Journal of Statistical Modelling: Theory and Applications
Vol. 5, No. 1, 2024, pp. 15-32
Yazd University Press 2024

JSMTA

Research Paper

Goodness of fit tests for nonincreasing densities on real positive data with nonparametric Bayesian methods

Soleiman Khazaei*
[1]Department of Statistics, Razi University, Kermanshah, Iran

Received: April 09, 2024/ Revised: July 11, 2024/ Accepted: July 23, 2024

Abstract: In this paper, we study a nonparametric Bayesian inference on the family of nonincreasing density functions on real positive data. One interesting problem is the goodness of fit test in such a context. In other words, we consider nonparametric Bayesian testing on the family of nonincreasing density in this domain. So, we define nonparametric hypothesis testing and compare two different testing approaches. The first approach is given based on the Bayes factor. This approach is the well-known Bayesian approach for testing, although its computation is complicated. Decision-theoretic considerations with the loss function drive the second approach for a given distance. This second approach has the advantage of considering the distance to the null hypothesis but needs the definition of a threshold. When no threshold is known as a priori, a possibility exists to calculate a p-value, and the method becomes more complicated to compute. We propose a hybrid algorithm to accelerate the computation of the p-value. The comparison of both approaches is performed based on a simulation study.

Keywords: Bayes factor; Loss function; Nonincreasing density; P-value; Testing hypotheses.
Mathematics Subject Classification (2010): 62G10, 62F15.

1 Introduction

In many applied areas of research, the nonparametric Bayesian methods are used increasingly, but their applications in hypothesis testing situations have become of interest recently. In particular, the well-known problem of the goodness of fit testing has

*Corresponding author: s.khazaei@razi.ac.ir received negligible attention compared to the estimation problem, see among others Verdinelli and Wasserman (1998), Robert and Rousseau (2003), and also Al labadi and Zarepour (2013) study the Bayesian nonparametric goodness of fit test for right censored data, Hart and Choi (2017) and Al labadi and Berry (2022) by using the Dirichlet process proposed the estimation of the extropy and consider the goodness of fit test procedure. First, we consider a testing procedure based on the Bayes factor (BF), while the latter, we use an approach based on loss functions. Improvement in Markov Chain Monte Carlo (MCMC) simulation methods has increased the use of Bayesian techniques in more complex and realistic models than was previously possible. Despite the significant development of these algorithms, the computation of Bayesian tests remains an issue, especially in large dimensional frameworks, as encountered in the goodness of fit tests. In particular, BF requires the computation of marginal likelihood, which is difficult, see for instance, Basu and Bery (2003). In this work, we are interested in goodness of fit tests where the density is known to be monotone nonincreasing on $\mathbb{R}^+$. Testing uniformity versus a monotone density has been studied by Woodroofe and Sun (1999) from a frequentist perspective. More precisely, they consider the problem of testing $H_0 : f = 1$ versus $H_1 : f \neq 1$ where f is monotone and nonincreasing function on $[0, 1]$.

The presented method, which involves nonparametric Bayesian inference and goodness -of-fit testing for nonincreasing density functions on real positive data, finds application in various fields where understanding the distribution of continuous, positive-valued variables is crucial. Here, there are some practical examples to illustrate its application:

Analysis of Environmental Data: Suppose we are interested in modeling the concentration of a pollutant (measured in parts per million, ppm) in a river over time. We hypothesize that the concentration levels should follow a nonincreasing trend as the pollutants disperse downstream. We want to test whether the observed pollutant concentrations at these stations follow a nonincreasing density function over time.

Analysis of Drug Dissolution Rates in Pharmaceutical Research: In pharmaceutical research and development, understanding the dissolution rates of drugs is critical for ensuring their effectiveness and safety. The dissolution rate refers to how quickly and completely a drug substance dissolves in the gastrointestinal tract, influencing its bioavailability and therapeutic efficacy. Investigate whether the observed dissolution rates follow a nonincreasing pattern over time. We hypothesize that as time progresses, the dissolution rates may decline due to factors such as saturation of the dissolution medium or changes in drug particle size.

Modeling Cell Growth Rates in Biotechnology: In biotechnology, understanding and modeling the growth rates of cells is crucial for optimizing processes such as fermentation or cell culture. Cell growth rates are typically positive values that decrease over time due to factors like nutrient depletion or waste accumulation. We can test the hypothesis if the observed cell growth rates follow a nonincreasing pattern over time.

We study Bayesian nonparametric testing on the family of monotone nonincreasing density functions on $\mathbb{R}^+$. Nonparametric estimation of the monotone nonincreasing density is a well-known problem and has been considered from theoretical and applied perspectives in the frequentist literature. In particular, the estimation of monotone density functions has applications in reliability and serves as a preliminary analysis in survival analysis. Monotone nonincreasing densities on $\mathbb{R}^+$ have a mixture representation, allowing for likelihood-based inference. See, for instance, the introduction in Balabdaou and Wellner (2007) for a review on the subject. Let $\mathcal{F}_0$ be the set of all monotone nonincreasing densities on $\mathbb{R}^+$ and Π is the prior probability measure on $\mathcal{F}_0$. Similar to Woodroofe and Sun (1999) given observations $\boldsymbol{x} = (x_1, ..., x_n)$ supposed to be Independent and identically distributed (i.i.d.) from some monotone nonincreasing density, f. Our goal is to consider the following test

$$H_0 : f = f_0 \quad \text{against} \quad H_1 : f \neq f_0 \quad \text{and} \quad f \searrow, \tag{1}$$

where f_0 is a given monotone nonincreasing density. Note that we do not restrict ourselves to $f_0 = 1$, although this is a trivial generalization. We compare two different approaches to test the above hypotheses. One is based on the BF, which is related to the penalized likelihood ratio test and can be written as

$$BF_{0/1} = \frac{f_0(\boldsymbol{x})}{m(\boldsymbol{x})}, \quad m(\boldsymbol{x}) = \int_{\mathcal{F}_0} f(\boldsymbol{x}) d\Pi(f), \quad f(\mathbf{x}) = \prod_{i=1}^{n} f(x_i). \tag{2}$$

The prior probability measure is the most commonly used in the Bayesian approach for testing, although its computation remains an open problem, given its complicated issue. Sections 2 and 3 verify this approach to test the defined hypothesis. Decision theoretic considerations drive the second approach. Consider for a given metric,d, the loss function

$$L(f, \delta) = \delta(\epsilon - d(f, f_0))\mathbb{I}_{d(f,f_0)<\epsilon} + (1 - \delta)(d(f, f_0) - \epsilon)\mathbb{I}_{d(f,f_0)>\epsilon}, \tag{3}$$

where in this paper, d is the L_1 distance. From Robert and Rousseau (2003)

$$\delta^{\Pi} = 1 \quad \text{iff} \quad E^{\Pi}\left(d(f, f_0)|\boldsymbol{x}\right) > \epsilon,$$

where $E^{\Pi}\left(d(f, f_0)|\boldsymbol{x}\right)$ denotes the posterior expectation of $d(f, f_0)$.

The second approach has the advantage of considering the distance to the null hypothesis but needs the definition of a threshold ϵ. In the case where there is no prior knowledge on the tolerance threshold ϵ, Robert and Rousseau (2003) and Rousseau (2007) calibrate it by computing a p-value associated with the test statistic $H(\boldsymbol{x}) = \mathbb{E}^{\Pi}[d(f, f_0)|\boldsymbol{x}]$ and Rousseau (2007) studies what it means in term of threshold. In other words, it is proved in Rousseau (2007) that using the p-value defined by $p_0(\mathbf{x}) = P_0^n\left[H(\mathbf{X}) \geq H(\mathbf{x})|\mathbf{x}\right]$ and accepting H_0 if $p_0(\mathbf{x}) \leq \alpha$ with α fixed corresponds to choosing a threshold of order the posterior concentration rate for estimating f_0 under the alternative. For a discussion on the implications of such a result, see Rousseau (2007).

