Non Parametric Full Bayesian Significance Testing for bayesian histograms

F. Corrêa Filho[1] R. B. Stern[1] J. M. Stern [1]

[1]Institute of Mathematics and Statistics
University of São Paulo

MaxEnt, July 2024

Table of Contents

1. Introduction — Parametric vs Nonparametric FBST
2. Bayesian Histograms
3. Entropy
4. Theorems
5. Simulations
6. Conclusion

Table of Contents

1. Introduction — Parametric vs Nonparametric FBST
2. Bayesian Histograms
3. Entropy
4. Theorems
5. Simulations
6. Conclusion

Parametric FBST

Let Θ be a finite dimensional space such as a subset of $\mathbb{R}^p$, $p \in \mathbb{N}$ The FBST [BSW08] consists on deciding whether to reject some statistical hypothesis $H_0 : \theta = \theta_0$ based on the posterior probability of the tangent set $T(\theta_0)$.

- First we define the posterior density $f_{\theta|\mathbf{x}_n}(\theta) = \frac{p(\mathbf{x}_n|\theta)p(\theta)}{\int p(\mathbf{x}_n|\theta)p(\theta)}$

Parametric FBST

Let Θ be a finite dimensional space such as a subset of $\mathbb{R}^p$, $p \in \mathbb{N}$ The FBST [BSW08] consists on deciding whether to reject some statistical hypothesis $H_0 : \theta = \theta_0$ based on the posterior probability of the tangent set $T(\theta_0)$.

- First we define the posterior density $f_{\theta|\mathbf{x}_n}(\theta) = \frac{p(\mathbf{x}_n|\theta)p(\theta)}{\int p(\mathbf{x}_n|\theta)p(\theta)}$
- Then we find the set in which $f_{\theta|\mathbf{x}_n}(\theta)$ is smaller than $f_{\theta|\mathbf{x}_n}(\theta_0)$. This is $T(\theta_0)$

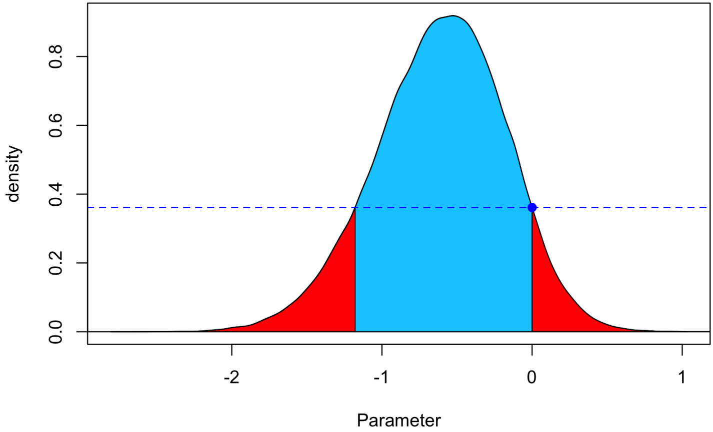
Parametric FBST

Let Θ be a finite dimensional space such as a subset of $\mathbb{R}^p$, $p \in \mathbb{N}$ The FBST [BSW08] consists on deciding whether to reject some statistical hypothesis $H_0 : \theta = \theta_0$ based on the posterior probability of the tangent set $T(\theta_0)$.

- First we define the posterior density $f_{\theta|\mathbf{x}_n}(\theta) = \frac{p(\mathbf{x}_n|\theta)p(\theta)}{\int p(\mathbf{x}_n|\theta)p(\theta)}$
- Then we find the set in which $f_{\theta|\mathbf{x}_n}(\theta)$ is smaller than $f_{\theta|\mathbf{x}_n}(\theta_0)$. This is $T(\theta_0)$
- Finally calculate the posterior probability of $T(\theta_0)$. For point hypothesis we shall write:

$$ev(\theta_0) = \int_{T(\theta_0)} f_{\theta|\mathbf{x}_n}(t)dt = \Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}(T(\theta_0))$$

FBST in a nutshell: Are there many θ that are less likely than θ_0 (regarding the posterior)? If there are many, $ev >> 0$ then we don't reject H_0. If there are few, $ev \approx 0$, then we reject H_0 (almost all Θ are more likely than θ_0)



Nonparametric FBST?

