physical sciences forum

MDPI

Proceeding Paper

Nonparametric Full Bayesian Significance Testing for Bayesian Histograms †

Fernando Corrêa *,‡, Julio Michael Stern ‡ and Rafael Bassi Stern ‡

Institute of Mathematics and Statistics, University of São Paulo, São Paulo 05508-040, Brazil; jstern@ime.usp.br (J.M.S.); rstern@ime.usp.br (R.B.S.)

* Correspondence: fptcorrea@gmail.com; Tel.: +55-1196065-9957

† Presented at the 43rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Ghent, Belgium, 1–5 July 2024.

‡ These authors contributed equally to this work.

Abstract

In this article, we present an extension of the Full Bayesian Significance Test (FBST) for nonparametric settings, termed NP-FBST, which is constructed using the limit of finite dimension histograms. The test statistics for NP-FBST are based on a plug-in estimate of the cross-entropy between the null hypothesis and a histogram. This method shares similarities with Kullback–Leibler and entropy-based goodness-of-fit tests, but it can be applied to a broader range of hypotheses and is generally less computationally intensive. We demonstrate that when the number of histogram bins increases slowly with the sample size, the NP-FBST is consistent for Lipschitz continuous data-generating densities. Additionally, we propose an algorithm to optimize the NP-FBST. Through simulations, we compare the performance of the NP-FBST to traditional methods for testing uniformity. Our results indicate that the NP-FBST is competitive in terms of power, even surpassing the most powerful likelihood-ratio-based procedures for very small sample sizes.

Keywords: nonparametrics; bayesian nonparametrics; significance testing

Academic Editor: Geert Verdoolaege

Published: 20 October 2025

Citation: Corrêa, F.; Stern, J.M.; Stern, R.B. Nonparametric Full Bayesian Significance Testing for Bayesian Histograms. *Phys. Sci. Forum* **2025**, *12*, 11. https://doi.org/10.3390/psf2025012011

1. Introduction and General Setting

Full Bayesian Significance Testing (FBST) [1] is a Bayesian method for testing if a parameter θ belongs to some set Θ_0. In traditional statistical setting, researchers analyze a collection of n observations $\mathbf{X}_n = (X_1, \ldots, X_n)$ that are presumed to conform to a specified distribution f_θ characterized by an unobserved parameter θ. A Bayesian statistician makes inferences about θ by updating a prior density $\pi(\theta)$, supported by the set of all possibilities Θ. After observing $\mathbf{X}_n$, one obtains a posterior density $f_{\theta|\mathbf{X}_n}$. Often, one needs to determine whether $f_{\theta|\mathbf{X}_n}$ supports a scientific hypothesis framed with respect to θ belonging to some subset $\Theta_0 \subset \Theta$, written $H_0 : \theta \in \Theta_0$. FBST tests H_0 by comparing the posterior density $f_{\theta|\mathbf{X}_n}$ of points inside and outside Θ_0. This comparison is represented by the posterior probability of the tangential set:

$$T(\Theta_0) = \{\theta : f_{\theta|X}(\theta) \leq \sup_{t \in \Theta_0} f_{\theta|X}(t)\}. \tag{1}$$

$T(\Theta_0)$ encompasses all points in the parameter space that exhibit lower posterior density compared to those in Θ_0. The FBST methodology posits that if the posterior probability of $T(\Theta_0)$ is low, the hypothesis H_0 should be rejected, as it is located in a region characterized by low posterior density.

Definition 1. *In a standard Bayesian statistical model, let Θ be a finite dimensional parametric space, $\mathbf{X}_n$ an observed sample, $\mathcal{L}$ the likelihood function, π be the prior distribution, and $f_{\theta|\mathbf{X}_n}$ the posterior density proportional to $\pi(\theta)\mathcal{L}(\theta)$. Also, let $\Pi_{\theta|\mathbf{X}_n}$ be the measure on Θ induced by $f_{\theta|\mathbf{X}_n}$. The Full Bayesian Significance Test (FBST) for testing $H_0 : \theta \in \Theta_0$ consists on rejecting the null hypothesis based on the e-value statistic*

$$ev(\Theta_0; \mathbf{X}_n) = \Pi_{\theta|\mathbf{X}_n}(T(\Theta_0)) \tag{2}$$

where the tangential set $T(\Theta_0)$ is given by Equation (1). H_0 is rejected if $ev(\Theta_0; \mathbf{X}_n) < \alpha$ for some fixed significance level $\alpha \in [0, 1]$.

