most well-known and widely used prior in Bayesian nonparametric inference. Consider a space $\mathfrak{X}$ with a $\sigma-$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 [16], a random probability measure $P = \{P(A)\}_{A\in\mathcal{A}}$ is called a Dirichlet process on $(\mathfrak{X}, \mathcal{A})$ with parameters a and H, if for any finite measurable partition $\{A_1, \ldots, A_k\}$ of $\mathfrak{X}$, the joint distribution of the vector $(P(A_1), \ldots P(A_k))$ has the Dirichlet distribution with parameters $(aH(A_1), \ldots, aH(A_k))$, where $k \geq 2$. We assume that if $H(A_j) = 0$, then $P(A_j) = 0$ with probability one. If P is a Dirichlet process with parameters a and H, we write $P \sim DP(a, H)$. The parameter a is known as the *concentration parameter* and the probability measure H is called the *base*(*centering*) *measure* of P.

An attractive feature of the Dirichlet process is the conjugacy property. If $X_1, \ldots, X_m$ is a sample from $P \sim DP(a, H)$, then the posterior distribution of P given $X_1, \ldots, X_m$ coincides with the distribution of the Dirichlet process with parameters a^* and H^*, where

$$a^* = a + m \qquad \text{and} \qquad H^*_m = \frac{a}{a+m}H + \frac{m}{a+m}F_m. \tag{1}$$

Here and throughout the paper, $F_m = \sum_{i=1}^{m} \delta_{X_i}/m$ is the empirical distribution and δ_{X_i} denotes the Dirac measure at X_i. We also use a “$*$" as a superscript to denote posterior quantities. Notice that the posterior base distribution H^* is a convex combination of the base distribution and the empirical distribution. The weight associated with the prior base distribution H is proportional to a, while the weight associated with the empirical distribution is proportional to the number of observations m. The posterior base distribution H^* approaches the prior base measure H for large values of a. On the other hand, for small values of a, H^* is close to the empirical distribution.

Following [16], $P \sim DP(a, H)$ has the following series representation

$$P = \sum_{i=1}^{\infty} J_i \delta_{Y_i}, \tag{2}$$

where $\Gamma_i = E_1 + \cdots + E_i$, $E_i \overset{i.i.d.}{\sim}$ exponential(1), $Y_i \overset{i.i.d.}{\sim} H$ independent of Γ_i, $L(x) = a \int_x^{\infty} t^{-1} e^{-t} dt$, $x > 0$, $L^{-1}(y) = \inf\{x > 0\colon L(x) \geq y\}$ and $J_i = L^{-1}(\Gamma_i) / \sum_{i=1}^{\infty} L^{-1}(\Gamma_i)$.

It follows clearly from (2) that a realization of the Dirichlet process is a discrete probability measure. This is true even when the base measure is absolutely continuous [9]. We point out that this discreteness property of the Dirichlet process has no more troublesome than that of the empirical process. Also, since data is always measured to finite accuracy, the true distribution being sampled from is discrete. This makes the discreteness property of P with no practical significant limitation. By imposing the weak topology, the support for the Dirichlet process is quite large. Specifically, the support for the Dirichlet process is the set of all probability measures whose support is contained in the support of the base measure. This means that whenever the support of the base measure is $\mathfrak{X}$, the space of all probability measures is the support of the Dirichlet process. See [16] and [18] for further discussion about the support of the Dirichlet process. In practice, it is difficult to work with (2) because there is no tractable form for the Lévy measure L and determining the random weights in the sum requires the computation of an infinite sum. Recently, Zarepour and Al-Labadi [37] derived an efficient approximation of the Dirichlet process with monotonically decreasing weights. Specifically, let X_n be a random variable with a Gamma$(a/n, 1)$ distribution. Define

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

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

$$P_n = \sum_{i=1}^{n} \frac{G_n^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right)}{\sum_{i=1}^{n} G_n^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right)} \delta_{\theta_i} \tag{3}$$

converges almost surely (a.s.) to P defined by (2). Note that $G_n^{-1}(p)$ is the $(1-p)$th quantile of the gamma$(a/n, 1)$ distribution. This provides the following algorithm.

Based on representation (3), the following algorithm outlines the steps required to generate a sample from the approximate Dirichlet process with parameters a and H:

Algorithm A: Simulating an approximation for the Dirichlet process

1. Fix a relatively large positive integer n.

2. Generate $\theta_i \overset{\text{i.i.d.}}{\sim} H$ for $i = 1, \ldots, n$.

3. For $i = 1, \ldots, n+1$, generate E_i from an exponential distribution with mean 1, independent of $(\theta_i)_{1 \leq i \leq n}$ and let $\Gamma_i = E_1 + \cdots + E_i$.

4. For $i = 1, \ldots, n$, compute $G_n^{-1}(\Gamma_i / \Gamma_{n+1})$.

5. Use representation (3) to find an approximate sample of the Dirichlet process.

For other simulation methods of the Dirichlet process, consult [10, 31, 36].

3. KOLMOGOROV DISTANCE

A well-known distance between two distributions is the Kolmogorov distance. For cdf's F and G this is defined as

$$d(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|.$$

Note that other distances such as the Cramér–von Mises distance and the Anderson–Darling distance could be employed in our approach, see [19].

The following lemma demonstrates that, as sample sizes get large, the Kolmogorov distance between posterior distributions of Dirichlet processes converges to the Kolmogorov distance between the true (population) distributions generated the data.

Lemma 1. *Let* $\mathbf{X} = X_1, \ldots, X_{m_1} \overset{i.i.d.}{\sim} F$ *and* $\mathbf{Y} = Y_1, \ldots, Y_{m_2} \overset{i.i.d.}{\sim} G$, *with* F *and* G *being continuous cdf's. Let* $P \sim DP(a_1, H_1)$ *and* $Q \sim DP(a_2, H_2)$. *Let* $P^* = P \mid \mathbf{X}$ *and* $Q^* = Q \mid \mathbf{Y}$. *Then, as* $m_1, m_2 \to \infty$, $d(P^*, Q^*) \overset{a.s.}{\to} d(F, G)$.