The p-value approach is computationally more demanding. We propose a hybrid algorithm following McVinish et al. (2009) to accelerate the computation of the p-value. This method is described in Section 4, and then the comparison of both approaches is based on a simulation study.

2 A nonparametric prior on the set of monotone nonincreasing densities on $\mathbb{R}^+$

To define a Bayesian procedure either based on the $BF_{0/1}$ in (2) or on some other loss functions as in (3), one needs to define a prior on $\mathcal{F}_0$. To do so, we use the well-known representation of monotone nonincreasing densities on $\mathbb{R}^+$. By Williamson (1956) and Lévy (1962) it is known that a density function f is monotone nonincreasing if and only if there exists P a probability measure on $\mathbb{R}^+$ such that $f = f_P$ with

$$f_P(\boldsymbol{x}) = \int_0^\infty \mathbb{I}_{(\boldsymbol{x}\leq\theta)} \frac{1}{\theta} dP(\theta). \tag{4}$$

Remember that we denoted $\mathcal{F}_0 = \{\text{monotone nonincreasing densities on } \quad \mathbb{R}^+\} = \{f_P : P \in \mathcal{M}\}$ where $\mathcal{M}$ is the set of probability measures on $\mathbb{R}^+$ and $\theta \in \mathbb{R}^+$.

Hence, a monotone nonincreasing density is a mixture of uniform distributions on $\mathbb{R}^+$, and a natural nonparametric family of priors is the Dirichlet Process Mixture (DPM) of uniform distributions corresponds to considering a Dirichlet Process (DP) on the mixing P on $\mathbb{R}^+$. The Dirichlet process Mixture of uniform distributions has the following hierarchical representation

$$\begin{aligned} X_i|\theta_i &\overset{ind.}{\sim} G(.|\theta_i) = \mathcal{U}(0,\theta_i) \quad \text{for} \quad i = 1,...,n, \\ \theta_1,...,\theta_n|P &\overset{i.i.d}{\sim} P, \\ P &\sim DP(.|\alpha, G_0), \end{aligned} \tag{5}$$

where $0 < \alpha < \infty$, G_0 is a probability distribution on $\mathbb{R}^+$ and $DP(\alpha, G_0)$ denotes the distribution of Dirichlet process with base measure αG_0. Note that the density of $G(.|\theta_i)$ is given by $g(x|\theta_i) = \frac{1}{\theta_i}\mathbb{I}_{0\leq x\leq\theta_i}$. Using the stick-breaking representation of the DP by Sethuraman (1994), If $V_j \overset{i.i.d}{\sim} Beta(1,\alpha), j = 1,2,...$, we can write

$$P(.) = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}(.), \qquad \theta_j \overset{i.i.d}{\sim} G_0, \tag{6}$$

where $\pi_1 = V_1$, $\pi_j = V_j \prod_{i=1}^{j-1}(1 - V_i)$ for $j = 2, 3, ...$ and the $\delta_{\theta_j}(.)$ stands for the Dirac massing on θ_j. Then f_P in (4) can be defined as

$$f_P(x) = \sum_{j=1}^{\infty} \pi_j \frac{1}{\theta_j} \mathbb{I}_{x\leq\theta_j}(x),$$

emphasizing the discrete nature of the DPM. Alternatively using Blackwell and MacQueen (1973) we can express the marginal distribution of $\theta_1, ..., \theta_n$ in the hierarchical representation (5), integrating out P. This leads to $\theta_1 \sim G_0$ and for all $i = 2, ..., n$ the conditional distribution of θ_i given $\theta_1, ..., \theta_{i-1}$ is

$$P(\theta_i \in .|\theta_1, ..., \theta_{i-1}) = \frac{\alpha}{n+i-1} G_0(.) + \sum_{j=1}^{k_{[i-1]}} \frac{n_{j,i-1}}{n+i-1} \delta_{\theta^*_{j,i-1}}(.), \tag{7}$$

where $\{\theta^*_{1,i-1}, ...\theta^*_{k_{[i-1]},i-1}\}$ are the set of $k_{[i-1]}$ distinct values in $\{\theta_1, ..., \theta_{i-1}\}$ and $n_{j,i-1}$ is the number of points in $\{\theta_1, ..., \theta_{i-1}\}$ equal to $\theta^*_{j,i-1}$.

In this section, we are interested in hypothesis testing (1). A first possibility is to compute a BF using a prior H_1 like DPM of uniform distributions described above. The BFs are the Bayesian answers to 0-1 types of loss functions and are, therefore, better suited for well-separated hypothesis testing problems. The BF for testing (1) based on observation $\boldsymbol{x} = (x_1, ..., x_n)$ is the ratio of the marginal likelihood under the null hypothesis to the marginal likelihood under the alternative hypothesis as defined by (2) with marginal likelihood, $m(\boldsymbol{x}) = \int_{\mathcal{M}} \left(\prod_{i=1}^n \int_{x_i}^{\infty} \frac{1}{\theta} dP(\theta)\right) d\Pi(P)$. Hence, it is necessary to compute the marginal likelihood to calculate the BF. Note that in the case of the BF approach, it is also easy to extend the hypothesis test to

$$H_0 : f_0 \in \mathcal{G}_0 \quad \text{versus} \quad H_1 : f_0 \notin \mathcal{G}_0, \quad f_0 \searrow, \tag{8}$$

where $\mathcal{G}_0$ is a given parametric model composed of monotone nonincreasing densities such as $\mathcal{G}_0 = \{f_\lambda(x); x > 0, \lambda \in \Lambda, f_\lambda \in \mathcal{F}\}$ and Λ is infinite dimensional space. Indeed, in this case, the marginal likelihood on $\mathcal{G}_0$ is relatively easy to compute, particularly when the prior belongs to a conjugate family of the model $(f_\lambda, \lambda \in \Lambda)$. Then we can write the BF as $BF_{0/1} = \int_{\mathbb{R}^+} f_\lambda(\boldsymbol{x}) d\Pi_0(\lambda)/m(\boldsymbol{x})$ where Π_0 is prior on $\mathbb{R}^+$. For the observations $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ large values of the $BF_{0/1}$ shows that there is strong evidence for H_0 based on the data, small values of $BF_{0/1}$ show otherwise. As n, the sample size increases indefinitely, we would expect to obtain appropriate information about the sampling density, say f_0, and the BF should also correctly be able to decide between H_0 and H_1. Dass and Lee (2006) showed that under weak conditions on the prior of the model and if the null model $\mathcal{G}_0$ consists of a single density f_0, the BF is consistent. Consistency of the BF when $\mathcal{G}_0$ is a parametric family is not so clear, especially under H_0, see Rousseau (2007) for a discussion on these issues. In the special case where $\lambda = 1$, and $f_0(\boldsymbol{x})$ is the exponential distribution with parameter one, the $BF_{0/1}$ is given by

$$BF_{0/1}(\boldsymbol{x}) = \frac{e^{-n\bar{\boldsymbol{x}}}}{m(\boldsymbol{x})} \qquad \boldsymbol{x} > 0. \tag{9}$$

Taking the logarithm of both sides of the expression (9), we have that

$$\log BF_{0/1} = -n\bar{\boldsymbol{x}} - \log m(\boldsymbol{x}) \quad \text{and} \quad \widehat{\log BF_{0/1}} = -n\bar{\boldsymbol{x}} - \widehat{\log m(\boldsymbol{x})},$$

where $\widehat{\log m(\boldsymbol{x})}$ is an estimation of $\log m(\boldsymbol{x})$. The difficulty here lies in the computation of $m(\boldsymbol{x})$.

2.1 Computation of the marginal likelihood $m(\boldsymbol{x})$

In a statistical model, estimation of the marginal likelihood can be a provocative task. In general, the marginal likelihood does not have a closed form solution except in some special cases where there exists a conjugate prior and also in practice there are some difficulties that hampers the estimation of the marginal likelihood. Even in reasonably simple statistical models computing the marginal likelihood can be difficult, see Calderhead and Girolami (2009). Mixture models are a typically case of this situation.