In other words, the e-value quantifies the credibility of a hypothesis H_0 using the maximum probability argument, whereby a system is optimally represented by its most probable realization. This probability is defined as the posterior density $f_{\theta|\mathbf{X}_n}$, which quantifies the continuous probability associated with a specific point $\theta \in \Theta_0$. The e-value directly addresses the question "What is the posterior probability of observing a θ with a posterior density exceeding that of any point in Θ_0?". A higher e-value signifies that $H_0 : \theta \in \Theta_0$ is deemed more credible, whereas a lower e-value suggests that H_0 is considered less credible.

In this paper, we extend this concept to a nonparametric framework for density estimation using histograms. Bayesian nonparametric approaches for density estimation can be divided into two main categories. The first type focuses on defining priors Π_0 in the infinite-dimensional space of probability densities. Upon observing the data $\mathbf{X}_n$, these priors are updated into infinite-dimensional posteriors, facilitating an adapted approach to Bayesian inference. Well-established examples of such priors include the Dirichlet Process Mixtures (DPM) and its extensions [2]. In contrast, the second type of Bayesian nonparametric approach employs regular finite-dimensional Bayesian modeling in parameter spaces $\Theta_{k(n)}$ that maintain a fixed dimension $k(n)$, while allowing $k(n)$ for gradual expansion as the sample sizes increase. This includes truncated versions of infinite-dimensional priors and histograms with a fixed number of bins that increases with the sample size. This paper specifically examines a variant of the FBST applied in the context of increasing dimensionality.

In this paper, we propose an FBST for the problem of Bayesian density estimation using Dirichlet-Multinomial models, interpreted as histograms where the number of bins increases with the sample size. This methodology is in alignment with the Bayesian frameworks outlined by [3–5]. Therefore, we will use consistent notation to leverage results from the existing literature. The primary advantages of leveraging the Dirichlet-Multinomial model include (1) the feasibility of deriving an explicit formula for the FBST test statistic in a nonparametric context, (2) the implicit relation between the formula for the FBST test statistic and the differential entropy estimation, and (3) the potential to extend frequentist consistency results from the literature to this method. These attributes collectively establish a robust framework for nonparametric hypothesis testing that is mathematically rigorous, interpretable through the lens of information geometry, and consistent from a frequentist standpoint.

This paper is structured as follows. Section 2 outlines the essential definitions and properties of our proposed methodology. Section 3 provides simulations demonstrating the statistical power of our test. Finally, Section 4 offers a discussion of our findings and potential avenues for future research. The proof of our results are presented in Appendix A.

2. FBST for Random Histograms

We start with a formal definition of our model. To maintain clarity, we will restrict our analysis to densities on $[0,1]$.

Definition 2. *For $k \in \mathbb{N}$, consider the set of densities with support on $[0,1]$, defined as*

$$\mathcal{H}_k = \left\{ f \in L_1([0,1]) : f(t) = k \sum_{i=1}^{k} \mathbf{1}_{I_i}(t) w_i \text{ and } \sum_{i=1}^{k} w_i = 1, w_i \geq 0 \right\}$$

where $I_j = [(j-1)/k, j/k)$ for $j = 1, 2, \ldots, k$. A random histogram θ is a random variable that selects a random element of $\mathcal{H}_{k(n)}$.

The distribution of θ is fully characterized by the distribution of the vector of random weights $W = (W_1, \ldots, W_k)$. Bayesian posterior inference on θ may also be conducted with respect to W if the likelihood is given by

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \theta(X_i) \propto w_1^{\sum_{i=1}^{n} \mathbf{1}_{I_1}(X_i)} \ldots w_k^{\sum_{i=1}^{n} \mathbf{1}_{I_k}(X_i)}$$

which corresponds to the assumption that the sample values $X_1, \ldots, X_n | \theta$ are conditionally independent and share an identical density θ. In this paper, we shall consider random histograms sampled implicitly by Dirichlet priors of the weights W. This approach guarantees that the posterior inference on θ is conjugate and computationally tractable, as it is equivalent to inference on a Dirichlet-Multinomial Bayesian model.

Proposition 1. *Consider θ a random histogram with weights $W \sim Dirichlet(\alpha_1, \ldots, \alpha_k)$. If $X_i \perp X_j | \theta$, $i \neq j$ and $X_i | \theta \sim \theta$, then the posterior $\theta | \mathbf{X}_n$ remains a random histogram with weights*

$$W | \mathbf{X}_n \sim Dirichlet(\alpha_1 + N_1, \ldots, \alpha_k + N_k) \tag{3}$$

where $N_j = \sum_{i=1}^{n} \mathbf{1}_{I_j}(X_i), 1 \leq j \leq k$.