Proof. From the triangle inequality we have

$$\begin{aligned} d(P^*_{m_1}, Q^*_{m_2}) &\leq d(P^*_{m_1}, H^*_{m_1}) + d(Q^*_{m_2}, H^*_{m_1}) \\ &\leq d(P^*_{m_1}, H^*_{m_1}) + d(Q^*_{m_2}, H^*_{m_2}) + d(H^*_{m_1}, H^*_{m_2}). \end{aligned}$$

It follows that,

$$d(P^*_{m_1}, Q^*_{m_2}) - d(H^*_{m_1}, H^*_{m_2}) \leq d(P^*_{m_1}, H^*_{m_1}) + d(Q^*_{m_2}, H^*_{m_2}).$$

Similarly,

$$\begin{aligned} d(H^*_{m_1}, H^*_{m_2}) &\leq d(P^*_{m_1}, H^*_{m_1}) + d(P^*_{m_1}, H^*_{m_2}) \\ &\leq d(P^*_{m_1}, H^*_{m_1}) + d(Q^*_{m_2}, H^*_{m_2}) + d(P^*_{m_1}, Q^*_{m_2}), \end{aligned}$$

$$d(H^*_{m_1}, H^*_{m_2}) - d(P^*_{m_1}, Q^*_{m_2}) \leq d(P^*_{m_1}, H^*_{m_1}) + d(Q^*_{m_2}, H^*_{m_2}).$$

Therefore,

$$|d(P^*_{m_1}, Q^*_{m_2}) - d(H^*_{m_1}, H^*_{m_2})| \leq d(P^*_{m_1}, H^*_{m_1}) + d(Q^*_{m_2}, H^*_{m_2}). \tag{4}$$

The proof of the lemma is complete since, as $m_1, m_2 \to \infty$, the right-hand side of (4) converges to zero [31, 4] and $d(H^*_{m_1}, H^*_{m_2}) \to d(F, G)$ by the continuous mapping theorem and Polyá's theorem [14]. □

Corollary 2. *Under the null hypothesis* $\mathcal{H}_0\colon F = G$, $d(P^*_{m_1}, Q^*_{m_2}) \to 0$ *as* $m_1, m_2 \to \infty$.

The following result allows the use of the approximation to the Dirichlet process when considering the prior and posterior distributions of the Kolmogorov distance.

Lemma 3. *Let* $P \sim DP(a_1, H_1)$ *and* $Q \sim DP(a_2, H_2)$. *Let* P_{n_1} *and* Q_{n_2} *be two approximations of* P *and* Q, *respectively, as defined in* (3). *Then, as* $n_1, n_2 \to \infty$, $d(P_{n_1}, Q_{n_2}) \overset{a.s.}{\to} d(P, Q)$.

Proof. Similarly to the proof of Lemma 1, the result follows since

$$\begin{aligned} |d(P_{n_1}, Q_{n_2}) - d(P, Q)| &\leq d(P_{n_1}, P) + d(Q_{n_2}, Q) \\ &\leq d(P_{n_1}, H_1) + d(P, H_1) + d(Q_{n_2}, H_2) + d(Q, H_2). \end{aligned}$$

Now, by Corollary 4.2 in [6], the right-hand side converges to zero as $n_1, n_2 \to \infty$. □

The next lemma shows that the Kolmogorov distance between two Dirichlet processes is independent of the base measures when they are identical. This result will play a key role in the proposed approach.

Lemma 4. *Let* $P \sim DP(a_1, H_1)$ *and* $Q \sim DP(a_2, H_2)$, *where* H_1 *and* H_2 *are continuous cumulative distribution functions. If* $H_1 = H_2$, *then the distribution of* $d(P, Q)$ *does not depend on* H_1 *and* H_2.

Proof. Since H_1 is nondecreasing, we have

$$\theta_i < t \quad \text{if and only if} \quad H_1(\theta_i) < H_1(t).$$

It follows from (2) that

$$P(t) = P\big((-\infty, t]\big) = \sum_{i=1}^{\infty} J_i \delta_{\theta_i}\big((-\infty, t]\big) = \sum_{i=1}^{\infty} J_i \delta_{H_1(\theta_i)}\big((0, H_1(t)]\big).$$

Observe that, since $(\theta_i)_{i\geq 1}$ is a sequence of i.i.d. random variables with continuous distribution H_1, for $i \geq 1$, we have $U_i \overset{d}{=} H_1(\theta_i)$, where $(U_i)_{i\geq 1}$ is a sequence of i.i.d. random variables with a uniform distribution on $[0,1]$. Hence, $P(t) = P_\lambda(H_1(t))$, where $P_\lambda \sim DP(a_1, \lambda)$ and λ is the Lebesgue measure on $[0,1]$. Similarly, $Q(t) = Q_\lambda(H_2(t))$, where $Q_\lambda \sim DP(a_2, \lambda)$. Thus,

$$d(P, Q) = \sup_{t\in\mathbb{R}} |P(t) - Q(t)| = \sup_{t\in\mathbb{R}} |P_\lambda(H_1(t)) - Q_\lambda(H_2(t))|.$$

If $H_1 = H_2$, and since they are continuous, we have

$$d(P, Q) = \sup_{t\in\mathbb{R}} \big|P_\lambda(H_1(t)) - Q_\lambda(H_2(t))\big| = \sup_{0\leq z\leq 1} \big|P_\lambda(z) - Q_\lambda(z)\big|.$$

This shows that the distribution of $d(P, Q)$ does not depend on the base measures H_1 and H_2 whenever $H_1 = H_2$. □

4. A BAYESIAN NONPARAMETRIC TWO-SAMPLE TEST

Given two samples $\mathbf{X} = X_1, \ldots, X_{m_1} \overset{i.i.d.}{\sim} F$ and $\mathbf{Y} = Y_1, \ldots, Y_{m_2} \overset{i.i.d.}{\sim} G$ with F and G being unknown continuous cumulative distribution functions, we want to test the null hypothesis $\mathcal{H}_0\colon F = G$. To this end, we use the priors $P \sim DP(a_1, H_1)$ and $Q \sim DP(a_2, H_2)$ so, by (1), $P^*_{m_1} = P \mid \mathbf{X} \sim DP(a_1 + m_1, H_1^*)$ and $Q^*_{m_2} = Q \mid \mathbf{Y} \sim DP(a_2 + m_2, H_2^*)$. By Lemma 1, $d(P^*_{m_1}, Q^*_{m_2})$ almost surely approximates $d(F, G)$. Thus, it seems reasonable to use a test procedure that specifies rejecting the null hypothesis when the distance between $P^*_{m_1}$ and $Q^*_{m_1}$ is large. Specifically, we reject $\mathcal{H}_0$ if, for some cut-off number U,

$$E^*\big(d(P^*_{m_1}, Q^*_{m_2})\big) > U, \tag{5}$$

where E^* is the expectation with respect to posterior probability measures.

The first step in computing (5) is to pick parameters for the two priors $DP(a_1, H_1)$ and $DP(a_2, H_2)$. Note that, by Lemma 4, it is necessary to set $H_1 = H_2$ for the approach to be independent of the choice of H_1 and H_2. Another important reason of this choice is to avoid prior-data conflict as discussed, for example, in [15] and [1]. Prior-data conflict here means that $DP(a_1, H_1)$ lies in the "tails" of $DP(a_2, H_2)$. Prior-data conflict will occur whenever there is only a tiny overlap between the effective support regions of P and Q. In this context, the existence of prior-data conflict can yield a failure in computing the distribution of $d(P, Q)$. To avoid this problem it is necessary that H_1 and H_2 share the same effective support (note that, P and Q have the same support as H_1 and H_2, respectively). This can certainly be secured by setting $H_1 = H_2$. The effect of prior-data conflict is demonstrated in Section 5, Table 2. On the other hand, by Lemma 4, the distribution of the distance $d(P, Q)$ is independent from the choice of the base measures. For simplicity we suggest setting $H_1 = H_2 = N(0, 1)$, although other choices are certainly possible.