Especially, in high dimensional models it becomes a crucial issue. Since $\mathcal{F}_0$ is infinite dimensional, $m(\boldsymbol{x})$ as defined in (2) cannot be obtained analytically. Recall that under the $DP(\alpha, G_0)$ prior on P which we denote $\Pi(.|G_0,\alpha)$ the marginal distribution of $\boldsymbol{x}$ has the form

$$m(\boldsymbol{x}) = m_{G_{0,\alpha}}(\boldsymbol{x}) = \int f_P(\boldsymbol{x}) d\Pi(P|G_{0,\alpha}), \tag{10}$$

where $f_P(\boldsymbol{x}) = \prod_{i=1}^n \int_{x_i}^\infty \frac{1}{\theta} dP(\theta)$. Clearly, analytical or exact calculation $m_{G_0,\alpha}(\boldsymbol{x})$ is not possible when n becomes large.

In the Ferguson (1973) they did some exact calculation for small sample sizes, which were extended by Brunner and Lo (1989), but in large sample sizes this is not feasible. In this work we use the method of Basu and Chib, see Basu and Bery (2003). The method of Basu and Chib is based on Importance Sampling, hereafter denoted (IS) approach. We first explain how the marginal likelihood function can be obtained as a by product of the Sequential Importance sampling (SIS) method. Recall that $\boldsymbol{x} = (x_1, ..., x_n)$ and that $\theta_{(n)} = (\theta_1, ..., \theta_n)$ is a decomposition of θ in the hierarchical formulation of the DPM model (5) and set G_0 the Inverse Gamma distribution $IG(a,b), a > 0, b > 0$ and denote its density by g_0. By the sequential method of Kong et al. (1994), we obtain an unbiased estimate of the marginal likelihood function, $m_{G_0,\alpha}(\boldsymbol{x})$, in the following way : We first generate θ_1 from the conditional predictive density

$$\pi(\theta_1|x_1) = \frac{f(x_1|\theta_1)g_0(\theta_1)}{m(x_1)} = \frac{\frac{b^a}{\Gamma(a)}(\frac{1}{\theta_1})^{a+2} e^{\frac{-b}{\theta_1}}}{m(x_1)} \mathbb{I}_{(x_1 \le \theta_1)}, \tag{11}$$

where $m(x_1)$ is the marginal likelihood function at x_1. $\boldsymbol{x}_{(i)} = (x_1, ..., x_i)$, $\theta_{(i)} = (\theta_1, ..., \theta_i)$. From (7), we compute $f(x_i|\boldsymbol{x}_{(i-1)}, \theta_{(i-1)})$

$$f(x_i|\boldsymbol{x}_{(i-1)}, \theta_{(i-1)}) = \frac{\alpha}{\alpha_i} \int_{x_i}^\infty \frac{b^a}{\Gamma(a)} \frac{e^{-\frac{b}{\theta}}}{\theta^{a+2}} d\theta + \sum_{j=1}^{K_{[i-1]}} \frac{n_{j,i-1}}{\alpha_i} (\frac{\mathbb{I}_{(x_i < \theta_j)}}{\theta_j}), \tag{12}$$

and then we compute the conditional probabilities by

$$\pi(\theta_i|\boldsymbol{x}_{(i)}, \theta_{(i-1)}) = c_i \left\{ \frac{\alpha}{\alpha_i} (\frac{\mathbb{I}_{(x_i < \theta_i)}}{\theta_i}) g_0(\theta_i) + \sum_{j=1}^{K_{[i-1]}} \frac{n_{j,i-1}}{\alpha_i} \delta_{\theta^*_{j,i-1}}(\theta_i) \right\}, \tag{13}$$

where g_0 is Inverse-gamma density function with parameter a and b, where $\alpha_i = \alpha + i - 1$, $n_{j,i-1}$ and $k_{[i-1]}$ are defined in (7) and c_i is the normalizing constant and we simulate θ_i from (13). Set $w_1(\theta_0) = w_1$ and for $i = 2, ..., n$,

$$w_i(\theta_{(i-1)}) = w_{i-1}(\theta_{(i-2)}) f(x_i|\boldsymbol{x}_{(i-1)}, \theta_{(i-1)}).$$

Then $w_n(\theta_{(n)}) = m(x_1) \prod_{i=2}^n f(x_i|\boldsymbol{x}_{(i-1)}, \theta_{(i-1)})$. In fact this kind of computation is commonly referred to as peeling. We repeatedly compute the predictive conditional probabilities (12) and the conditional probabilities (13) independently M times. Choice of suitable M, the number of replications, is discussed at the end of this section. For $m = 1, ..., M$, we consider the results of the above computations as $\theta^m = (\theta_1^m, ..., \theta_n^m)$

and $w_n^{(m)} = w_n(\theta^m)$. Here we note that θ^m is a sample coming from a proposal distribution $\pi^*(\theta^m|\boldsymbol{x})$, which is not equal to the actual conditional distribution $\pi(\theta^m|\boldsymbol{x})$, see Kong et al. (1994). Hence, $w_n^{(m)}$ can be rewritten as $\frac{\pi(\theta^m|\boldsymbol{x})}{\pi^*(\theta^m|\boldsymbol{x})} = \frac{w_n^{(m)}}{m_{G_{0,\alpha}}(\boldsymbol{x})}$ and then

$$w_n^{(m)} = \frac{\pi(\theta^m|\boldsymbol{x})}{\pi^*(\theta^m|\boldsymbol{x})} m_{G_{0,\alpha}}(\boldsymbol{x}), \tag{14}$$

where $m_{G_{0,\alpha}}(\boldsymbol{x})$ is given by (10) and $\pi^*(\theta^m|\boldsymbol{x}) = \pi(\theta_1^m|x_1)\prod_{i=2}^n \pi(\theta_i^m|\boldsymbol{x}_{(i)}, \theta_{(i-1)}^m)$.

Since $m_{G_{0,\alpha}}(\boldsymbol{x})$ in (14) is independent of θ^m, the expression (14) can be used to obtain an estimate of the marginal likelihood function $m_{G_{0,\alpha}}(\boldsymbol{x})$. By a simple calculation we obtain $\mathbb{E}^{\pi^*}[w^{(m)}(\theta^m)] = \int w^{(m)}(\theta^m)\pi^*(\theta^m|\boldsymbol{x})d\theta^m = m_{G_{0,\alpha}}(\boldsymbol{x})$. So, an unbiased estimate of marginal likelihood function for $m_{G_{0,\alpha}}(\boldsymbol{x})$ is given by

$$\hat{m}_{G_0,\alpha}(\boldsymbol{x}) = \frac{1}{M}\sum_{m=1}^{M} w^{(m)}(\theta^m).$$

Hence, the mean of the $w^{(m)}$ is a consistent Monte carlo estimate of the marginal likelihood function. We can apply this basic idea to the DPM model, but in applying SIS method to the DPM model.