A usual approach for Bayesian nonparametric inference on a histogram is letting $k(n)$ grow slowly with the sample size n. This may be interpreted as a data-dependent prior; the full parameter space being considered is the set of all densities and, contingent on n, random histograms puts mass only on specific subsets of this set. One could define priors that do not depend on n, but this would come at a heavy computational cost. Moreover, meaningful and computationally sound inference might be conducted both in frequentist and Bayesian perspective if we also require the priors of w to depend on n [2,4].

Fixing n and $k(n)$, the original definition of $ev(\Theta_0; \mathbf{X}_n)$ may be adapted to conduct tests regarding $\theta | \mathbf{X}_n$. Given that there exists a bijection between an element of $\mathcal{H}_{k(n)}$ and its corresponding weights $(W_1, \ldots, W_{k(n)})$, the FBST test statistic may be defined in terms of the Dirichlet distribution defined in Equation (3). However, this approach comes at the price of being able to only test hypothesis of the form $\Theta_0 \subset \mathcal{H}_{k(n)}$. Therefore, if a researcher is interested in testing a hypothesis framed in terms of a general Θ_0, our proposed procedure specifies a test statistic based on its finite-dimensional counterpart. As $k(n)$ is permitted to increase with the sample size, this translation process becomes increasingly negligible.

Definition 3 (FBST for random histograms)**.** *Let θ be a random histogram defined by Dirichlet weights, and $\mathbf{X}_n | \theta$ represent an i.i.d. sample drawn from θ. The FBST test statistic for testing a hypothesis $H_0 : \theta \in \Theta_0$, where Θ_0 is an arbitrary set of densities on $[0,1]$, is given by*

$$ev(\Theta_0; \mathbf{X}_n) = \Pi_{\theta|\mathbf{X}_n}\left(\sum_{j=1}^{k(n)} (N_j + \alpha_j)\log(W_j) \leq \sup_{\mathbf{w}\in\mathcal{S}(\Theta_0)} \sum_{i=1}^{k(n)} (N_j + \alpha_j)\log w_j\right) \quad (4)$$

where $\mathcal{S}(\Theta_0)$ *denotes probabilities attributed to the sets* $I_1, \ldots, I_{k(n)}$ *by each element of* Θ_0*:*

$$\mathcal{S}(\Theta_0) = \left\{\left(\int_{I_1} f(t)dt, \int_{I_2} f(t)dt, \ldots, \int_{I_{k(n)}} f(t)dt\right) : f \in \Theta_0\right\}$$

The FBST for random histograms may be interpreted in the context of information theory. Let p and q be two m dimensional probability vectors. We recall that the cross-entropy divergence $H(p,q)$ between those vectors is given by $-\sum_{j=1}^{m} p_i \log q_i$ and the Kullback–Leibler divergence $D_{KL}(p||q)$ is given by $\sum p_i \log\left(\frac{p_i}{q_i}\right)$. By introducing $\hat{w}_n = \left(\frac{N_1+\alpha_j}{n+\sum_{j=1}^{k(n)}\alpha_j}, \ldots, \frac{N_{k(n)}+\alpha_j}{n+\sum_{j=1}^{k(n)}\alpha_j}\right)$, Equation (3) can be articulated as

$$ev(\Theta_0; \mathbf{X}_n) = \Pi_{\theta|\mathbf{X}_n}\left(-H(\hat{w}_n, W) \leq \sup_{w\in\mathcal{S}(\Theta_0)} -H(\hat{w}_n, w)\right), \quad (5)$$

$$ev(\Theta_0; \mathbf{X}_n) = \Pi_{\theta|\mathbf{X}_n}\left(D_{KL}(\hat{w}_n||W) \geq \inf_{w\in\mathcal{S}(\Theta_0)} D_{KL}(\hat{w}_n||w)\right). \quad (6)$$

These equations demonstrate that the application of the FBST definition leads to statistical tests grounded in an information-theoretic measure of divergence between a distribution of the sample into $k(n)$ bins and the expected value of counts on those same bins under the assumption that θ is some hypothesized density f. Indeed, in the context of this particular test, a related concept has emerged in the literature on goodness-of-fit testing, notably in G-tests [6] and other methodologies rooted in frequentist nonparametric estimates of the continuous variant of the Kullback–Leibler divergence for probability densities [7]. Both tests utilize a χ^2 asymptotic distribution under the null hypothesis. For the FBST, there are specific rates of increase of $k(n)$ that ensure the presence of analogous results.