The selection of a_1 and a_2 is also important. It is possible to consider several values of a_1 and a_2. In general, it is highly recommended to choose $a_i \leq 0.5m_i$, $i = 1, 2$, otherwise the prior may become too influential. For example, setting $a_i = 0.5m_i$ in (1) gives

$$H^*_{m_i} = \frac{1}{3}H + \frac{2}{3}F_{m_i}, \tag{6}$$

which means the chance to draw a sample from the collected data is two times of the chance to generate a new value from H. It is noticed that, for most purposes, setting $a_1 = a_2 = 1$ is considered adequate.

This issue is further explored in Table 1 in Section 5, where the sensitivity of approach with respect to the choice of the concentration parameter of the Dirichlet process is discussed.

To compute the threshold U, in an old version of the paper [2], suggested to set U to be $d_{0.975}$, the 97.5% quantile of a "reference" distance between the following two Dirchlet processes

$$P^r_{m_1} \sim DP\big(1+m_1, H=N(0,1)\big) \quad \text{and} \quad Q^r_{m_2} \sim DP\big(1+m_2, H=N(0,1)\big). \tag{7}$$

The proposed forms of the Dirichlet processes in (7) ensure that any discrepancy between the distance $d(P^r_{m_1}, Q^r_{m_2})$ and the posterior distance $d(P^*_{m_1}, Q^*_{m_2})$ is due to the difference between the distributions of the two samples. To clarify this point, notice that any Dirichlet process is centered about its mean. The deviation from the mean is controlled via the concentration parameter. Now, $P^r_{m_1}$ and $Q^r_{m_2}$ have the same concentration parameters as $P^*_{m_1}$ and $Q^*_{m_2}$, respectively. Also, by Lemma 4, since the distribution of the distance is independent of the base measures, we expect that $E(d(P^r_{m_1}, Q^r_{m_2})) \approx E^*\big(d(P^*_{m_1}, Q^*_{m_2})\big)$ a.s. whenever $F = G$. See also Proposition 5 below for the consistency of the approach. It follows that, we reject $\mathcal{H}_0$ if

$$E\big(d(P^*_{m_1}, Q^*_{m_1})\big) > d_{0.975} \tag{8}$$

and we do not reject $\mathcal{H}_0$ otherwise. Clearly, by Lemma 4, $d_{0.975}$ does not depend on the choice of the base measures.

Instead of computing $d_{0.975}$ in (8) it seems more rational to use a test statistic that compares the expectation of the posterior distance with the expectation of the prior distance. That is, reject the null hypothesis if

$$\big|E^*\big(d(P^*_{m_1}, Q^*_{m_2})\big) - E\big(d(P^r_{m_1}, Q^r_{m_2})\big)\big| > \epsilon \tag{9}$$

and do not reject the null hypothesis otherwise. One can set values of ϵ by simulating data under the null hypothesis and then take the empirical 0.95 quantile of the distribution of $\big|E^*\big(d(P^*_{m_1}, Q^*_{m_2})\big) - E\big(d(P^r_{m_1}, Q^r_{m_2})\big)\big|$. This method is also used in [23] and is known as "the Bayes, non-Bayes compromise" described in [20]. See also Fig. 1 for histograms of samples from the distribution of $\big|E^*\big(d(P^*_{m_1}, Q^*_{m_2})\big) - E\big(d(P^r_{m_1}, Q^r_{m_2})\big)\big|$, when $m_1 = m_2 = 50$ and several values of a_1 and a_2.

The following proposition establishes the consistency of the approach to the two-sample problem as sample sizes increase. So the procedure performs correctly as sample sizes increase when $\mathcal{H}_0$ is true.

Proposition 5. *Let $P^r_{m_1}$ and $Q^r_{m_2}$ be as defined in* (7). *As $m_1, m_2 \to \infty$,* (i) *if $\mathcal{H}_0$ is true, then*

$$\big|E^*\big(d(P^*_{m_1}, Q^*_{m_2}) - E(d(P^r_{m_1}, Q^r_{m_2}))\big)\big| \xrightarrow{a.s.} 0$$

and (ii) *if $\mathcal{H}_0$ is false, then*

$$\big|E^*\big(d(P^*_{m_1}, Q^*_{m_2})\big) - E\big(d(P^r_{m_1}, Q^r_{m_2})\big)\big| \xrightarrow{a.s.} c > 0.$$

Proof. Note that, by the triangle inequality, we have

$$d(P^r_{m_1}, Q^r_{m_2}) \leq d(P^r_{m_1}, H) + d(Q^r_{m_1}, H).$$

The right-hand side of this inequality converges a.s. to 0 [25, 4]. That is, $d(P^r_{m_1}, Q^r_{m_2}) \xrightarrow{a.s.} 0$. By Corollary 2, $d(P^*_{m_1}, Q^*_{m_2}) \xrightarrow{a.s.} 0$. Now (i) follows from the continuous mapping theorem and Theorem 25.8 in [8]. To prove (ii), notice that, if $\mathcal{H}_0$ is false, then

$$d(P^*_{m_1}, Q^*_{m_2}) \xrightarrow{a.s.} d(F, G) = c > 0.$$

By arguments similar to (i), we get (ii). □

The following algorithm summarizes the steps required for a Bayesian nonparametric test for the two-sample problem.

Algorithm B:
Bayesian nonparametric test for two samples $\mathbf{X} = X_1, \ldots, X_{m_1}$ and $\mathbf{Y} = Y_1, \ldots, Y_{m_2}$