The weights $w^{(m)}$ are highly variable, see Basu and Bery (2003) . To overcome this problem they consider the collapsed Sequential Importance Sampling(SIS) approach developed in the context of a DPM model and later the method extended to multinomial and non-exchangeable beta-binomial models by Quintana and Newton (1998). The idea of the collapsed SIS method is to integrate out the θ_i's for $i = 1, ..., n$ which collapses the space in which the sequential sampling operates to the set of possible cluster memberships. Since the θ_i's integrate out analytically given the clustering structure, this approach has less variability due to the Rao-Blackwellization effect, see MacEachern et al. (1999). Now, we estimate the marginal likelihood function $m_{G_{0,\alpha}}(x)$ using the collapsed SIS as described in Basu and Bery (2003). Denote by k_i the membership index of θ_i, i.e. $k_i = j$ if and only if $\theta_i = \theta_j$ $j, i = 1, ..., n$ and $k_1 = 1$ and $k_{(n)} = (k_1, ..., k_n)$. The idea is to simulate $k_{(n)}$ in the place of $\theta = (\theta_1, ..., \theta_n)$. Let $s_1 = m(x_1) = \int_{x_1}^{\infty} \frac{b^a}{\Gamma(a)}(\frac{1}{\theta})^{(a+2)}e^{-\frac{b}{\theta}}d\theta$ and for $i = 2, ..., n$ we compute sequentially the prequential predictive density of x_i in following step 1 of the collapsed SIS approach. We set $\boldsymbol{x}_j^{i-1} = \{x_l; l \leqslant i-1, k_l = j\}$, $N_j^{i-1} = \max\{x_l; l \leqslant i-1, k_l = j\}$ and n_j^{i-1} the cardinal of the set $\boldsymbol{x}_j^{i-1}$, then $K_j^{i-1}(\theta|\boldsymbol{x}_j^{i-1})$, the posterior distribution of θ, based on the prior and on $\boldsymbol{x}_j^{i-1}$ those latent observations in the jth cluster is

$$K_j^{i-1}(\theta|\boldsymbol{x}_j^{i-1}) = \frac{\prod_{l\in\boldsymbol{x}_j^{i-1}} \frac{1}{\theta}\mathbb{I}_{(x_l\leq\theta)}g_0(\theta)}{\int \prod_{l\in\boldsymbol{x}_j^{i-1}} \frac{1}{\theta}\mathbb{I}_{(x_l\leq\theta)}g_0(\theta)d\theta} = \frac{\mathbb{I}_{(\theta\geqslant N_j^{i-1})}(\frac{1}{\theta})^{n^*+1}e^{-\frac{b}{\theta}}}{\int_{N_j^{i-1}}(\frac{1}{\theta})^{n_j^{i-1}+a+1}e^{-\frac{b}{\theta}}d\theta}, \tag{15}$$

where $n^* = n_j^{i-1} + a$. Using (13) and (15) we then consider:
1. Compute for each $i = 2, ..., n$,

$$s_i \quad = \quad f(x_i|\boldsymbol{x}_{(i-1)}, k_{(i-1)}, G_0)$$

$$= \frac{\alpha}{\alpha_i}\int_0^\infty f(x_i|\theta)dG_0(\theta) + \sum_{j=1}^{k_{[i-1]}} \frac{n_j^{i-1}}{\alpha_i}\int_0^\infty f(x_i|\theta)K_j^{i-1}(\theta|\boldsymbol{x}_j^{i-1})d\theta.$$

Now we move to the next step of the collapsed SIS method, where we apply the method to drew a k_i from the following joint distribution:
2. Let $k_{\max}^{i-1} = \max\{k_l; l \leqslant i-1\}$ drew k_i from

$$\begin{aligned} p(k_i = j|\boldsymbol{x}_{(i)}, k_{(i-1)}) &= \begin{cases} \frac{n_{j,i}}{\alpha_i}\int_{x_i}\frac{1}{\theta}dK_j^{i-1}(\theta|\boldsymbol{x}_j^{i-1}), & 1 \leqslant j \leqslant k_{\max}^{i-1} \\ \frac{\alpha}{\alpha_i}\int_{x_i}\frac{1}{\theta}dG_0(\theta), & j = k_{\max}^{i-1}+1 \end{cases} \\ &= \begin{cases} \frac{(n^*)n_j^{i-1}}{b(\alpha_i)}\frac{P_{G_0(n^*+1,b)}(\theta \geqslant N_j^{i-1} \vee x_i)}{P_{G_0(n^*,b)}(\theta \geqslant N_j^{i-1})}, & 1 \leqslant j \leqslant k_{\max}^{i-1} \\ \frac{\alpha a P_{G_0(a+1,b)}(\theta \geqslant x_i)}{(\alpha_i)b}, & j = k_{\max}^{i-1}+1. \end{cases} \end{aligned} \tag{16}$$

In other words either the k_i's come from the set of current individual cluster labels with probabilities given in the first line of the (16) or is equal to $k_{\max}^{i-1}+1$ (a new cluster label) with probability given in the second line of (16). Start with $k_{(n)}^{(0)}$ an initial value and iterate for $m = 1, ..., M$. By step 1 and step 2 we obtain $(k_{(n)}^{(1)}, ..., k_{(n)}^{(M)})$ and $s_i^{(m)} = f(x_i|\boldsymbol{x}_{(i-1)}, k_{(i-1)}^{(m)}, G_0)$, $i = 1, ..., n$. Then we compute $w^{(m)}(k_{(n)}^{(m)}) = s_1^{(m)}\prod_{i=2}^n s_i^{(m)}$ and finally $\hat{m}_{G_0,\alpha}(\boldsymbol{x}) = \bar{w} = \frac{1}{M}\sum_{m=1}^M w^{(m)}(k_{(n)}^{(m)})$. Figure 1 illustrates the stabilization of the estimate of the marginal likelihood estimation of the DPM model based on the collapsed SIS method. As the graph shows the estimate stabilizes up to the second decimal place quite quickly. For $M = 100000$ and $n = 10, 100, 500, 1000$ we perform the algorithm and we compute estimation of the logarithm of the BF to perform the goodness of Fit test (1). In the simulation study we generate the repeated samples $\boldsymbol{x}^t = (x_1^t, ..., x_n^t), t = 1, ..., T$ from $f_\theta(x) = \theta e^{-\theta x}$ so that $\theta = 1$ corresponds to H_0 and for all $\theta \neq 1, f_\theta$ is in H_1. Then to estimate $\mathbb{E}_\theta[\log BF_{0/1}(\boldsymbol{x})]$, we have iterated the algorithm for $T = 50$ times to compute $\mathbb{E}_\theta[\log BF_{0/1}(\boldsymbol{x})] = \frac{1}{T}\sum_{t=1}^T \log\widehat{BF}_{0/1}(\boldsymbol{x}^t)$. These estimates are listed in Table 1. As it is shown when the observations are generated from the exponential distribution with parameter $\theta = 1$, the logarithm of the BF takes maximum value and it decreases much more slowly in n for values of θ that are close to 1 and for $\theta = 1.2$ the $\log BF_{0/1}$ is still increasing between $n = 100$ and $n = 1000$ whereas for $\theta = 0.2, 2$ it has strongly decreased. When $n = 10$ the estimation of $\log BF_{0/1}$ under H_0 for all parameters is negative although for $\theta = 1$ it takes maximum value. In this case we can compute exactly the logarithm of the BF and compare it with the estimation from collapsed sequential method. The calculation shows that the Collapsed sequential method accurately estimates the $\log BF_{0/1}$ with maximum digression 0.002 which it is defined by $|\frac{\log BF_{0/1}}{\log BF_e} - 1|$, where $\log BF_e$ is the exact value of logarithm of the BF. Therefore when the sample size is small we would reject the null hypothesis whatever the value of θ, even when $\theta = 1$ (at least on average). This is a surprising result since one would expect that the smaller n the harder it is to detect departure from the null.

histograms of the logarithm of BF's over repeated samples under f_θ, when $n = 100, 500, 1000$, these are shown in Figures 2-4 respectively. Although the means have

Table 1: Estimate of logarithm of the BF to test the model (1) when the observations are generated from exponential distribution with different parameter values θ.