Theorem 1. *If (1)* $\mathbf{X}_n$ *is an independent and identically distributed sample of* X_1 *with density* f^* *Lipchitz continuous on* $[0,1]$*; (2)* θ *is a random histogram satisfying* $M > \alpha_i > m$ *for all i and fixed quantities m, M and (3)* $k(n) = \frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}$*, for any* $\epsilon > 0$*, then the FBST for random histograms with* $H_0 : \theta = f_0$ *satisfies*

1. $ev(\Theta_0; \mathbf{X}_n) \to^{\mathcal{L}} Unif(0,1)$ *if* $f_0 = f^*$ *and*
2. $ev(\Theta_0; \mathbf{X}_n) \to^{\mathbb{P}} 0$ *if* $f_0 \neq f^*$*, where* $\to^{\mathbb{P}}$ *denotes convergence in probability with respect to* $\mathbf{X}_n$*.*

One particular virtue of Equation (4) is the simplicity of the optimization step in the FBST. This fact is due the convexity of the cross-entropy functional. Also, this optimization will be able to reject false null hypotheses as the sample size grows larger, as we exemplify for the case of fixed parametric families.

Theorem 2. *Let* P_α *be a parametric family of differentiable distribution functions* f_α*, such that* $\min ||f_\alpha - f^*||_2^2 = \epsilon > 0$ *and* $\Theta_0^k = \{(F_\alpha(I_1), ..., F_\alpha(I_{k(n)})) : F_\alpha \in P_\alpha\}$ *is the corresponding subset of the* $k(n)$ *dimensional simplex. Then, the FBST on histograms for goodness-of-fit of this parametric family satisfies* $ev(\Theta_0^k) \to 0$ *in* f^* *probability.*

This procedure is similar to other nonparametric methods that do not rely on maximum likelihood estimates for testing, but instead optimize specific statistics. This idea dates back to Berkson's suggestion to minimize chi-squared rather than maximize likelihood [8], although there have been few attempts to directly optimize test statistics, such as the Kolmogorov–Smirnov statistic [9]. This may be because optimizing usual test statistics for goodness-of-fit, such as Kolmogorov–Smirnov, Anderson–Darling, and Cramer–von Mises [10], requires specialized optimization procedures, like the one developed in [9].

Alternatively, the most common approach for testing adherence to a parametric family of distributions involves estimating parameters by maximum likelihood and then deriving the null distribution of an existing test through resampling [11]. Our new test, as we will demonstrate in simulations, could also require corrections when the optimization suggested by Theorem 2 is used.

3. Simulations

In this section, we will compare the statistical power of our test with that of other available techniques through simulations. Both simple and composite null hypotheses will be considered. Simulations will be conducted using the R programming language [12] and its public repository of packages. For simple hypotheses, the following tests are compared:

1. The e-value for histograms, as defined in Definition 3, adopting with $\alpha_{i,j} = 1$, with the number of bins defined as the hypothesis of Theorem 1;
2. Classic Kolmogorov–Smirnov (KS) test, as described in [6];
3. Alternative versions of KS, AD, and CV, constructed by [10], implemented by the R Package [13].

Following [10], we shall compare the NP-FBST of Definition 3 with

$$k(n) = \max\{2, n^{1/6}\log(n)^{-1/6-1}\}$$

using sample sizes $n = 10, 20, 30, 50, 70, 100, 150, 200, 300$. The null hypothesis tested shall be $H_0 : f^* = Unif(0,1)$ and H_1 will be simulated in 4 scenarios: $Beta(1.6, 1.6)$, $Beta(1.3, 1.3)$, $Beta(0.8, 0.8)$, and $Beta(0.6, 0.6)$. We calculate the statistical power as the % of correct rejection of H_0, with rejection at the 5% level, on 500 Monte Carlo sample. The results are summarized in Figure 1.

Analyzing Figure 1, we observe the following:

- For $\alpha > 1$ and $\beta > 1$, the NP-FBST power may be much more powerful for small sample sizes, but it is still competitive for large sample sizes.
- For $\alpha < 1$ and $\beta < 1$, which are non-Lipchitz alternative hypotheses, the test performs worse than Zhang's alternatives [10]. However, it still shows comparable or superior power compared to the usual Kolmogorov–Smirnov statistic.
- For sample sizes below 1000, $k(n)$ will usually be very small, such as 2 or 3, so the test is just a regular multinomial test.

To showcase the approximation properties of the NP-FBST based on Lemma A1, we shall simulate one last example, but this time we will adopt $k(n) = \log_2(n)$, usually referred as Sturge's Law, for histogram binning [14]. This is, of course, appropriate asymptotically, as $\log_2(n) \in o(n^{1/6}/\log(n)^{-1/6-\epsilon})$ for all $\epsilon > 0$. However, this produces very competitive statistical power for testing if a $Beta(2, 2)$ is a Irwin-Hall(2), as highlighted in Figure 2.