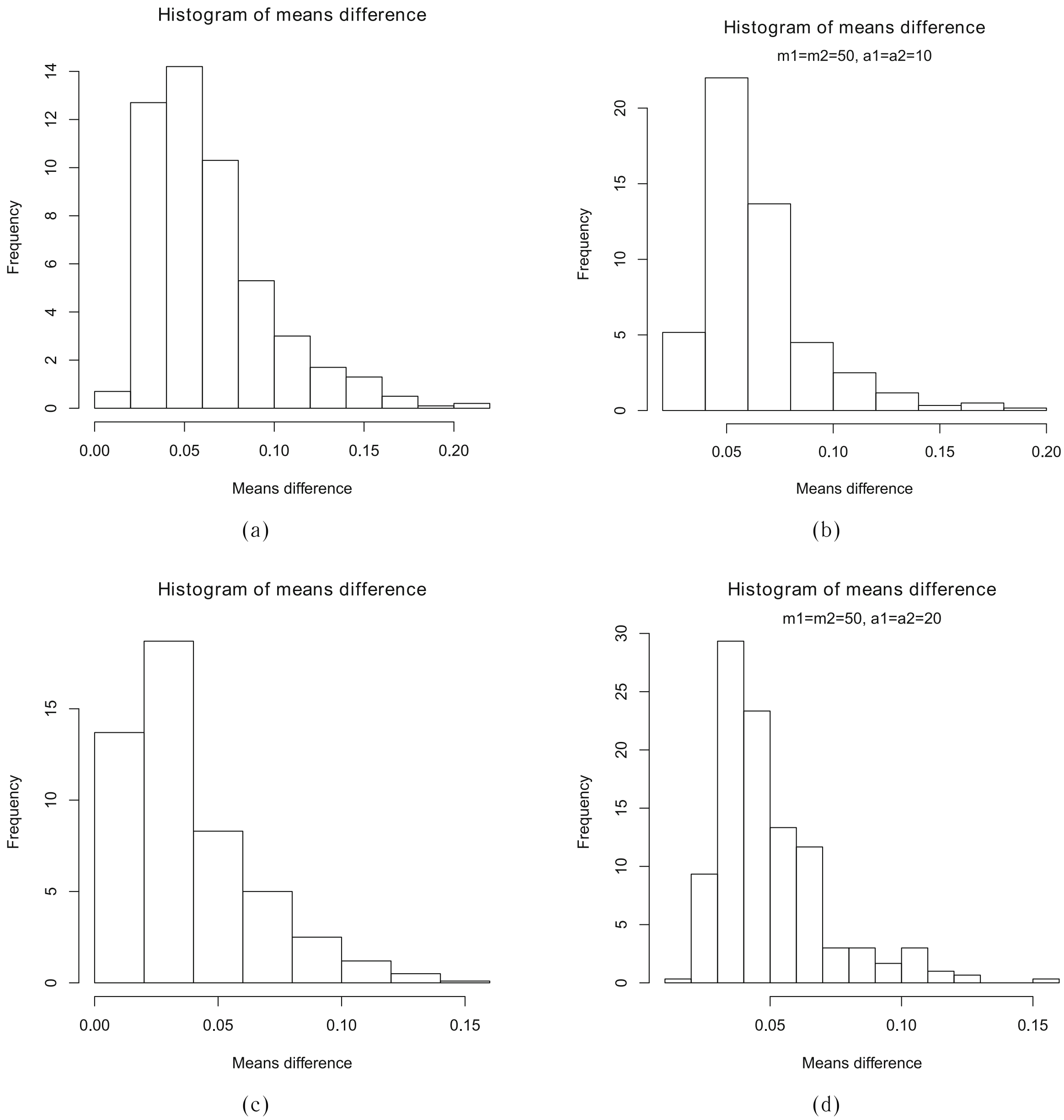

Fig. 1. Histograms of samples from the distribution of $|E^*(d(P^*_{m_1}, Q^*_{m_2})) - E(d(P^r_{m_1}, Q^r_{m_2}))|$, when $m_1 = m_2 = 50$. (a) $a_1 = a_2 = 1$; (b) $a_1 = a_2 = 5$; (c) $a_1 = a_2 = 10$; (d) $a_1 = a_2 = 20$.

1. Use Algorithm A to generate approximate samples from the two posterior processes

$$P^*_{m_1} = P \mid \mathbf{X} \sim DP\Big(a_1 + m_1, \frac{a_1}{a_1 + m_1} H_1 + \frac{1}{a_1 + m_1} \sum_{i=1}^{m_1} \delta_{X_i}\Big)$$

and

$$Q^*_{m_2} = Q \mid \mathbf{Y} \sim DP\Big(a_2 + m_2, \frac{a_2}{a_2 + m_2} H_2 + \frac{1}{a_2 + m_2} \sum_{i=1}^{m_2} \delta_{Y_i}\Big).$$

Here we recommend to take $a_1 = a_2 = 1$ and $H_1 = H_2 = N(0, 1)$. Denote the approximated processes of $P^*_{m_1}$ and $Q^*_{m_2}$ by $P^*_{n_1,m_1}$ and $Q^*_{n_2,m_2}$, respectively.