	θ					
n	0.2	0.8	0.9	1	1.2	2
10	-25.77	-1.74	-1.08	-0.42	-1.58	-2.13
100	-153.56	1.19	1.94	3.14	2.43	-10.34
500	-181.04	0.74	10.79	14.85	6.40	-28.34
1000	-253.56	0.26	6.40	128.14	8.88	-114.34
5000	$-\infty$	-7.34	-1.40	$+\infty$	-6.18	-201.21

increased slowly the histograms for $\theta = 1.2$ and 0.8 have shifted slightly towards negative values, in the sense that the left hand tails of the histograms are more spread out when $n = 500$ and $n = 1000$ than when $n = 100$. But as we see in Table 1 when $n = 5000$ the $\log BF_{0/1}$ makes a clear distinction between the models even if the values of θ are close to one. According to Jeffreys' scale of evidence when $\theta = 0.8, 0.9, 1$ or 1.2 we accept the null hypothesis for $10 < n \leq 1000$, since that there is not enough reason to reject the H_0.

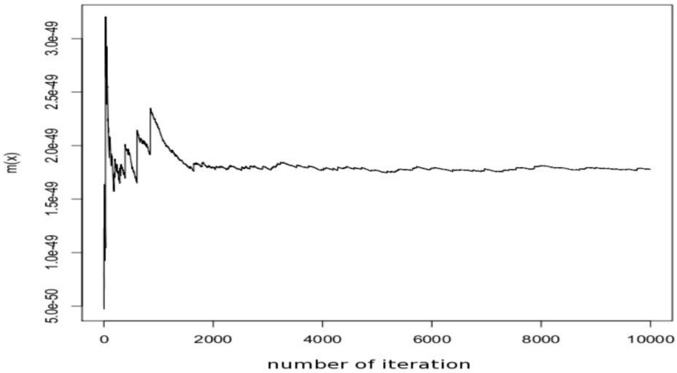

Figure 1: The logarithm of marginal likelihood estimation of the DPM model versus number of iteration when $n = 1000$ and $f_0(x) = e^{-x}, x > 0$.

We now present the approach based on the loss function in (3).

3 The loss function approach

In the goodness of fit setting measures of goodness of fit typically summarize the discrepancy between observed values and the expected values under the model in question, when there does not exist visible separation between the models and BF might not be the most appropriate answer. In this section we consider an alternative to the BF for the testing problem (1) using a loss function which also takes into account the distance between f and f_0 so that accepting H_0 if f is close to f_0 is not as serious as accepting when it is very different. In this section we investigated testing procedure which is the Bayesian answer to a distance based on loss function (3) which is a prior more satisfying in cases where the usual question is: Is f_0 a reasonable approximation of f? Consider the L_1 distance on densities, $d(f_P, f_0) = \int |f_P - f_0|dx$. Often the problem of interest is to check if a given f_0 is a reasonable model for the observations, i.e.

$$H_0 : d(f_P, f_0) \leq \epsilon, \qquad H_1 : d(f_P, f_0) \geq \epsilon, \tag{17}$$

so that ϵ determines what reasonable means. In the goodness of fit framework the second definition of hypothesis will be more specifically studied, however we encounter with the question, how we can choose ϵ small enough? Choosing ϵ is difficult. Berger and Delampady (1987) obtained a bound in term of ϵ and Rousseau (2007) obtained a range of value of ϵ corresponding to the use of defined p-value in three different situations.

To answer (17), we consider loss function in (3). The Bayesian answer to the loss function in (3) is given by

$$\delta^{\Pi}(x) = 1 \quad \text{iff} \quad H(\boldsymbol{x}) \geqslant \epsilon$$

where $H(\boldsymbol{x}) = \mathbb{E}^{\Pi}[d(f, f_0)|\boldsymbol{x}]$ and $d(f, f_0) = |f - f_0|_1$, i.e. we accept H_0 if and only if $H(\mathbf{x}) \leq \epsilon$ and reject H_0 if and only if $H(\boldsymbol{x}) \geq \epsilon$. To decide which hypothesis will be accepted we need to define the threshold ϵ, when such threshold is unknown a priori, one possibility is to consider a p-value associated to the test statistic $H(\boldsymbol{x})$ to calibrate the threshold problem as proposed by Robert and Rousseau (2003). Hence, we compute

$$P(\boldsymbol{x}) = P_0(H(X) \geq H(\boldsymbol{x})|\boldsymbol{x}), \tag{18}$$

where P_0 is the model under f_0 and $\boldsymbol{x}$ is the observed values. This allows for the calibration of the test statistics $H(\boldsymbol{x})$ into a known scale. We now describe how to compute such p-values in the context of monotone nonincreasing decreasing densities.

3.1 Computation of the Bayesian p-values

To compute the p-value, $P(\boldsymbol{x})$, described above we need first to compute the test statistic $H(\boldsymbol{x})$ for a given data set $\boldsymbol{x} = (x_1, ..., x_n)$, under the nonparametric prior Π on $\mathcal{F}_0$ the set of non increasing densities or $\mathcal{M}$ the set of probability measures on $\mathbb{R}^+$. Recall that the prior is defined by

$$f_P(\boldsymbol{x}) = \prod_{i=1}^{n} \int \frac{1}{\theta} \mathbb{I}_{x_i \leq \theta} dP(\theta), \quad P \sim DP(\alpha, IG(a, b)). \tag{19}$$

Since $H(\boldsymbol{x}) = \int_{\mathcal{M}} d(f_P, f_0) d\Pi(P|\boldsymbol{x})$, we cannot use samples of the posterior of P which are only based on the marginal representation of the Dirichlet Process (Blackwell and Mackqueen). We therefore consider simulations of the posterior based on the retrospective sampling algorithm by Papaspiliopoulos and Roberts (2008) and the slice sampler of Kalli et al. (2011). The output of the retrospective sampling are in the form $(K^t, V^t, Z^t)_{t=1}^T$ where $K^t = (K_1^t, ..., K_n^t), V^t = (V_1^t, ..., V_{k_{\max}^t}^t)$ and $Z^t = (Z_1^t, ..., Z_{k_{\max}^t}^t)$ where T is the number of iteration and $k_{\max}^t = \max(K^t)$. We represent (19) for $\boldsymbol{x} = (x_1, ..., x_n)$ as

$$x_i|Z, K \overset{ind.}{\sim} g(x_i|Z_{K_i}) = U(0, Z_{K_i}) \quad \text{for} \quad i = 1, ..., n,$$

$$\begin{aligned}
K_i|\pi &\overset{i.i.d.}{\sim} \sum_{j=1}^{\infty}\pi_j\delta_j(\cdot),\\
Z_j &\overset{i.i.d.}{\sim} G_0 = IG(a,b)\ a,b>0 \quad \text{for} \quad j=1,2,...,\\
\pi_1 &= V_1, \quad \pi_j = V_j\prod_{i=1}^{j-1}(1-V_i), \quad V_j \overset{i.i.d.}{\sim} Beta(1,\alpha), \quad \text{for } j=2,..., \qquad (20)\\
P(\cdot) &= \sum_{j=1}^{\infty}\pi_j\delta_{Z_j}(\cdot).,
\end{aligned}$$

where we can derive simple expressions for the full conditional distributions of $K|\boldsymbol{x}, Z, V$ and $Z, V|K, \boldsymbol{x}$ and thus use a Gibbs algorithm. The parameters in (20) are the classification variable $K=(K_1,...,K_n)$ (that is K_i is an allocation variable related to X_i, where $K_i=j$ if and only if $\theta_i = Z_j$) the cluster parameters Z_j, the cluster probabilities $\pi_j, j=1,2,...$ and finally the random measure P.