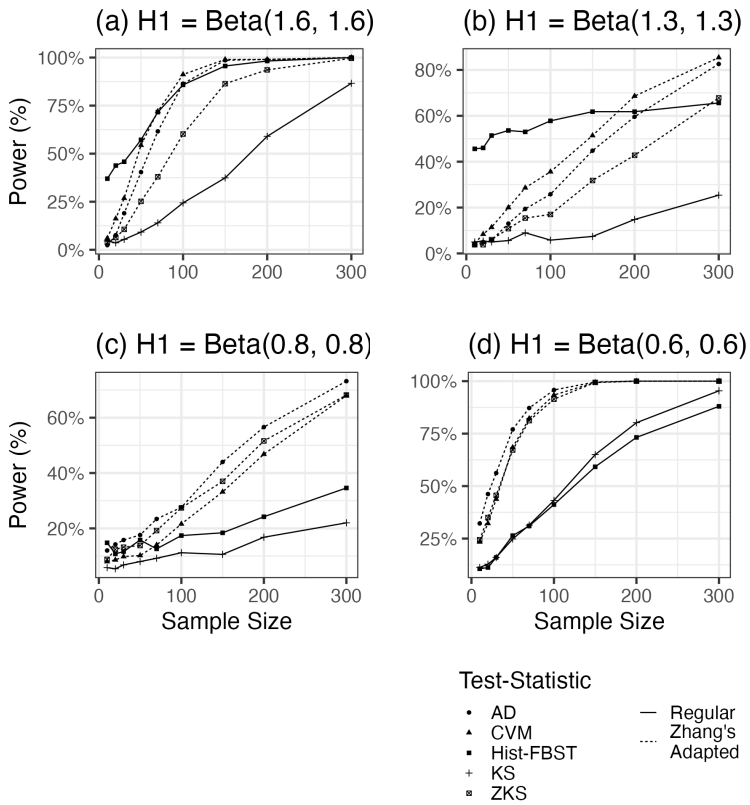

Figure 1. Simulated power for uniformity tests of several test statistics under specific simulated H_1. All power percentages were calculated considering 500 Monte Carlo samples. (**a**) $H_1 : Beta(1.6, 1.6)$, (**b**) $H_1 : Beta(1.3, 1.3)$, (**c**) $H_1 : Beta(0.8, 0.8)$, (**d**) $H_1 : Beta(0.6, 0.6)$.

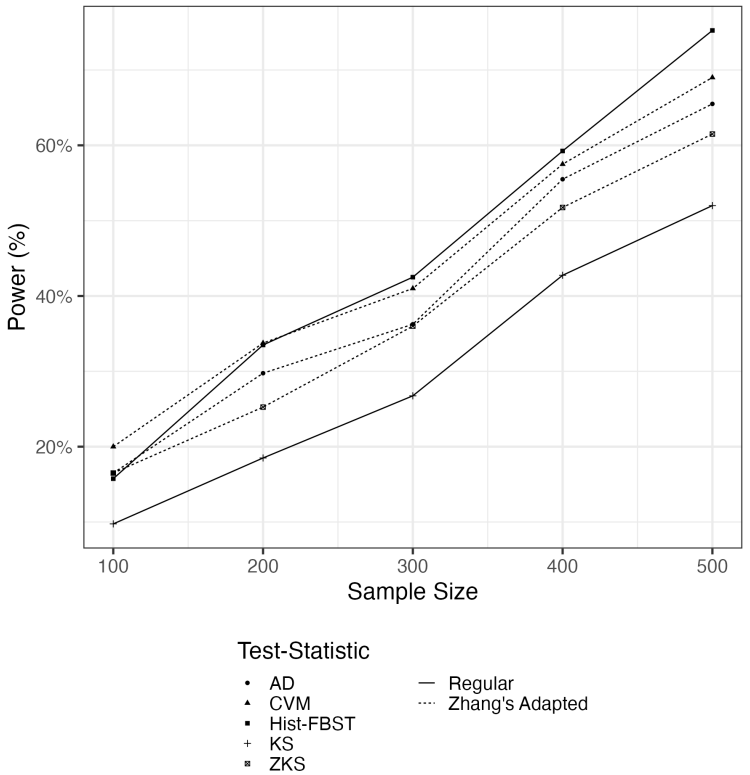

Figure 2. Simulated power for several tests with $H_0 : f^* = Beta(2, 2)$ under $H_1 : f^* =$ Irwin-Hall(2).

4. Conclusions

In this paper, we presented a new nonparametric Bayesian procedure extending the usual FBST for Bayesian histograms. We summarize our results as practical and theoretical.