$f_{\theta|\mathbf{x}_n}$ is the density of $\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}$ with reference to the Lebesgue Measure

Nonparametric FBST?

$f_{\theta|\mathbf{x}_n}$ is the density of $\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}$ with reference to the Lebesgue Measure

Remark

When Θ is infinite dimensional there is no Lebesgue Measure (translation invariant Borel measures exists, but are pathologic) $\implies$ we don't have a canonical version of the density of $\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}$ even when we have some posterior measure over infinite dimensional parameters spaces (e.g. Gaussian Process, Bayesian Kernel Density Estimation, Mixture of Dirichlet etc)

Nonparametric FBST?

$f_{\theta|\mathbf{x}_n}$ is the density of $\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}$ with reference to the Lebesgue Measure

Remark

When Θ is infinite dimensional there is no Lebesgue Measure (translation invariant Borel measures exists, but are pathologic) $\implies$ we don't have a canonical version of the density of $\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}$ even when we have some posterior measure over infinite dimensional parameters spaces (e.g. Gaussian Process, Bayesian Kernel Density Estimation, Mixture of Dirichlet etc)

FBST depends on some notion of posterior density. In infinite dimensions we must look for alternatives

Table of Contents

1. Introduction — Parametric vs Nonparametric FBST
2. Bayesian Histograms
3. Entropy
4. Theorems
5. Simulations
6. Conclusion

Bottom up approach — Bayesian Histograms

Let's start with an example of an histogram

Let $\mathbf{X}_n \sim Beta(2, 2)$ be an i.i.d. sample of size n and $k(n) = round(\log_2(n))$ the number of bins as a function of the sample size.

Bottom up approach — Bayesian Histograms

Let's start with an example of an histogram
Let $\mathbf{X}_n \sim Beta(2,2)$ be an i.i.d. sample of size n and $k(n) = round(\log_2(n))$ the number of bins as a function of the sample size. For example we could have, $n = 1000$, $k(n) = 10$. Consider the counts of the observed sample $\mathbf{x}_n$ in discrete bins $I_1 = [0, 1/10)$, $I_2 = [1/10, 2/20)$, ..., $I_10 = (9/10, 1]$ are
$(n_1, n_2, ..., n_{10}) = (50, 114, 161, 173, 196, 165, 99, 42)$

Bottom up approach — Bayesian Histograms

Let's start with an example of an histogram
Let $\mathbf{X}_n \sim Beta(2,2)$ be an i.i.d. sample of size n and $k(n) = round(\log_2(n))$ the number of bins as a function of the sample size. For example we could have, $n = 1000$, $k(n) = 10$. Consider the counts of the observed sample $\mathbf{x}_n$ in discrete bins $I_1 = [0, 1/10)$, $I_2 = [1/10, 2/20)$, ..., $I_10 = (9/10, 1]$ are
$(n_1, n_2, ..., n_{10}) = (50, 114, 161, 173, 196, 165, 99, 42)$
Bayesian histograms perform inference on the probability of each bin:

$$(Prior)(\theta_1, ..., \theta_{10}) \sim Dirichlet(\alpha_{1,10}, ..., \alpha_{10,10})$$

$$(Likelihood)(n_1, ..., n_{10})|(\theta_1, ..., \theta_{10}) \sim Multinom(n, (\theta_1, ..., \theta_{10}))$$

$$(\theta_1, ..., \theta_{10})|(n_1, ..., n_{10}) \sim Dirichlet(n_1 + \alpha_{1,10}, ..., n_{10} + \alpha_{10,10})$$

Bottom up approach — Bayesian Histograms (formal)

Let $k : \mathbb{N}^* \to \mathbb{N}^*$ be an increasing function of the sample size n

Bottom up approach — Bayesian Histograms (formal)

Let $k : \mathbb{N}^* \to \mathbb{N}^*$ be an increasing function of the sample size n
Also, consider a partition of $[0, 1]$ defined by $I_j = [(j-1)/k(n), j/k(n))$ for $j = 1, 2, \ldots, k(n)$. Let $\mathcal{S}_{k(n)}$ be the unit simplex in $\mathbb{R}^{k(n)}$ and denote by:

$$\mathcal{H}_{k(n)} = \left\{ f : f(x) = k(n) \sum_{j=1}^{k(n)} w_j \mathbb{I}_{I_j}(x), w \in \mathcal{S}_{k(n)} \right\} \tag{1}$$

the set of histograms which are densities with $k(n)$ bins.