To generate (K^t, V^t, Z^t) we use the algorithm proposed by MacEachern et al. (1999). Now, to be able to compute $d(f_P, f_0)$ at each iteration we need to compute $(f_P^t(y_j))_{j=1}^J$ where $(y)_{j=1}^J$ is a grids on $\mathbb{R}^+$ at each iteration t. To compute $f_P^t(y)$ we use Papaspiliopoulos and Roberts (2008) which

$$f_P^t(y_j) \overset{d}{=} \sum_{j=1}^{\max(K^t)} \frac{\pi_j^t}{Z_j^t}\mathbb{I}_{(y_j \leqslant Z_j^t)} + \tilde{f}_P^t(y_j)\prod_{j=1}^{\max(K^t)}(1-V_j^t), \qquad (21)$$

where $\pi_j^t = V_j^t\prod_{i=1}^{j-1}(1-V_i^t), Z_j^t \sim IG(a,b), a,b>0$ and $\tilde{f}_P^t$ is draw from the prior. Thus, when the output $(K^t, V^t, Z^t)_{t=1}^T$ come from the retrospective sampling algorithm and $(\tilde{f}_P^t)_{t=1}^T$ are drown independently from the prior, this representation of f_P^t are asymptotically distributed from the posterior. We thus need to simulate $\tilde{f}_P$ from the prior , which is done using Guglielmi and Tweedie (2001). Also for computing $H(\boldsymbol{x})$ one of the difficulties in the above computation comes from the fact that the Z_j^t's, $t=1,...,T$ are truncated $IG(a,b)$ random variables. To generate Z_j^t's from truncated $IG(a,b)$ so we use Damien and Walker (2001) describe. Obviously, for a given data set $\boldsymbol{x}$, obtaining $H(\boldsymbol{x})$ is computationally demanding. It is thus not feasible to compute $H(\boldsymbol{x})$ for many different data sets $\boldsymbol{x}$ as would be necessary to compute the p-value $P(\boldsymbol{x})$. To do so, we use an IS approximation in a similar manner to McVinish et al. (2009). Consider for $m=1,...,M, \boldsymbol{x}^{new,m} = (x_1^{new,m},...,x_n^{new,m})$ samples independently distributed from the model defined by H_0 in (17), that is, $\boldsymbol{x}^{new,m} \overset{ind.}{\sim} \exp(1)$. Since the support of the posterior distribution given new data is the same as the support of the posterior distribution given old data, we can use two data sets to approximate $H(\boldsymbol{x}^{new,m})$ by a MCMC run under the posterior $d\Pi(P|\boldsymbol{x}^{new,*})$. Choose the data set $\boldsymbol{x}^{new,*}$ such that $\min(x_1^{new,*},...,x_n^{new,*}) \leq \min(x_1^{new,m},...,x_n^{new,m})$ for $m=1,...,M$. Hence

$$H(\boldsymbol{x}^{new,m}) = \frac{\int d(f_P,f_0)W(\boldsymbol{x}^{new,m},\boldsymbol{x}^{new,*})d\Pi(P|\boldsymbol{x}^{new,*})}{\int W(\boldsymbol{x}^{new,m},\boldsymbol{x}^{new,*})d\Pi(P|\boldsymbol{x}^{new,*})},$$

where

$$W(\boldsymbol{x}^{new,m},\boldsymbol{x}^{new,*}) = \frac{f_P(\boldsymbol{x}^{new,m})}{f_0(\boldsymbol{x}^{new,*})} = \frac{\prod_{i=1}^n f_P(x_i^{new,m})}{\prod_{i=1}^n f_P(x_i^{new,*})}, \qquad (22)$$

$$f_P(x_i^{new,m}) \quad = \quad \int_{x_i^{new,m}} \frac{1}{\theta} dP(\theta), \quad \text{for } m = 1, ..., M \quad \text{and} \quad f_P(\boldsymbol{x}^{new,*}) \neq 0.$$

According to McVinish et al. (2009), an estimation of the expected value is given by

$$\hat{\mathbb{E}}^{\Pi}[d(f_P, f_0)|\boldsymbol{x}^{new,m}] = \frac{\sum_{t=1}^{T} d(f_P^t, f_0) W_t(\boldsymbol{x}^{new,m}, \boldsymbol{x}^{new,*})}{\sum_{t=1}^{T} W_t(\boldsymbol{x}^{new,m}, \boldsymbol{x}^{new,*})}, \tag{23}$$

where T is the number of MCMC iteration. As a result of the strong law of large numbers this IS estimator with normalized weights is consistent for $\mathbb{E}^{\Pi}[d(f_P, f_0)|\boldsymbol{x}^{new,m}]$ see Tierney (1994).

Let $\boldsymbol{x}^0 = (x_1^0, ..., x_n^0)$ be a vector of observations comes from model (17) under null hypothesis, that is $\boldsymbol{x}^0 \overset{i.i.d}{\sim} \exp(1)$ and also $\boldsymbol{x}^{new} = (x_1^{new}, ..., x_n^{new})$ be a future observation where $\boldsymbol{x}^{new} \overset{i.i.d}{\sim} \exp(1)$. Since the parameter under H_0 is known, the p-value is defined by

$$P(\boldsymbol{x}^0) = P_{H_0}[H(\boldsymbol{x}^{new}) \geq H(\boldsymbol{x}^0)|\boldsymbol{x}^{\mathbf{0}}].$$

Thus a proper approximation to the p-value can be obtained as

$$\hat{P}(\boldsymbol{x}^0) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{I}(\hat{H}(\boldsymbol{x}^{new,m}) \geq \hat{H}(\boldsymbol{x}^0)). \tag{24}$$

Let we choose G grids on $[0, 1]$ and define $G_{[0,1]} = \{y_1, ..., y_G\}$ and $G_{\mathbb{R}^+} = \{x_1 = F_0^{-1}(y_1), ..., x_G = F_0^{-1}(y_G)\}$ be a grids on $\mathbb{R}^+$. To estimate the $H(\boldsymbol{x}^1)$ or $H(\boldsymbol{x}^{new,m})$ it is necessary estimate

$$\hat{d}(f_P^t, f_0) = \frac{1}{G} \sum_{g=1}^{G} |\frac{\hat{f}_P^t(x_g)}{f_0(x_g)} - 1|. \tag{25}$$

To do this we use retrospective MCMC sampling following to generate the samples Z^t, V^t, P^t for number of iteration $T = 500000$. We use the thinning-Burnin to make a new T=10000. we propose the following Pseudo-code representation to compute the estimation of $H(\boldsymbol{x})$. $\forall j \in \mathbb{N}$ we generate Z, V and update the K. By algorithm 2 from Papaspillopoulos and Roberts(2008).

Algorithm 3.1. *Retrospective MCMC to estimate* $H(\boldsymbol{x}^{new,m})$.

```
Step 1 Initialisation of $Z^0, V^0$ and $K^0$
Step 2
    Step 2.1 For $t \in \{0, 1, ..., T-1\}$ by using Papaspiliopoulos and Roberts (2008):
    Step 2.2 Generate $(Z^t, V^t)$ given $X^n$ and $K^t$.
    Step 2.3 Generate $K^t$ given $(X^t, Z^t, V^t)$.
    Step 2.4 Generate $\tilde{f}_P(x_g)$ from the prior.
    Step 2.5 Generate $(f^t(x_g))_P^G$.
Step 3
    Step 3.1Compute the estimated loss $\hat{d}(f_P^t, f_0)$ using (25). Using the thinning + Burnin based on a very long MCMC chain to make a much small chain independently new $T$.
```

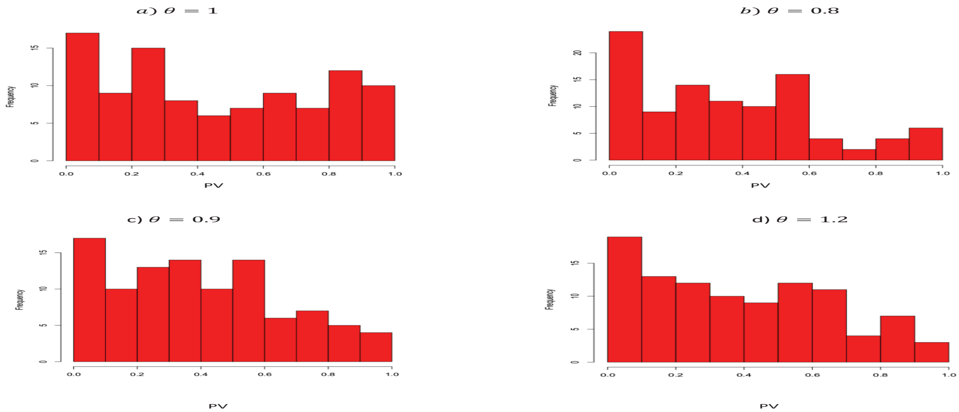

Figure 2: Histograms of p-values when $n = 100$ observations are generated from an exponential distribution with parameter θ.