On the practical and applied front, we draw the following conclusions:

- For small sample sizes, our method is competitive in terms of statistical power, even compared to sophisticated alternatives such as Zhang's tests [10].
- For larger sample sizes, the very slow sample size growth required by Theorem 1 harms the statistical power. Therefore, other binning rules could be considered. Further research shall look for an adaptable number of bins. Desirable binning rules should be larger than $n^{1/6}\log(n)^{1/6+\epsilon}$ for small sample sizes, but smaller for large sample sizes. Our simulations suggest that the usual $k(n) = O(\log_2(n))$, known as Sturge's Law, is a competitive alternative for moderate sample sizes lower than 1000.
- Unlike previous attempts, our method is computationally inexpensive with competitive statistical power.

From a theoretical perspective, we derive the following conclusions:

- The natural Dirichlet-Multinomial formulation of Bayesian histograms induces statistical tests based on estimates of Kullback–Leibler divergences. This formulation logically follows from the definition of the FBST, and the same logic could be applied to other Bayesian density estimation methods. The frequentist properties of other versions of this NP-FBST remain to be studied in future works, but our results highlight the Kullback–Leibler divergence as a possible "canonical" statistic for nonparametric versions of the FBST for density estimation.
- Our results show that taking the limit of a slowly increasing finite-dimensional parameter space is a viable strategy for building nonparametric versions of the FBST. The frequentist properties of the FBST are intimately related to the Bernstein–von Mises theorem. Therefore, if these types of Gaussian approximations are available, our arguments should also hold. In fact, all the main references of this paper build specific growth rates of the dimension of the parameter space and could be used to find other versions of nonparametric FBSTs [3–5].
- For composite hypotheses, our method can be used both for testing based on the maximum likelihood estimate of nuisance parameters and for directly optimizing the test statistics, which may be interpreted as a weighted likelihood function. Usual numerical methods for optimizing the likelihood will work for our test statistic, which is not the case for other usual statistics such as Anderson–Darling, Cramer–von Mises, or Kolmogorov–Smirnov.

For future research, we highlight that adaptively choosing the number of bins $k(n)$ is crucial, as the statistical power is heavily influenced by this quantity. Additionally, Theorem 1 requires a Lipschitz continuous data-generating density, a usual assumption for histograms, but excludes unbounded densities, which are important from a practical and theoretical point of view. Extending our results to Hölder continuous densities is particularly important but requires the derivation of other versions of Bernstein–von Mises theorems.

Author Contributions: Conceptualization, J.M.S. and R.B.S.; methodology, R.B.S.; software, F.C.; validation, R.B.S. and J.M.S.; formal analysis, R.B.S.; investigation, F.C.; writing—original draft preparation, F.C.; writing—review and editing, J.M.S. and R.B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CAPES grant number 88887.613569/2021-00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All code and simulated datasets are available on https://github.com/azeloc/histograms.maxent.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FBST	Full Bayesian Significance Testing
NP	Nonparametric

Appendix A. Proofs

The following lemma is the main ingredient of the proof. As $k(n)$ grows slowly, we are able to employ a normal approximation to $\sqrt{n}\hat{J}(\theta - \hat{h})$, where J is the square root of the Fisher information matrix, evaluated at $\hat{h}$.

Lemma A1 (Adapted from Theorem 2.4 and the discussion of Section 3 of [3])**.** *Consider*

- f^* *a Lipchitz p.d.f. on [0, 1];*
- f^* *the posterior obtained using the model defined in Definition 2, with* $k(n) = n^{1/6}\log(n)^{-1/6-\epsilon}$ *with a fixed* $\epsilon > 0$*;*
- $J(\theta)$ *as the square root of the multinomial Fisher information matrix evaluated at* $\theta \in \mathcal{S}^{k(n)}$*, and* $\hat{J} = J(\hat{h})$ *the sample estimate of* $J(\theta)$*;*
- $\hat{D}$ *as the diagonal matrix with* $\hat{h}$ *in its entries.*

Then:

1. $f_{\theta|\mathbf{X}_n}(\{\theta : ||\theta - \theta_0|| \geq \epsilon\}) \to 0$ *in* f^* *probability;*
2. $\sqrt{n}\hat{J}(\hat{h} - \theta_0) \to N(0,1)$ *in distribution;*
3. *the largest eigenvalue of* $\hat{J}^2$ *is* $O(k(n)^2)$*;*
4. $$\int \left| f_{\theta|\mathbf{X}_n}(\theta) - \Phi(v; \hat{h}, \hat{J}^2/n) \right| d\theta \to 0 \text{ in } f^* \text{ probability} \tag{A1}$$