Bottom up approach — Bayesian Histograms (formal)

Let $k : \mathbb{N}^* \to \mathbb{N}^*$ be an increasing function of the sample size n
Also, consider a partition of $[0,1]$ defined by $I_j = [(j-1)/k(n), j/k(n))$ for $j = 1, 2, \ldots, k(n)$. Let $\mathcal{S}_{k(n)}$ be the unit simplex in $\mathbb{R}^{k(n)}$ and denote by:

$$\mathcal{H}_{k(n)} = \left\{ f : f(x) = k(n) \sum_{j=1}^{k(n)} w_j \mathbb{I}_{I_j}(x), w \in \mathcal{S}_{k(n)} \right\} \tag{1}$$

the set of histograms which are densities with $k(n)$ bins.
We'll define a prior on $\mathcal{H}_{k(n)}$ by defining a prior on $(w_1, \ldots, w_{k(n)})$.
Conditionally on $k(n)$, we consider the following Dirichlet priors:

$$w \sim \mathcal{D}(\alpha_{1,k(n)}, \ldots, \alpha_{k(n),k(n)}), \quad c_1 \leq \alpha_{j,k} \leq c_2 \tag{2}$$

A Bayesian histogram posterior on $\mathcal{H}_{k(n)}$ is the conjugated Dirichlet-Multinomial posterior based how many sample points ended up in I_j, the counts n_j

Bottom up approach — Bayesian Histograms (formal)

We define the posterior random histogram as:

$$h(t) = k(n) \sum \theta_j \mathbb{I}_{I_j}(t), \tag{3}$$

where $(\theta_1, \ldots, \theta_{k(n)})$ follows the posterior Dirichlet distribution previously highlighted. The posterior estimate of the histogram is $\hat{h}$ pluging the MAP on 3:

$$\hat{h}(t) = \sum \frac{n_i + \alpha_{i,k(n)}}{n + \sum \alpha_{j,k(n)}} \mathbb{I}_{I_j}(t) \tag{4}$$

Bayesian Histograms (example)

The shaded area represents the uncertainty encoded in the Dirichlet posterior, the blue bars represents $\hat{h}$ and the red line represents f^*, a $Beta(2,2)$. Sample size $N = 1000$, $k(n) = 16$.

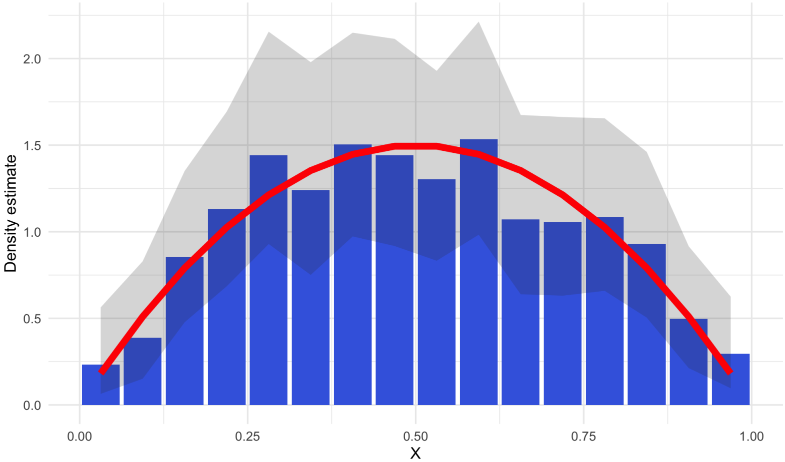


Table of Contents

1. Introduction — Parametric vs Nonparametric FBST
2. Bayesian Histograms
3. Entropy
4. Theorems
5. Simulations
6. Conclusion