2. Compute $d(P^*_{n_1,m_1}, Q^*_{n_2,m_2})$.

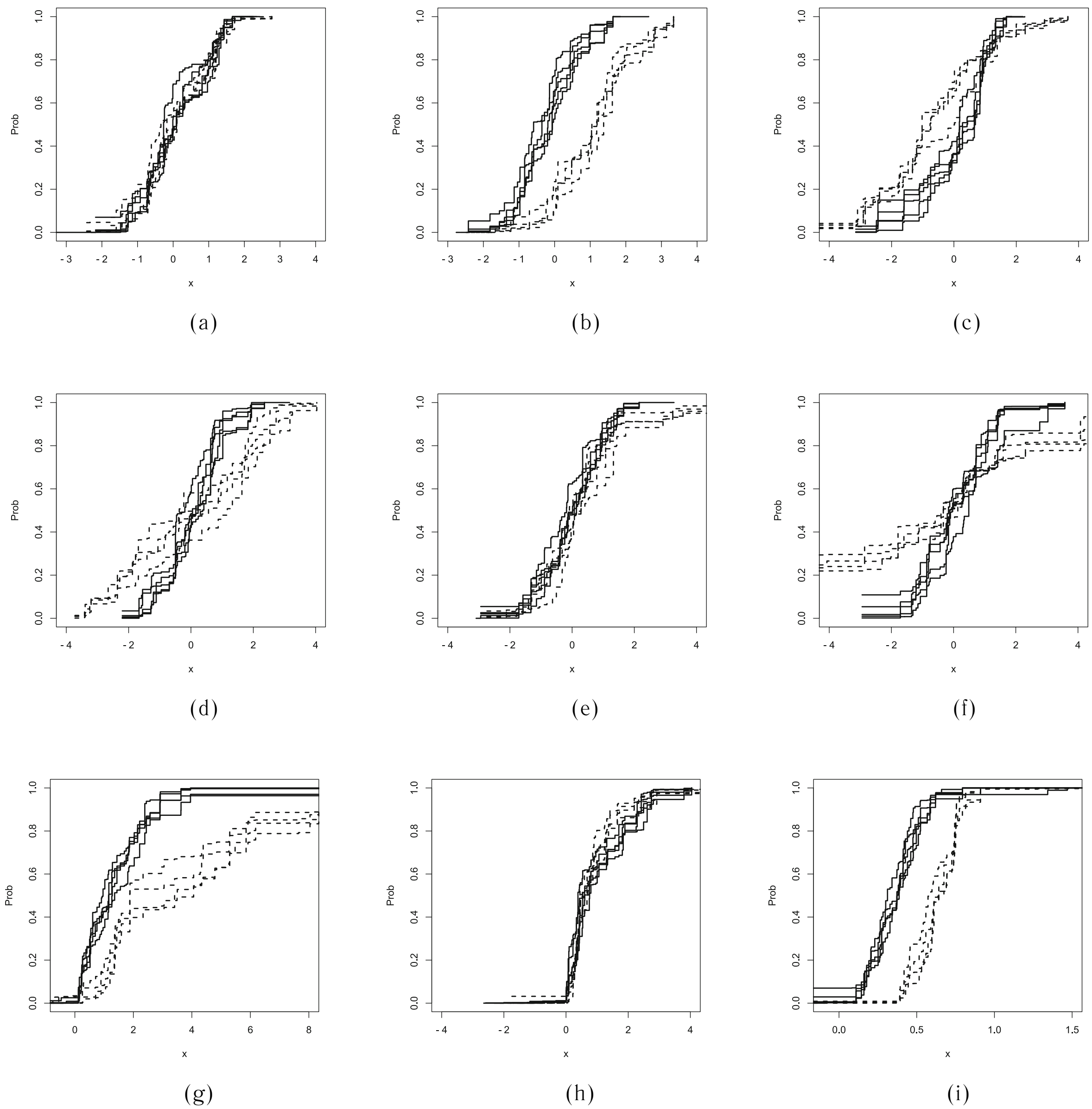

Fig. 2. The solid lines represent sample paths of the posterior Dirichlet process given the first sample and the dashed lines represent sample paths of the posterior Dirichlet process given the second sample. (a) $\mathbf{X} \sim N(0,1)$ and $\mathbf{Y} \sim N(0,1)$; (b) $\mathbf{X} \sim N(0,1)$ and $\mathbf{Y} \sim N(1,1)$; (c) $\mathbf{X} \sim N(0,1)$ and $\mathbf{Y} \sim N(0,2)$; (d) $\mathbf{X} \sim N(0,1)$ and $\mathbf{Y} \sim 0.5N(-2,1) + 0.5N(2,1)$; (e) $\mathbf{X} \sim N(0,1)$ and $\mathbf{Y} \sim t_3$; (f) $\mathbf{X} \sim N(0,1)$ and $\mathbf{Y} \sim t_{0.5}$; (g) $\log \mathbf{X} \sim N(0,1)$ and $\log \mathbf{Y} \sim N(1,1)$; (h) $\mathbf{X} \sim \mathcal{E}(1)$ and $\mathbf{Y} \sim \mathcal{E}(1)$; (i) $\mathbf{X} \sim Beta(4,6)$ and $\mathbf{Y} \sim 0.2 + Beta(4,6)$.

3. Repeat steps (1) and (2) to obtain R_1 i.i.d. samples of $d(P^*_{n_1,m_1}, Q^*_{n_2,m_2})$. For large n_1, n_2 and R_1, the empirical distribution of these values is an approximation to the distribution of $d(P^*_{m_1}, Q^*_{m_2})$.

4. Compute $\overline{d}$, the average of the R_1 values generated at step (3).

5. Compute $d(P^r_{n_1,m_1}, Q^r_{n_2,m_2})$ by repeating steps (1)–(3) with $P^*_{m_1}$, $Q^*_{m_2}$, $P^*_{n_1,m_1}$, $Q^*_{n_2,m_2}$, R_1 and $\overline{d}$ are replaced by $P^r_{m_1}$, $Q^r_{m_2}$, $P^r_{n_1,m_1}$, $Q^r_{n_2,m_2}$, R_2 and $\overline{d}_r$, respectively. Here $P^r_{m_1}$ and $Q^r_{m_2}$ are defined in (7).

6. Compute $\delta = |\overline{d} - \overline{d}_r|$.

7. If $\delta > \epsilon$, then there is a sufficient evidence to reject $\mathcal{H}_0$. Otherwise, we do not reject the null hypothesis $\mathcal{H}_0$.

The results of the next section are based on a simple implementation of this algorithm. The R codes are available from the authors.

5. EXAMPLES

In this section, the proposed method is assessed through two examples, where simulated samples from a variety of distributions are considered. The following notation is used for the distributions in the tables, namely, $N(\mu, \sigma^2)$ is the normal distribution with mean μ and standard deviation σ, t_r is the t distribution with r degrees of freedom, $\mathcal{E}(\lambda)$ is the exponential distribution with mean $1/\lambda$ and $U(a, b)$ is the uniform distribution over $[a, b]$. For all cases we set $n_1 = n_2 = 1000$ in Algorithm A and $R_1 = R_2 = 2000$, and $m_1 = m_2 = 50$ in Algorithm B. The results are also compared with the (frequentist) Kolmogorov–Smirnov (KS) test and the Bayesian nonparametric test in [23]. To calculate the distance in the frequentist KS test, the R function "ks.test" is used. The cut-off of the frequentist KS test can be obtained from standard tables. See, for example, [21]. As for test-B in [23], we use the code provided by the same authors and posted at `http://www.stats.ox.ac.uk/~caron/code/polyatreetest/index.html`.

Example 1. Consider samples generated from the distributions in Table 1, where each sample is of size 50 (Cases 1–9). To study the sensitivity of the approach to the choice of concentration parameters, various values of a_1 and a_2 are considered. The results are reported in Table 1. In particular, when $a_1 = a_2 = 1, 5, 10, 20$, we get $d_r = 0.165, 0.158, 0.151, 0.145$, respectively. Recalling that we want $\delta = |\overline{d} - \overline{d}_r| < \epsilon$ when $\mathcal{H}_0$ is true and $\delta = |\overline{d} - \overline{d}_r| > \epsilon$ when $\mathcal{H}_0$ is false, it is seen that the methodology is powerful in every instance. For example, in Case 1, since $\delta = 0.027 < \epsilon = 0.132$, the two sampling distributions are identical. On the other hand, in Case 2, since $\delta = 0.397 > \epsilon = 0.132$, the two samples are drawn from two different distributions. Notice that, in all cases, the appropriate conclusion is attained with $a_1 = a_2 = 1$. The other values of a_1 and a_2 considered in Table 1 support our conclusions. We point out that the frequentist Kolmogorov–Smirnov fails to recognize the difference between the two samples generated in Case 6 (i.e., $x \sim N(0, 1)$ and $y \sim t_{0.5}$). On the other hand, for the other cases, the three tests gave the same conclusion. In particular, in Case 5, the three methods fail to reject the null hypothesis. This could be explained since the actual Kolmogorov distance between $N(0, 1)$ and t_3 is close to 0.049, which is quite a bit smaller than the other cases in Table 1. For example, the $t_{0.5}$ distribution has Kolmogorov distance from the $N(0, 1)$ around 0.200 while the $N(-1, 1)$ has Kolmogorov distance from the $N(0, 1)$ nearby 0.338. Here, the R code "distrMod" is used to calculate the exact Kolmogorov distance.

Figure 2 provides a plot of 5 sample paths for each of the posterior Dirichlet processes given the first sample and the posterior Dirichlet process given the second sample. In Case 1 (Fig. 2-a), Case 5 (Fig. 2-e) and Case 8 (Fig. 2-h), the plots of the sample paths for the two posterior processes move toward each other. This suggests that the null hypothesis is not rejected. On the other hand, in the other cases, the plots of the sample paths for the posterior processes deviate from each other. This supports the rejection of the null hypothesis.

It is also interesting to consider the effect of prior-data conflict on the methodology. As discussed in Section 4, prior-data conflict will occur whenever there is only a tiny overlap between H_1 and H_2. Table 2 gives the outcomes of particular samples of sizes $m_1 = m_2 = 50$ with various choices of H_1 and H_2. Obviously, when $H_1 = H_2$, we get the correct conclusion but not otherwise. For instance, when $\mathbf{X} \sim N(-5, 1)$ and $\mathbf{Y} \sim N(5, 1)$ with $H_1 = N(-5, 1)$ and $H_2 = N(5, 1)$, we get $\delta = 0$, which yields accepting the null hypothesis. Obviously, this conclusion is incorrect. On the other hand, when $H_1 = H_2$, δ is close to 0.8 and the right conclusion is attained. This illustrates the importance of setting $H_1 = H_2$ in the priors $DP(a_1, H_1)$ and $DP(a_1, H_1)$.