Proof of Theorem 1. Let N_n and Φ_n be the measure and density induced by the Gaussian approximation of Equation (A1). It follows from Equation (A1) that $|\Pi_n(A) - N_n(A)| \to 0$ in probability for all measurable A. Equation (A1) also implies that the following expectation vanishes in probability:

$$\int f_{\theta|\mathbf{X}_n}(\theta) \frac{\left| f_{\theta|\mathbf{X}_n}(\theta) - \Phi_n(\theta) \right|}{f_{\theta|\mathbf{X}_n}(\theta)} d\theta = \mathbb{E}_{\theta|\mathbf{X}_n}\left[\frac{\left| f_{\theta|\mathbf{X}_n}(\theta) - \Phi_n(\theta) \right|}{f_{\theta|\mathbf{X}_n}(\theta)} \right] \to 0.$$

Now let $B_\epsilon = \{\theta : |\frac{\Phi_n(\theta)}{f_{\theta|\mathbf{X}_n}(\theta)} - 1| \leq \epsilon\}$. By Markov's inequality it follows that

$$\Pi_{\theta|\mathbf{X}_n}(B_\epsilon^c) \leq \frac{\mathbb{E}_{\theta|\mathbf{X}_n}\left[\frac{|f_{\theta|\mathbf{X}_n}(\theta) - \Phi_n(\theta)|}{f_{\theta|\mathbf{X}_n}(\theta)} \right]}{\delta} \to 0$$

in f^* probability for all ϵ.

Now, consider

$$T'(\theta_0) = \{\theta : \Phi_n(\theta) \leq \Phi_n(\theta_0)\}$$

and for a fixed $\gamma > 0$

$$C_\gamma = \{||\theta - \theta_0|| < \gamma\}.$$

One may verify that $(T'(\theta_0) \cap B_{\epsilon_1} \cap C_\delta) \subset T(\theta_0)$ and $(T(\theta_0) \cap B_{\epsilon_2} \cap C_\delta) \subset T'(\theta_0)$ for any choice of ϵ_1, ϵ_2 and δ. It follows that

$$\Pi_{\theta|\mathbf{X}_n}(T'(\theta_0)) = \Pi_{\theta|\mathbf{X}_n}(T'(\theta_0) \cap (B_\epsilon \cap C_\delta)) + \Pi_{\theta|\mathbf{X}_n}(T'(\theta_0) \cap (B_\epsilon \cap C_\delta)^c) \leq$$

$$\Pi_{\theta|\mathbf{X}_n}(T(\theta_0)) + \Pi_{\theta|\mathbf{X}_n}(B_\epsilon^c \cup C_\delta^c)$$

and analogously

$$\Pi_{\theta|\mathbf{X}_n}(T(\theta_0)) \leq \Pi_{\theta|\mathbf{X}_n}(T'(\theta_0)) + \Pi_{\theta|\mathbf{X}_n}(B_\epsilon^c \cup C_\delta^c).$$

Therefore

$$|\Pi_{\theta|\mathbf{X}_n}(T(\theta_0)) - \Pi_{\theta|\mathbf{X}_n}(T'(\theta_0))| \leq \Pi_{\theta|\mathbf{X}_n}(B_\epsilon^c) + \Pi_{\theta|\mathbf{X}_n}(C_\delta^c) \to 0$$

in probability. Finally we conclude that the last convergence implies that

$$|\Pi_{\theta|\mathbf{X}_n}(T(\theta_0)) - N_n(T'(\theta_0))| \leq$$

$$|\Pi_{\theta|\mathbf{X}_n}(T(\theta_0)) - \Pi_{\theta|\mathbf{X}_n}(T'(\theta_0))| + |\Pi_{\theta|\mathbf{X}_n}(T'(\theta_0)) - N_n(T'(\theta_0))| \to 0$$

in probability. This conclusion ensures that the e-value statistic may be approximated by $N_n(T'(\theta_0))$. Now, note that as $N_n(T'(\theta_0))$ is based on a quadratic form, this probability shall might expressed in terms of the χ^2 distribution:

$$\begin{aligned} N_n(T'(\theta_0)) = \\ N_n\Big(n(\theta - \hat{h})^T \hat{J}^{-2}(\theta - \hat{h}) \geq (\theta_0 - \hat{h})^T \hat{J}^{-2}(\theta_0 - \hat{h})\Big) = \\ 1 - \chi^2_{k(n)}\Big((\theta_0 - \hat{h})^T \hat{J}^{-2}(\theta_0 - \hat{h})\Big). \end{aligned} \tag{A2}$$

Also, note that χ^2 may be approximated in distribution by a Gaussian with same mean and variance. Therefore we may rewrite

$$N_n(T'(\theta_0)) \approx 1 - \Phi\left(\frac{(\theta_0 - \hat{h})^T \hat{J}^{-2}(\theta_0 - \hat{h}) - k(n)}{\sqrt{2k(n)}}\right) \tag{A3}$$

and also, as $\hat{h}$ is approximate Gaussian, $(\theta_0 - \hat{h})^T \hat{J}^{-2}(\theta_0 - \hat{h})$ is approximate $\chi^2_{k(n)}$ and therefore, $\frac{(\theta_0 - \hat{h})^T \hat{J}^{-2}(\theta_0 - \hat{h}) - k(n)}{\sqrt{2k(n)}} \to^D N(0,1)$. Therefore, by the continuous mapping theorem $1 - \Phi\left(\frac{(\theta_0 - \hat{h})^T \hat{J}^{-2}(\theta_0 - \hat{h}) - k(n)}{\sqrt{2k(n)}}\right)$ converges to a uniform distribution.