Why bayesian histograms? — Relation to entropy

Log likelihood in this model is cross-entropy $\implies$ e-value test statistic for $H_0 : \theta = p$ is the Kullback-Leibler Divergence between $\hat{\theta}$ and p, one element of the simplex

$$ev(p) = \Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}(f_{\theta|\mathbf{x}_n}(\theta) \leq f_{\theta|\mathbf{x}_n}(p)) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(\prod_{i=1}^{k(n)} \theta_i^{\alpha_{i,k(n)}+n_i} \leq \prod_{i=1}^{k(n)} p_i^{\alpha_{i,k(n)}+n_i}\right) =$$

Why bayesian histograms? — Relation to entropy

Log likelihood in this model is cross-entropy $\implies$ e-value test statistic for $H_0 : \theta = p$ is the Kullback-Leibler Divergence between $\hat{\theta}$ and p, one element of the simplex

$$ev(p) = \Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}(f_{\theta|\mathbf{x}_n}(\theta) \leq f_{\theta|\mathbf{x}_n}(p)) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(\prod_{i=1}^{k(n)} \theta_i^{\alpha_{i,k(n)}+n_i} \leq \prod_{i=1}^{k(n)} p_i^{\alpha_{i,k(n)}+n_i}\right) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(\sum_{i=1}^{k(n)} (\alpha_{i,k(n)} + n_i) \log \theta_i \leq \sum_{i=1}^{k(n)} (\alpha_{i,k(n)} + n_i) \log p_i\right) =$$

Why bayesian histograms? — Relation to entropy

Log likelihood in this model is cross-entropy $\implies$ e-value test statistic for $H_0 : \theta = p$ is the Kullback-Leibler Divergence between $\hat{\theta}$ and p, one element of the simplex

$$ev(p) = \Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}(f_{\theta|\mathbf{x}_n}(\theta) \leq f_{\theta|\mathbf{x}_n}(p)) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(\prod_{i=1}^{k(n)} \theta_i^{\alpha_{i,k(n)}+n_i} \leq \prod_{i=1}^{k(n)} p_i^{\alpha_{i,k(n)}+n_i}\right) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(\sum_{i=1}^{k(n)}(\alpha_{i,k(n)} + n_i)\log\theta_i \leq \sum_{i=1}^{k(n)}(\alpha_{i,k(n)} + n_i)\log p_i\right) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(H(\hat{\theta},\theta) > H(\hat{\theta},p)\right) = \Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(D_{KL}(\hat{\theta},\theta) > D_{KL}(\hat{\theta},p)\right) \tag{5}$$

Why bayesian histograms? — Relation to entropy

Log likelihood in this model is cross-entropy $\implies$ e-value test statistic for $H_0 : \theta = p$ is the Kullback-Leibler Divergence between $\hat{\theta}$ and p, one element of the simplex

$$ev(p) = \Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}(f_{\theta|\mathbf{x}_n}(\theta) \leq f_{\theta|\mathbf{x}_n}(p)) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(\prod_{i=1}^{k(n)} \theta_i^{\alpha_{i,k(n)}+n_i} \leq \prod_{i=1}^{k(n)} p_i^{\alpha_{i,k(n)}+n_i}\right) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(\sum_{i=1}^{k(n)}(\alpha_{i,k(n)}+n_i)\log\theta_i \leq \sum_{i=1}^{k(n)}(\alpha_{i,k(n)}+n_i)\log p_i\right) =$$

$$\Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(H(\hat{\theta},\theta) > H(\hat{\theta},p)\right) = \Pi_{\theta|\mathbf{X}_n=\mathbf{x}_n}\left(D_{KL}(\hat{\theta},\theta) > D_{KL}(\hat{\theta},p)\right) \tag{5}$$

The continuous version of the last line might be used on any density estimation problem!

Why bayesian histograms? — Consistency results

Frequentist consistency theorems for FBST relies on Bernstein-Von Mises type of theorems

Why bayesian histograms? — Consistency results

Frequentist consistency theorems for FBST relies on Bernstein-Von Mises type of theorems

Lemma adapted from [Gho00]

If f^* is a Lipchitz continuous density on $[0,1]$ and $k(n) = O\left(\frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}\right)$ for any $\epsilon > 0$ then:

1. θ posterior distribution is approximately a multivariate gaussian $N(\hat{\theta}, \hat{F}/\sqrt{n})$, where $\hat{F}$ is the Fisher Information Matrix evaluated at $\hat{\theta}$. The approximation holds in the total variation distance between both measures.
2. $||\hat{\theta} - \theta^{f^*}|| \to 0$ in f^* probability
3. $\hat{\theta}$ is also a approximately multivariate gaussian $N(\theta^{f^*}, \hat{F}/\sqrt{n})$