Example 2. In this example, we explore the performance of the proposed test as sample sizes increase. We consider samples from the distributions $\mathbf{X} \sim N(0, 1)$, $\mathbf{Y} \sim N(0, 1)$ (Case 1) and $\mathbf{X} \sim N(0, 1)$, $\mathbf{Y} \sim N(1, 1)$ (Case 2). The results are summarized in Table 3. It follows that the null hypothesis is not rejected in Case 1 but rejected in Case 2. Clearly, the proposed approach works well even with small sample sizes.

Table 1. The proposed Bayesian nonparametric test against the (frequentist) Kolmogorov–Smirnov test and the Holmes et al. test [23]. Here d_{KS} is the (frequentist) Kolmogorov–Smirnov distance.

Samples	$a_1 = a_2$	$\overline{d}(\overline{d}_r)$	$\delta(\epsilon)$	d_{KS}(cut-off)	Holmes et al.
$\mathbf{X} \sim N(0,1)$	1	0.192(0.165)	0.027(0.132)	0.100(0.272)	Fail to reject
$\mathbf{Y} \sim N(0,1)$	5	0.186(0.158)	0.028(0.122)		
	10	0.173(0.151)	0.022(0.116)		
	20	0.166(0.145)	0.0210(0.088)		
$\mathbf{X} \sim N(0,1)$	1	0.562(0.165)	0.397(0.132)	0.540(0.272)	Reject
$\mathbf{Y} \sim N(1,1)$	5	0.524(0.158)	0.366(0.122)		
	10	0.488(0.151)	0.337(0.116)		
	20	0.424(0.145)	0.279(0.088)		
$\mathbf{X} \sim N(0,1)$	1	0.351(0.165)	0.186(0.132)	0.300(0.272)	Reject
$\mathbf{Y} \sim N(0,4)$	5	0.329(0.158)	0.171(0.122)		
	10	0.307(0.151)	0.156(0.116)		
	20	0.272(0.145)	0.127(0.088)		
$\mathbf{X} \sim N(0,1)$	1	0.378(0.165)	0.213(0.132)	0.340(0.272)	Reject
$\mathbf{Y} \sim 0.5(N(-2,1)$	5	0.352(0.158)	0.194(0.122)		
$+N(2,1))$	10	0.327(0.151)	0.176(0.116)		
	20	0.309(0.145)	0.164(0.088)		
$\mathbf{X} \sim N(0,1)$	1	0.200(0.165)	0.035(0.132)	0.120(0.272)	Fail to reject
$\mathbf{Y} \sim t_3$	5	0.194(0.158)	0.036(0.122)		
	10	0.183(0.151)	0.032(0.116)		
	20	0.165(0.145)	0.020(0.088)		
$\mathbf{X} \sim N(0,1)$	1	0.340(0.165)	0.175(0.132)	0.320(0.272)	Reject
$\mathbf{Y} \sim t_{0.5}$	5	0.317(0.158)	0.159(0.122)		
	10	0.295(0.151)	0.144(0.116)		
	20	0.258(0.145)	0.113(0.088)		
$\log \mathbf{X} \sim N(0,1)$	1	0.491(0.165)	0.326(0.132)	0.480(0.272)	Reject
$\log \mathbf{Y} \sim N(1,1)$	5	0.452(0.158)	0.294(0.122)		
	10	0.419(0.151)	0.268(0.116)		
	20	0.357(0.145)	0.212(0.088)		
$\mathbf{X} \sim \mathcal{E}(1)$	1	0.225(0.165)	0.060(0.132)	0.160(0.272)	Fail to reject
$\mathbf{Y} \sim \mathcal{E}(1)$	5	0.208(0.158)	0.050(0.122)		
	10	0.201(0.151)	0.050(0.116)		
	20	0.181(0.145)	0.036(0.088)		
$\mathbf{X} \sim B\text{eta}(4,6)$	1	0.671(0.165)	0.506(0.132)	0.660(0.272)	Reject
$\mathbf{Y} \sim 0.2 + B\text{eta}(4,6)$	5	0.632(0.158)	0.474(0.122)		
	10	0.577(0.151)	0.426(0.116)		
	20	0.492(0.145)	0.347(0.088)		

Table 2. Study of prior-data conflict of the proposed Bayesian nonparametric test for various choices of base measures H_1 and H_2.

Distribution	H_1	H_2	$\overline{d}$	$\overline{d}_r$	$\delta = \lvert\overline{d} - \overline{d}_r\rvert$
$\mathbf{X} \sim N(0,1)$	$N(0,1)$	$N(0,1)$	0.210	0.165	0.045
$\mathbf{Y} \sim N(0,1)$	$N(-5,1)$	$N(5,1)$	0.204	1	0.796
	$U(10,20)$	$N(0,1)$	0.208	1	0.792
	$U(10,20)$	$U(10,20)$	0.206	0.162	0.044
	$\mathcal{E}(1)$	$\mathcal{E}(1)$	0.205	0.166	0.039
$\mathbf{X} \sim N(-5,1)$	$N(0,1)$	$N(0,1)$	0.993	0.165	0.828
$\mathbf{Y} \sim N(5,1)$	$N(-5,1)$	$N(5,1)$	1	1	0
	$U(10,20)$	$N(0,1)$	1	1	0
	$U(10,20)$	$U(10,20)$	0.999	0.162	0.837
	$\mathcal{E}(1)$	$\mathcal{E}(1)$	0.980	0.166	0.814

Table 3. The proposed Bayesian nonparametric test against the (frequentist) Kolmogorov–Smirnov test and the Holmes et al. test [23] when considering small sample sizes. Here d_{KS} is the Kolmogorov–Smirnov distance obtained by the frequentist Kolmogorov–Smirnov test.

Sample Sizes	$\mathbf{X} \sim N(0,1), \mathbf{Y} \sim N(0,1)$			
	$\overline{d}(\overline{d}_r)$	$\delta(\epsilon)$	d_{KS}(cut-off)	Holmes et al.
$m_1 = m_2 = 5$	0.486(0.412)	0.074(0.244)	0.400(0.800)	Fail to reject
$m_1 = m_2 = 10$	0.407(0.326)	0.081(0.196)	0.333(0.600)	Reject
$m_1 = m_2 = 15$	0.337(0.279)	0.058(0.114)	0.385(0.467)	Fail to reject
$m_1 = m_2 = 20$	0.295(0.242)	0.053(0.143)	0.211(0.400)	Reject
$m_1 = m_2 = 30$	0.271(0.203)	0.068(0.112)	0.207(0.333)	Reject
$m_1 = m_2 = 50$	0.271(0.165)	0.027(0.132)	0.180(0.272)	Fail to reject
$m_1 = m_2 = 100$	0.145(0.122)	0.023(0.060)	0.081(0.192)	Fail to reject
$m_1 = m_2 = 200$	0.118(0.090)	0.028(0.038)	0.075(0.136)	Fail to reject

6. ASYMPTOTIC THEORY

Similar to the two-sample frequentist's Kolmogorov–Smirnov test, it is possible to construct a test based on the fact that the two independent processes $\sqrt{m_1}(P^*_{m_1} - H^*_{m_1})$ and $\sqrt{m_2}(P^*_{m_2} - H^*_{m_2})$ converge jointly in distribution to the two independent Brownian bridges B_F and B_G, where F and G are the "true" distributions generating the data. In particular, the next lemma establishes a direct connection between the frequentist two-sample Kolmogorov–Smirnov test and the one that relies on the Dirichlet process. Recall that a Gaussian process is called a *Brownian bridge with parameter measure F*, denoted by $\{B_F(t), t \in \mathbb{R}\}$, if $E(B_F(t)) = 0$ and $\mathrm{Cov}(B_F(s), B_F(t)) = F(\min(s,t)) - F(s)F(t)$, for any $t, s \in \mathbb{R}$ [26].