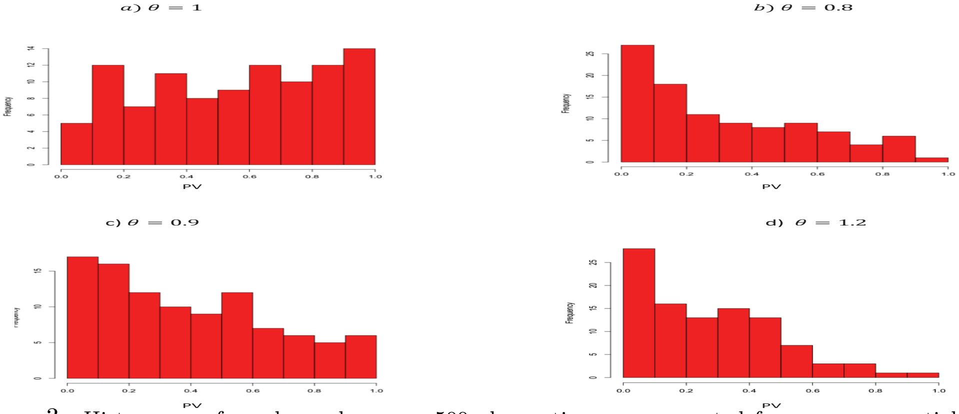

Figure 3: Histograms of p-values when $n = 500$ observations are generated from an exponential distribution with parameter θ.

Step 3.2 *$\boldsymbol{x}^m$ from the null hypotheses for $m = 1, ..., M$.*
Step 3.3 *$W(\boldsymbol{x}^{new,m}, \boldsymbol{x}^{new,*})$ using* (22)*.*
Step 3.3 *$H(\boldsymbol{x}^{new,m})$ using* (23)*.*

To do so we generate repeatedly samples $\boldsymbol{x}^m = (x_1^m, ..., x_n^m)$ for $m = 1, ..., M$ where $M = 1000$ and with different sample sizes $n = 100, 500, 1000$ from the density $f_0(x) = \theta e^{-\theta x}\mathbb{I}_{x>0}$ so that when $\theta = 1$ it corresponds to null hypothesis and we use empirical estimation of p-value is given by (24). The result of estimation of the p-value for different sample size under H_0 are listed in Table 2 as we see according to the values of the p-value we don't have enough reasons to reject H_0 when $n = 10, 100, 500, 1000$. The distribution of the p-value is uniform on $[0, 1]$ when H_0 is true and this uniformity defines a proper p-value which allows for its common interpretation across problems. Hence as we see in the Figures (a) for 5-7 the histograms seem to be uniform. We therefore investigate better the behavior of the p-value around $\theta = 1$ that is when $\theta = 0.8, 0.9$ and 1.2. In fact in these cases we compute the P-value under H_0 and we

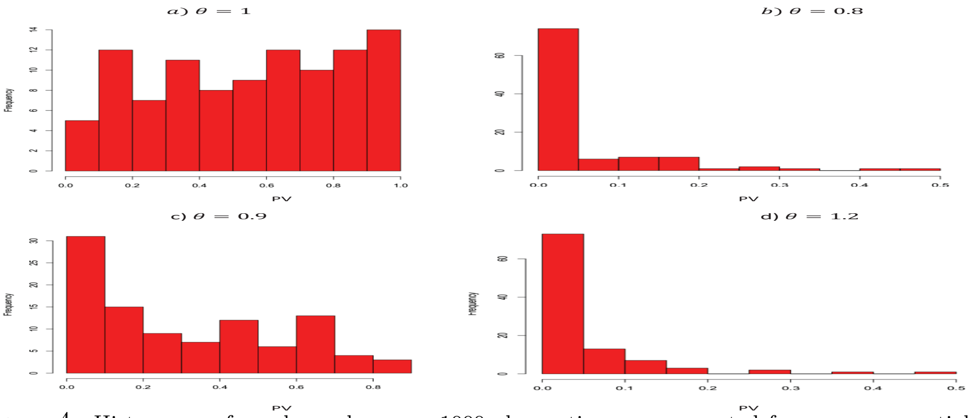

Figure 4: Histograms of p-values when $n = 1000$ observations are generated from an exponential distribution with parameter θ.

see that the distribution of the p-value is not uniform on $[0, 1]$ and this nonuniformity is more clearer when the sample size increases see Figures (b), (c) and (d) of the Figures 5-7.

Table 2: Estimation of the expected p-value for different values of the parameter θ.

	θ		
n	0.8	1	1.2
10	0.5923	0.4441	0.3534
100	0.3702	0.4541	0.3885
500	0.1412	0.5432	0.1399
1000	0.0536	0.5780	0.0487

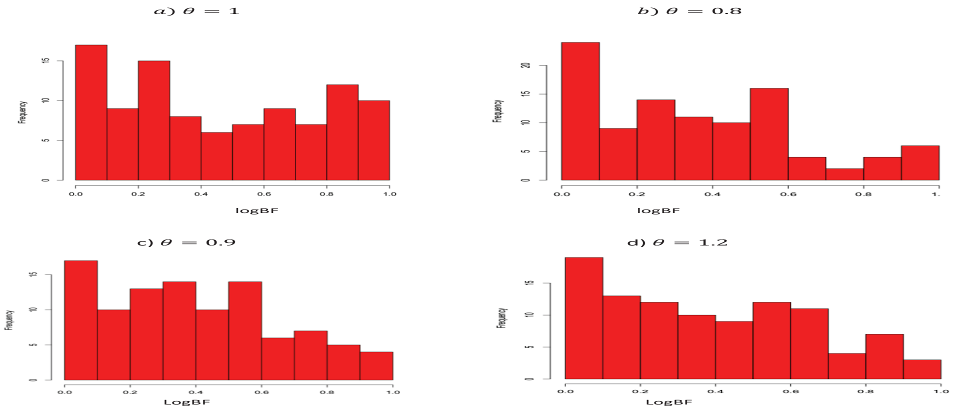

Figure 5: Histogram of the logarithm of the BF when $n = 100$ observations are generated from an exponential distribution with parameter θ.

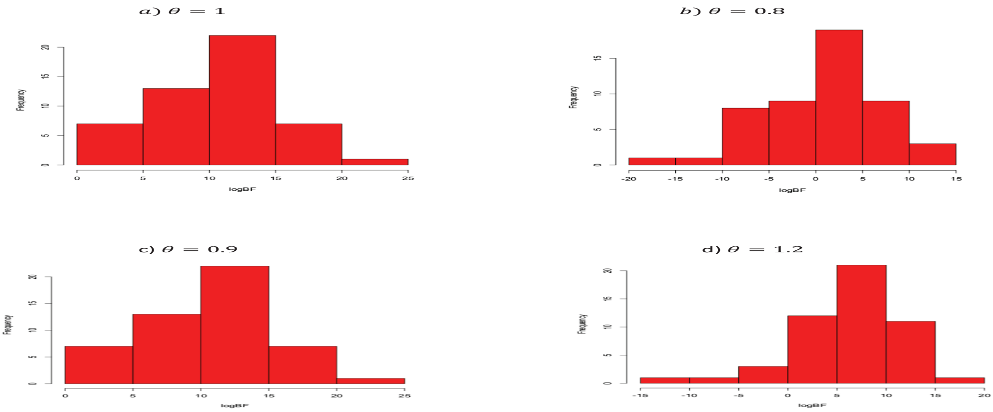

Figure 6: Histogram of the logarithm of the BF when $n = 500$ observations are generated from an exponential distribution with parameter θ.