Why bayesian histograms? — Consistency results

Frequentist consistency theorems for FBST relies on Bernstein-Von Mises type of theorems

Lemma adapted from [Gho00]

If f^* is a Lipchitz continuous density on $[0,1]$ and $k(n) = O\left(\frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}\right)$ for any $\epsilon > 0$ then:

1. θ posterior distribution is approximately a multivariate gaussian $N(\hat{\theta}, \hat{F}/\sqrt{n})$, where $\hat{F}$ is the Fisher Information Matrix evaluated at $\hat{\theta}$. The approximation holds in the total variation distance between both measures.
2. $||\hat{\theta} - \theta^{f^*}|| \to 0$ in f^* probability
3. $\hat{\theta}$ is also a approximately multivariate gaussian $N(\theta^{f^*}, \hat{F}/\sqrt{n})$

The proof is very sophisticated, but it mostly guarantees that the lowest eigenvalue of the Fisher Information Matrix is large in comparison to $1/\sqrt{n}$ so the model is *approximately* parametric

Table of Contents

1 Introduction — Parametric vs Nonparametric FBST

2 Bayesian Histograms

3 Entropy

4 Theorems

5 Simulations

6 Conclusion

Sketch of proof — Theorem 1

Let $ev(\theta^f)$ be the FBST for testing $H_0 : \theta = \theta^f$ and θ^f be the projection of a continuous density f on probabilities for $k(n)$ bins. If f^* is a Lipchitz continuous density on $[0, 1]$ and $k(n) = O\left(\frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}\right)$ then:

1. $ev(\theta^{f^*}) \rightarrow^L Unif(0, 1)$
2. If $g \neq f^*$ then $ev(\theta^g) \rightarrow^P 0$ in f^* probability

Sketch of proof — Theorem 1

Let $ev(\theta^f)$ be the FBST for testing $H_0 : \theta = \theta^f$ and θ^f be the projection of a continuous density f on probabilities for $k(n)$ bins. If f^* is a Lipchitz continuous density on $[0,1]$ and $k(n) = O\left(\frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}\right)$ then:

1. $ev(\theta^{f^*}) \rightarrow^L Unif(0,1)$
2. If $g \neq f^*$ then $ev(\theta^g) \rightarrow^P 0$ in f^* probability

1. By 1. we are able to control Type I error at any desired level rejecting the null hypothesis rejecting when $ev(\theta^{f^*}) < \alpha$

Sketch of proof — Theorem 1

Let $ev(\theta^f)$ be the FBST for testing $H_0 : \theta = \theta^f$ and θ^f be the projection of a continuous density f on probabilities for $k(n)$ bins. If f^* is a Lipchitz continuous density on $[0, 1]$ and $k(n) = O\left(\frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}\right)$ then:

1. $ev(\theta^{f^*}) \to^L Unif(0, 1)$
2. If $g \neq f^*$ then $ev(\theta^g) \to^P 0$ in f^* probability

1. By 1. we are able to control Type I error at any desired level rejecting the null hypothesis rejecting when $ev(\theta^{f^*}) < \alpha$
2. By 2. we are able to detect differences between density g and any Lipchitz continuous f^*

Theorem 1

1. By the lemma, the distribution of θ is approximately Gaussian, so we are manipulating $f_{\theta|\mathbf{x}_n}(\theta)$ that is a transformation of a Gaussian vector.

Theorem 1

1. By the lemma, the distribution of θ is approximately Gaussian, so we are manipulating $f_{\theta|\mathbf{x}_n}(\theta)$ that is a transformation of a Gaussian vector.
2. Then, we apply a Taylor approximation on $f_{\theta|\mathbf{x}_n}$ around $\hat{\theta}$ and the tangent region becomes a comparison ellipsoids.