For part 2, note that under $H1$ $(\theta_0 - \hat{h})^T \hat{J}^{-2}(\theta_0 - \hat{h})$ might be approximated by a non-central chi-square with mean $k(n) + \sum(\theta_i^* - \theta_0^i)^2 \approx k(n) + n\int(f_0(t) - f^*(t))dt$, which is asymptotically larger than $\sqrt{k(n)}$. Hence,

$$\frac{(\theta_0 - \hat{h})^T \hat{J}^{-2}(\theta_0 - \hat{h}) - k(n)}{\sqrt{2k(n)}} \to \infty,$$

and then $N_n(T'(\theta_0))$ must converge to 0. □

Proof of Theorem 2. By Lemma A1 we have that the largest eigenvalue of J^{-2} is of order $k(n)^2$, and therefore the normal approximation obtained at Equation (A3) becomes

$$ev(\Theta_0^k) \approx 1 - \Phi\left(\frac{(\theta^{\alpha_*} - \hat{h})^T \hat{f}^{-2}(\theta^{\alpha_*} - \hat{h}) - k(n)}{\sqrt{2k(n)}}\right)$$

Now we note that the condition of the theorem implies that $\frac{(\theta^{\alpha_*} - \hat{h})^T \hat{f}^{-2}(\theta^{\alpha_*} - \hat{h}) - k(n)}{\sqrt{2k(n)}}$ is asymptotically larger than $\sqrt{k(n)}$. Then it follows that the RHS of the above quantity converges to 0 in f^* probability. □

References

1. de B. Pereira, C.A.; Stern, J.M.; Wechsler, S. Can a significance test be genuinely Bayesian? *Bayesian Anal.* **2008**, *3*, 79–100. [CrossRef]
2. Ghosal, S.; van der Vaart, A.W. *Fundamentals of Nonparametric Bayesian Inference*; Number 44 in Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, UK, New York, NY, USA, 2017; pp. xxiv+646. [CrossRef]
3. Ghosal, S. Asymptotic Normality of Posterior Distributions for Exponential Families when the Number of Parameters Tends to Infinity. *J. Multivar. Anal.* **2000**, *74*, 49–68. [CrossRef]
4. Castillo, I.; Nickl, R. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Stat.* **2014**, *42*, 1941–1969. [CrossRef]
5. Castillo, I.; Rousseau, J. A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Stat.* **2015**, *43*, 2353–2383. [CrossRef]
6. D'Agostino, R.B.; Stephens, M.A. *Goodness-of-Fit Techniques*; Marcel Dekker, Inc.: New York, NY, USA, 1986.
7. Song, K.S. Goodness-of-fit tests based on Kullback-Leibler discrimination information. *IEEE Trans. Inf. Theory* **2002**, *48*, 1103–1117. [CrossRef]
8. Berkson, J. Minimum Chi-Square, not Maximum Likelihood! *Ann. Stat.* **1980**, *8*, 457–487. [CrossRef]
9. Zvi Drezner, O.T.; Zerom, D. A Modified Kolmogorov—Smirnov Test for Normality. *Commun. Stat.-Simul. Comput.* **2010**, *39*, 693–704. [CrossRef]
10. Zhang, J. Powerful Goodness-of-Fit Tests Based on the Likelihood Ratio. *J. R. Stat. Soc. Ser. (Stat. Methodol.)* **2002**, *64*, 281–294. [CrossRef]
11. Babu, G.J.; Rao, C.R. Goodness-of-Fit Tests When Parameters Are Estimated. *Sankhyā Indian J. Stat. (2003–2007)* **2004**, *66*, 63–74.
12. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
13. Cui, N.; Zhou, M. *DistributionTest: Powerful Goodness-of-Fit Tests Based on the Likelihood Ratio*, 2020. R package version 1.1. Comprehensive R Archive Network (CRAN). Available online: https://CRAN.R-project.org/package=DistributionTest (access on 31 July 2024).
14. Sturges, H.A. The Choice of a Class Interval. *J. Am. Stat. Assoc.* **1926**, *21*, 65–66. [CrossRef]