Lemma 6. *Consider the two-sample problem considered in Section* 3. *Suppose that* $m_1, m_2 \to \infty$, $m_1/(m_1 + m_2) \to \gamma \in (0,1)$. *If the hypothesis* H_0 *holds, then*

$$\sqrt{\frac{m_1 m_2}{m_1 + m_2}} d\big(P^*_{m_1}(t), Q^*_{m_2}(t)\big) \xrightarrow{d} 2 \sup_{x \in \mathbb{R}} |B_F(t)|, \tag{10}$$

where B_F *is a Brownian bridge with parameter measure* F.

Table 4. The proposed Bayesian nonparametric test against the (frequentist) Kolmogorov–Smirnov test and the Holmes et al. test [23] when considering small sample sizes. Here d_{KS} is the Kolmogorov–Smirnov distance obtained by the frequentist Kolmogorov–Smirnov test.

Sample Sizes	$\mathbf{X} \sim N(0,1), \mathbf{Y} \sim N(1,1)$			
	$\overline{d}(\overline{d}_r)$	$\delta(\epsilon)$	d_{KS}(cut-off)	Holmes et al.
$m_1 = m_2 = 5$	0.869(0.412)	0.457(0.244)	1.00(0.800)	Reject
$m_1 = m_2 = 10$	0.773(0.326)	0.447(0.196)	0.800(0.600)	Reject
$m_1 = m_2 = 15$	0.550(0.279)	0.271(0.114)	0.550(0.467)	Reject
$m_1 = m_2 = 20$	0.552(0.242)	0.310(0.143)	0.550(0.400)	Reject
$m_1 = m_2 = 30$	0.475(0.203)	0.272(0.112)	0.467(0.333)	Reject
$m_1 = m_2 = 50$	0.562(0.165)	0.397(0.132)	0.280(0.272)	Reject
$m_1 = m_2 = 100$	0.447(0.122)	0.325(0.06)	0.420(0.192)	Reject
$m_1 = m_2 = 200$	0.409(0.090)	0.319(0.038)	0.39(0.136)	Reject

Proof. Note that

$$\sqrt{\frac{m_1 m_2}{m_1+m_2}}\big(P^*_{m_1}(t) - Q^*_{m_2}(t)\big) = \sqrt{\frac{m_2}{m_1+m_2}}\sqrt{m_1}\big(P^*_{m_1}(t) - H^*_{m_1}(t)\big)$$
$$- \sqrt{\frac{m_1}{m_1+m_2}}\sqrt{m_2}\big(Q^*_{m_2}(t) - H^*_{m_2}(t)\big) + \sqrt{\frac{m_1 m_2}{m_1+m_2}}\big(H^*_{m_1}(t) - H^*_{m_2}(t)\big). \quad (11)$$

The first two terms in (11) converge respectively to $\sqrt{1-\gamma}B_F(t)$ and $-\sqrt{\gamma}B_G(t)$ [25]. On the other hand, the last term in (11) is equal to

$$\sqrt{\frac{m_1 m_2}{m_1+m_2}}\left(\frac{a}{a+m_1}H(t) + \frac{m_1}{a+m_1}F_{m_1}(t) - \frac{a}{a+m_2}H(t) - \frac{m_2}{a+m_2}G_{m_2}(t)\right)$$
$$= \sqrt{\frac{m_1 m_2}{m_1+m_2}}\left(\frac{m_1}{a+m_1}F_{m_1}(t) - \frac{m_2}{a+m_2}G_{m_2}(t)\right) \quad (12)$$
$$+ \sqrt{\frac{m_1 m_2}{m_1+m_2}}\,\frac{a(m_2-m_1)H(t)}{(a+m_1)(a+m_2)}. \quad (13)$$

It follows from a well-known result in empirical processes [34] that, under H_0, (12) converges in distribution to $\sqrt{1-\gamma}B_F(t) - \sqrt{\gamma}B_G(t)$. On the other hand, taking $m_2 = km_1$, for some $k > 0$, simplifies (13) to

$$a(k-1)\sqrt{\frac{k}{1+k}}\frac{m_1^{3/2}}{(a+m_1)(a+km_1)}H(t),$$

which clearly converges to zero as $m_1 \to \infty$. Hence

$$\sqrt{\frac{m_1 m_2}{m_1+m_2}}\big(P^*_{m_1}(t) - Q^*_{m_2}(t)\big) \to 2(\sqrt{1-\gamma}B_F(t) + \sqrt{\gamma}B_G(t)) = 2B_F(t), \quad (14)$$

where the last equality in (14) holds since $B_F(t)$ is the sum of two independent Gaussian processes and

$$E(B_F(t)) = E\left(\sqrt{1-\gamma}B_F(t) + \sqrt{\gamma}B_G(t)\right) = 0,$$

$$\mathrm{Cov}(B_F(s), B_F(t)) = \mathrm{Cov}\left(\sqrt{1-\gamma}B_F(s) + \sqrt{\gamma}B_G(s), \sqrt{1-\gamma}B_F(t) - \sqrt{\gamma}B_G(t)\right)$$
$$= (1-\gamma)\,\mathrm{Cov}(B_F(s), B_F(t)) + \gamma\,\mathrm{Cov}(B_G(s), B_G(t))$$
$$= \mathrm{Cov}(B_F(s), B_F(t)) = F(\min(s,t)) - F(s)F(t).$$

Now the continuous mapping theorem completes the proof of the lemma. □

Note that the result derived in Lemma 6 is, somehow, close to the following well-known result in the theory of empirical processes [34]:

$$\sqrt{\frac{m_1 m_2}{m_1 + m_2}} d\big(F_{m_1}(t), G_{m_2}(t)\big) \xrightarrow{d} \sup_{x\in\mathbb{R}} |B_F(t)|. \tag{15}$$

This is a fairly surprising result as one expects that, asymptotically, the right-hand sides of (10) and (15) should converge to the same limit, which is not the case.