Theorem 1

1. By the lemma, the distribution of θ is approximately Gaussian, so we are manipulating $f_{\theta|\mathbf{x}_n}(\theta)$ that is a transformation of a Gaussian vector.
2. Then, we apply a Taylor approximation on $f_{\theta|\mathbf{x}_n}$ around $\hat{\theta}$ and the tangent region becomes a comparison ellipsoids.
3. As Gaussian ellipsoids are χ^2 distributed the results follows.

NP-FBST for composite hypothesis

Theorem 2

Let $\mathcal{F}_0$ be a set of densities f such that $\inf ||f^* - f|| > \delta$ for some $\delta > 0$, Θ_0 the projection of $\mathcal{F}_0$ on Θ_0 and $ev(\Theta_0)$ be the FBST for testing $H_0 : \theta \in \Theta_0$. If f^* is a Lipchitz continuous density on $[0, 1]$ and $k(n) = O\left(\frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}\right)$ then $ev(\Theta_0) \to^P 0$ in f^* probability

NP-FBST for composite hypothesis

Theorem 2

Let $\mathcal{F}_0$ be a set of densities f such that $\inf ||f^* - f|| > \delta$ for some $\delta > 0$, Θ_0 the projection of $\mathcal{F}_0$ on Θ_0 and $ev(\Theta_0)$ be the FBST for testing $H_0 : \theta \in \Theta_0$. If f^* is a Lipchitz continuous density on $[0, 1]$ and $k(n) = O\left(\frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}\right)$ then $ev(\Theta_0) \to^P 0$ in f^* probability

1. This theorem qualifies FBST as a test for goodness-of-fit to parametric families, median tests, mean tests etc. Is suffices to note that on all those examples the densites being tested are not arbitrary close fo f^*

NP-FBST for composite hypothesis

Theorem 2

Let $\mathcal{F}_0$ be a set of densities f such that $\inf ||f^* - f|| > \delta$ for some $\delta > 0$, Θ_0 the projection of $\mathcal{F}_0$ on Θ_0 and $ev(\Theta_0)$ be the FBST for testing $H_0 : \theta \in \Theta_0$. If f^* is a Lipchitz continuous density on $[0, 1]$ and $k(n) = O\left(\frac{n^{1/6}}{(\log n)^{1/6+\epsilon}}\right)$ then $ev(\Theta_0) \to^P 0$ in f^* probability

1. This theorem qualifies FBST as a test for goodness-of-fit to parametric families, median tests, mean tests etc. Is suffices to note that on all those examples the densites being tested are not arbitrary close fo f^*
2. This results does not show how to control Type I error, so pratical applications should apply e.g. bootstrap techniques

Uniformity tests — Power comparisons

Theorem 1 might be fully applied to test if some sample where sampled from a uniform density. We shall compare the power of our procedure to the power of other existing procedures such as Kolmogorov Smirnov, Anderson-Darling and its more powerful versions by [Zha02]

Uniformity tests — Power comparisons

Theorem 1 might be fully applied to test if some sample where sampled from a uniform density. We shall compare the power of our procedure to the power of other existing procedures such as Kolmogorov Smirnov, Anderson-Darling and its more powerful versions by [Zha02]

1. For small samples, our procedure is very competitive regarding statistical power

Uniformity tests — Power comparisons

Theorem 1 might be fully applied to test if some sample where sampled from a uniform density. We shall compare the power of our procedure to the power of other existing procedures such as Kolmogorov Smirnov, Anderson-Darling and its more powerful versions by [Zha02]

1. For small samples, our procedure is very competitive regarding statistical power
2. For larges sample sizes Zhang's alternatives usually perform better, but are not so flexible

Theorem 1 might be fully applied to test if some sample where sampled from a uniform density. We shall compare the power of our procedure to the power of other existing procedures such as Kolmogorov Smirnov, Anderson-Darling and its more powerful versions by [Zha02]

1. For small samples, our procedure is very competitive regarding statistical power
2. For larges sample sizes Zhang's alternatives usually perform better, but are not so flexible
3. Our procedure is never worse than Kolmogorov-Smirnov