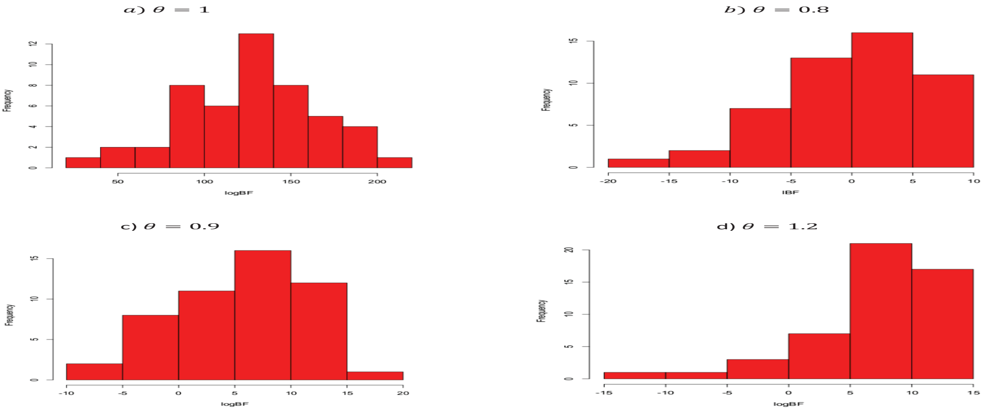

Figure 7: Histogram of the logarithm of the BF when $n = 100$ observations are generated from an exponential distribution with parameter θ.

4 Discussion and conclusions

In this paper, we have studied the goodness of fit test where the density is known to be monotone nonincreasing on $\mathbb{R}^+$. To do that we compare two different approaches. The two approaches we have studied are the BFs and loss function approach, calibrated by a p-value. The former is the most common Bayesian testing procedure. The range of this factor is considered a degree of credibility of the hypotheses. The BFs are consistent for hypothesis testing and model selection. This property of the BF is basic; that is, if one of the models under consideration is true, then statistical methods should guarantee the selection of the true model if enough information is observed. The use of the BF guarantees consistency, while the use of classical selection tools like p-value does not guarantee consistency under H_0. The BF will choose the model that is closest to the true model in terms of the Kullback-Leibler divergence. The above results on the BF for goodness of fit test of the model versus nonparametric family show that the logarithm of the BF takes maximum value when the observations are generated from the null hypotheses. The marginal likelihood estimation decreases much more slowly in n for values of θ that are close to one than for values of θ that are further away. When the sample size increases the logarithm of the BF tends to increase while when the sample size is small for instance for $n = 10$ in average the BF's cannot distinguish the difference between H_0 and H_1 even for θ far from $\theta = 1(H_0)$. When n becomes larger, $\log BF$ on average still cannot differentiate H_0 and H_1 while θ is close to 1 $(0.8, 1.2)$ since on average its logarithm is still positive. Hence in this case n has to be greater than 5000 for $\log BF$ to see the difference. As Table 2 shows, the results of the second method for goodness-of-fit testing between the model and the alternative model indicate that estimations of the frequentist expectation of the posterior risk decrease at a rate that seems to be slightly slower than $n^{1/3}$. Looking at the average of the p-value is not enough to see if it gives a discriminating method between H_0 and H_1 since for $n \leq 500$ the means of p-values under $\theta = 0.8$ or 1.2 are greater than 0.1. However for $n = 1000$ the means under $\theta = 0.8$ and $\theta = 1.2$ are smaller than 0.05. To refine the comparisons we have looked at the histograms of both, p-value and $\log BF$. These histograms show quite significantly that the p-values are more discriminating for smaller values of n. Hence, we propose the following scale of evidence: i. If $n < 5000$ the p-value is better than the BF approach, ii. If $n \geq 5000$ the BF is better than the p-value approach.

References

Al Labadi, L. and Berry, S. (2022). Bayesian estimation of extropy and goodness of fit tests. *Journal of Applied Statistics*, **49**(2):357–370.

Al labadi, L. and Zarepour, M. (2013). A Bayesian nonparametric goodness of fit test for right censored data based on approximate samples from the beta-Stacy process. *Canadian Journal of Statistics*, **41**(3):466–487.

Basu, S. and Chib, S. (2003). Marginal likelihood and bayes factors for dirichiet process mixture models. *Journal of the American Statistical Association*, **98**(461):224–235.

Balabdaoui, F. and Wellner, J.A. (2007). Estimation of a k-monotone density: limit distribution theory and the spline connection, *The Annals of Statistics*, **35**(6):2536–2564.

Berger, J.O. and Delampady, M. (1987). Testing precise hyphoteses. *Statistical Science*, **2**(3):317–352.

Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via polya urn schemes. *The annals of Statistics*, **1**(2):353–355.

Brunner, L.J. and Lo, A.Y. (1989). Bayes methods for a symmetric unimodal density and its mode, *The Annals of Statistics*, **17**(4):1550–1566.

Calderhead, B. and Girolami, M. (2009). Estimating Bayes factors via thermo-dynamic integration and population MCMC, *Computational Statistics and Data Analysis*, **48**(12):4028–4045.

Damien, P and Stephen G. Walker (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, **10**(2):206–215.

Dass, S.C. and Lee, J. (2006). A note on the consistency of Bayes factors for testing point null versus non-parametric alternatives. *Journal of Statistical Planning and Inference,* **119**(1):143–152.

Ferguson, T. (1973). Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**(2):209–230.

Guglielmi A. and Tweedie R.L. (2001). Markov chain Monte Carlo estimation of the law of the mean of a Dirichlet process. *Bernoulli*, **7**(4):573–592.

Hart, J.D. and Choi, T. (2017). Nonparametric goodness of fit via cross-validation Bayes factors. *Bayesian Analysis*, **12**(3):653–677.

Kalli, M., Griffin, J.E., and Walker, S.G. (2011). Slice sampling mixture models. *Statistics and Computing*, **21**(1):93–105.

Kong, A., Liu, J.S., and Wong, W.H. (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, **89**(425):278–288.

Lévy, P. (1962). Extensions d'un théorème de D. Dugué et M. Girault. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **1**:159–173.

Lo, A.Y., Brunner, L.J. and Chan, A. T. (1996). *Weighted Chinese restaurant processes and Bayesian mixture models.* Research Report, Hong Kong: The University of Science and Technology, ISMT Department.

MacEachern, S.N., Clyde, M., and Liu, J.S. (1999). Sequential importance sampling for non-parametric Bayes models : the next generation. *Canadian Journal of Statistics*, **27**(2):251–267.

McVinish, R., Rousseau, J. and Mengersen, K. (2009). Bayesian goodness-of-fit testing with mixtures of triangular distributions. *Scandinavian Journal of Statistics*, **36**(2):337–354.

Papaspiliopoulos, O. and Roberts, G.O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, **95**(1):169–186.

Quintana, F.A. and Newton, M.A. (1998). Assessing the order of dependence for partially exchangeable binary data. *Journal of the American Statistical Association*, **93**(441):194–202.

Robert, C.P. and Rousseau, J. (2003). A mixture approach to Bayesian goodness of fit. *Les cahiers du CEREMADE (2005-31).* http://www.ceremade.dauphine.fr/preprints /CMD/2002-9.ps.gz

Rousseau, J. (2007). Approximating interval hypotheses: P-values and Bayes factors. *Bayesian Statistics*, **8**:417–452.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**:639–650.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, **22**(4):1701–1786.

Verdinelli, I. and Wasserman, L. (1998). Bayesian goodness-of-fit testing using infinite-dimensional exponential families. *The Annals of Statististics*, **26**(4):1215–1241.

Williamson, R.E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Mathematical Journal*, **23**(2):189–207.

Woodroofe, M. and Sun, J. (1999). Testing uniformity versus a monotone density. *The Annals of Statististics*, **27**(1):338–360