7. CONCLUDING REMARKS

An approach based on the Kolmogorov distance and approximate samples from the Dirichlet process is proposed to assess the equality of two unknown distributions. The current study may lead to further research directions. For instance, it would be interesting to study the effect of selecting other distances such as the Wasserstein (or Kantorovich) distance, the Cramér—von Mises distance and the Anderson—Darling distance on the proposed approach. It is also interesting to make an extensive comparison study between the proposed approach and other Bayesian and frequentist tests. Another important extension is the generalization of the approach to construct a test for multivariate distributions. Finally, constructing a test based on the result derived in Lemma 6 and comparing it with the Kolmogorov—Smirnov test seems motivating. We leave this direction for future work.

8. ACKNOWLEDGMENTS

We would like to thank the editors and an anonymous referee for their careful review of the paper and their extremely helpful comments.

Research of the second author is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

1. L. Al-Labadi and M. Evans, "Optimal Robustness Results for Relative Belief Inferences and the Relationship to Prior-Data Conflict", Bayesian Analysis **12**, 705—728 (2016).
2. L. Al-Labadi, M. Masuadi, and M. Zarepour, "Two-Sample Bayesian Nonparametric Goodness-of-Fit Test" (2015), https://arxiv.org/abs/1411.3427.
3. L. Al-Labadi and M. Zarepour, "A Bayesian Nonparametric Goodness of Fit Test for Right Censored Data Based on Approximate Samples from the Beta—Stacy Process", Canadian J. Statist. **41** (3), 466—487 (2013).
4. L. Al-Labadi and M. Zarepour, "On Asymptotic Properties and Almost Sure Approximation of the Normalized Inverse-Gaussian Process", Bayesian Analysis **8**, 553—568 (2013).
5. L. Al-Labadi and M. Zarepour, "On Simulations from the Two-Parameter Poisson-Dirichlet Process and the Normalized Inverse-Gaussian Process", Sankhya A **76**, 158—176 (2014).
6. L. Al-Labadi and M. Zarepour, "Goodness of Fit Tests Based on the Distance between the Dirichlet Process and Its Base Measure", J. Nonparam. Statist. **26**, 341—357 (2014).
7. J. O. Berger and A. Guglielmi, "Bayesian Testing of a Parametric Model versus Nonparametric Alternatives", J. Amer. Statist. Assoc. **96**, 174—184 (2001).
8. P. Billingsley, *Probability and Measure*, 2nd ed. (Wiley, New York, 1995).
9. D. Blackwell and J. B. MacQueen, "Ferguson Distributions via Polya Urn Schemes", Ann. Statist. **1**, 353—355 (1973).
10. L. Bondesson, "On Simulation from Infinitely Divisible Distributions", Advances in Appl. Probab. **14**, 885—869 (1982).
11. K. M. Borgwardt and Z. Ghahramani, Z. "Bayesian Two—Sample Tests" (2009), http://arxiv.org/abs/0906.4032.
12. C. Carota and G. Parmigiani, "On Bayes Factors for Nonparametric Alternatives", in *Bayesian Statistics 5*, Ed. by J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith (Oxford Univ. Press, London, 1996).
13. Y. Chen and T. Hanson, "Bayesian Nonparametric k-Sample Tests for Censored and Uncensored Data", Comp. Statist. and Data Analysis **71**, 335—346 (2014).
14. A. Dasgupta, *Asymptotic Theory of Statistics and Probability* (Springer, New York, 2008).
15. M. Evans and H. Moshonov, "Checking for prior-data conflict", Bayesian Analysis **1** (4), 893—914 (2006).

16. T. S. Ferguson, "A Bayesian Analysis of Some Nonparametric Problems", Ann. Statist. **1**, 209–230 (1973).
17. J. P. Florens, J. F. Richard, and J. M. Rolin, *Bayesian Encompassing Specification Tests of a Parametric Model against a Nonparametric Alternative*, Techn. Rep. 9608, Univ. Catholique de Louvain, Inst. de Statistique (1996).
18. J. K. Ghosh and R. V. Ramamoorthi, *Bayesian Nonparametrics* (Springer, New York, 2003).
19. A. Gibbs and E. F. Su, "On choosing and bounding probability metrics", Int. Statist. Review **70**, 419–435 (2002).
20. I. J. Good, "The Bayes/Non-Bayes Compromise: A Brief Review", J. Amer. Statist. Assoc. **87**, 597–606 (1992).
21. H. L. Harter and D. B. Owen, *Selected Tables in Mathematical Statistics* (Markham Publ. Co., Chicago, 1972), Vol. I.
22. P. Hsieh, "A Nonparametric Assessment of Model Adequacy Based on Kullback–Leibler Divergence", Statist. and Comput. **23**, 149–162 (2011).
23. C. C. Holmes, F. Caron, J. E. Griffin, and D. A. Stephens, "Two-Sample Bayesian Nonparametric Hypothesis Testing", Bayesian Analysis *2*, 297–320 (2015).
24. L. Huang and M. Ghosh, "Two-Sample Hypothesis Hypothesis Testing Problems under Lehmann Alternatives and Polya Tree Priors", Statist. Sinica **24**, 1717–1733 (2014).
25. L. F. James, "Large Sample Asymptotics for the Two-Parameter Poisson–Dirichlet Process", in *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, Ed. by B. Clarke and S. Ghosal (Inst. of Math. Statist., Ohio, 2008)
26. N. Kim and P. Bickel, "The Limit Distribution of a Test Statistic for Bivariate Normality", Statistica Sinica **13**, 327–349 (2003).
27. M. Lavine, "Some Aspects of Polya Tree Distributions for Statistical Modelling", Ann. Statist. **20**, 1222–1235 (1992).
28. L. Ma and W. H. Wong, "Coupling Optional Pólya Trees and the Two Sample Problem", J. Amer. Statist. Assoc. **106**, 1553–1565 (2011).
29. R. McVinish, J. Rousseau, and K. Mengersen, "Bayesian Goodness of Fit Testing with Mixtures of Triangular Distributions", Scand. J. Statist. **36**, 337–354 (2009).
30. P. Muliere and K. Tardella, "Approximating Distributions of Random Functionals of Ferguson–Dirichlet Prior", Canadian J. Statist. **26**, 283–297 (1998).
31. J. Sethuraman, "A Constructive Definition of Dirichlet Priors", Statistica Sinica **4**, 639–650 (1994).
32. K. Shang and C. Reilly, "Nonparametric Bayesian Analysis Censoring", Commun. Statist. – Theory and Methods (2017) DOI: 10.1080/03610926.2017.1288249.
33. T. B. Swartz, "Nonparametric Goodness-of-Fit" Commun. Statist. – Theory and Methods **28**, 2821–2841 (1999).
34. A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics* (Springer, New York, 1996).
35. K. Viele, *Evaluating fit using Dirichlet processes*, Techn. Rep. 384 (Univ. Kentucky, Dept. Statist., 2000).
36. R. L. Wolpert and K. Ickstadt, "Simulation of Lévy Random Fields", in *Practical Nonparametric and Semiparametric Bayesian Statistics*, Ed. by D. Day, P. Nuller, and D. Sinha (Springer, 1998).
37. M. Zarepour and L. Al-Labadi, "On a Rapid Simulation of the Dirichlet Process", Statist. and Probab. Lett. **82**, 916–924 (2012).