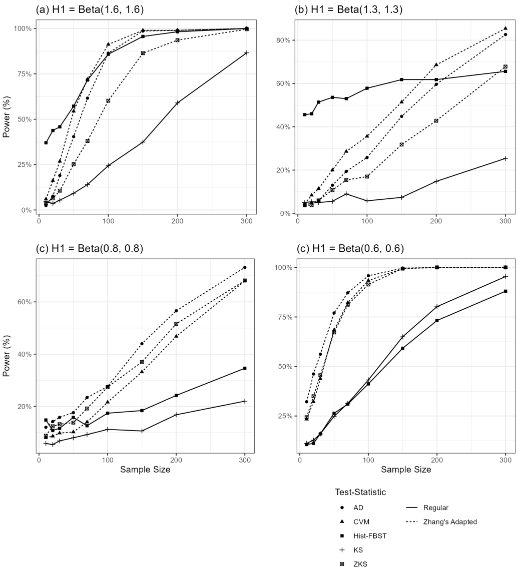
Theorem 1 might be fully applied to test if some sample where sampled from a uniform density. We shall compare the power of our procedure to the power of other existing procedures such as Kolmogorov Smirnov, Anderson-Darling and its more powerful versions by [Zha02]

1. For small samples, our procedure is very competitive regarding statistical power
2. For larges sample sizes Zhang's alternatives usually perform better, but are not so flexible
3. Our procedure is never worse than Kolmogorov-Smirnov
4. The Lipchitz continuity hypothesis seems very important and hurts the statistical power of the test

Table of Contents

1. Introduction — Parametric vs Nonparametric FBST
2. Bayesian Histograms
3. Entropy
4. Theorems
5. Simulations
6. Conclusion

Uniformity tests — Power comparison



Caveats

The results are useful in their current state but there are some important remarks to be made:

1. The bin size grows slowly, so the choice of ϵ is important
2. Boostrap methods for obtaining the null distribution might be more reliable than Theorem 1, as our results shows that the test is well suited for small sample sizes
3. The Lipchitz is restrictive. There are other Bernstein Von Mises Theorems [CN14] for bayesian histograms that allow for less smooth f^*, but then $k(n)$ would grow even slower
4. The null distribution for composite hypothesis remains to be studied

Table of Contents

1 Introduction — Parametric vs Nonparametric FBST

2 Bayesian Histograms

3 Entropy

4 Theorems

5 Simulations

6 Conclusion

Conclusion

1. When adapting the FBST for nonparametric settings the cross-entropy funcional its a natural choice of surprise functions. This is the case for histograms but could also be used for e.g. kernel density estimates.

Conclusion

1. When adapting the FBST for nonparametric settings the cross-entropy funcional its a natural choice of surprise functions. This is the case for histograms but could also be used for e.g. kernel density estimates.
2. The same approach applied to the Dirichlet-Multinomial model could be applied to other exponential family parametric distributions, as the Bernstein Von Mises results already exists.

Conclusion

1. When adapting the FBST for nonparametric settings the cross-entropy funcional its a natural choice of surprise functions. This is the case for histograms but could also be used for e.g. kernel density estimates.
2. The same approach applied to the Dirichlet-Multinomial model could be applied to other exponential family parametric distributions, as the Bernstein Von Mises results already exists.
3. For practical applications more refined algorithms for building $k(n)$ should be developed.

Bibliography I

Carlos A. de B. Pereira, Julio Michael Stern, and Sergio Wechsler. "Can a significance test be genuinely Bayesian?" In: *Bayesian Analysis* 3.1 (2008), pp. 79–100. DOI: 10.1214/08-BA303. URL: https://doi.org/10.1214/08-BA303.

Ismaël Castillo and Richard Nickl. "On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures". In: *The Annals of Statistics* 42.5 (2014), pp. 1941–1969. DOI: 10.1214/14-AOS1246. URL: https://doi.org/10.1214/14-AOS1246.

Bibliography II

Subhashis Ghosal. “Asymptotic Normality of Posterior Distributions for Exponential Families when the Number of Parameters Tends to Infinity”. In: *Journal of Multivariate Analysis* 74.1 (2000), pp. 49–68. URL: https://EconPapers.repec.org/RePEc:eee:jmvana:v:74:y:2000:i:1:p:49-68.

Jin Zhang. “Powerful Goodness-of-Fit Tests Based on the Likelihood Ratio”. In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64.2 (2002), pp. 281–294. ISSN: 13697412, 14679868. URL: http://www.jstor.org/stable/3088800 (visited on 05/16/2024).