Afrika Statistika

Afrika Statistika
Vol. 15 (2), 2020, pages 2349 - 2369.
DOI: http://dx.doi.org/10.16929/as/2020.2349.162

ISSN 2316-090X

Nonparametric ϕ-Divergence Estimation and Test for Model Selection

Freedath Djibril Moussa [1], Jean de Dieu Nkurunziza [2] and Papa Ngom [2,3]

[1]Faculté des Sciences et Techniques, Abomey-Calavi, Bénin
[2]Institut de Mathématiques et de Sciences Physiques
[3] LMA, Université Cheikh Anta Diop, Dakar, Sénégal

Received on June 25, 2020. Accepted on October 08, 2020.

Copyright © 2020, Afrika Statistika and The Statistics and Probability African Society (SPAS). All rights reserved

Abstract. In this paper, we study a bias reduced kernel density estimator and derive a nonparametric ϕ-divergence estimator based on this density estimator. We investigate the asymptotic properties of these two estimators and we formulate an asymptotically standard normal test for model selection.

Résumé. (Abstract in French). Dans cet article, nous étudions l'estimateur de densité à noyau avec un biais réduit et nous dérivons un estimateur nonparamétrique de la ϕ-divergence basé sur cet estimateur de densité. Nous investiguons les propriétés asymptotiques de ces deux estimateurs et nous formulons un test asymptotiquement normal standard pour la sélection de modèle.

Key words: nonparametric estimation; ϕ-divergence; strong consistency; asymptotic normality; hypothesis testing; model Selection.
Mathematics Subject Classification Objects : 62F10; 62F12; 62G07; 62G10; 62G20.

---

* Corresponding Author : Jean de Dieu Nkurunziza (jean.d.d.nkurunziza@aims-senegal.org)
Freedath Djibril Moussa : freedath.djibrilmoussa@gmail.com
Papa Ngom : papa.ngom@ucad.edu.sn

The authors.

Freedath Djibril Moussa, Ph.D., is assistant professor at the Department of Mathematics at the Faculty of Sciences and Technics, University of Abomey-Calavi.

Jean de Dieu Nkurunziza, Ph.D., has completed his Ph.D. at the Institute of Mathematics and Physical Sciences (IMSP), University of Abomey-Calavi.

Papa Ngom, Ph.D., is full professor at the Department of Mathematics and Computer Sciences, Université Cheikh Anta Diop, Dakar, Séénégal.

1. Introduction

Let $X_1, ..., X_n$ be independent and identically distributed (*iid*) random variables and assume that the common distribution function of these variables has an unknown density f. On can use non-parametric approach for the estimation of f. A widely used non-parametric estimator is the kernel density estimator given by Rosenblatt (1956) and Parzen (1962). The asymptotic properties of this estimator have been intensively investigated and many kernel-types estimators have been proposed. Dony and Einmahl (2006) showed the uniform consistency of kernel density estimator with general bandwidth sequences. Bouzebda and Elhattab (2011) established the uniform in bandwidth consistency of kernel-type estimators of Shannon Entropy. Einmahl and Mason (2005) proved the uniform in bandwidth consistency of kernel-type function estimators. Dhaker *et al.* (2016) proposed a strong uniformly consistent kernel-type estimator of divergence measures. Rudemo (1982) and Bowman *et al.* (1984) introduced a convenient method for the choice of optimal bandwidth in practice for kernel density estimator using cross-validation. Xiaoran and Jingjing (2014) focused on improving the convergence rate of the kernel density estimator by formulating a bias reduced kernel density estimator. In this paper, we investigate the properties of this estimator and prove its asymptotic normality and its strong consistency. Next, we use the bias reduced kernel density estimator to derive a nonparametric estimator of the ϕ-divergence proposed by Csiszár (1963) and developed by Ali and Silvey (1966). This general family of divergences which include the Kullback and Leibler (1951) divergence (KLD) and the Hellinger (1977) divergence , measures the dissimilarity between those two probability distributions and is a key tool for model selection. For more details on divergence measures, see Pardo (2006) and Basu *et al.* (1998). In Ba *et al.* (2017) , a general normal asymptotic theory for divergence measures estimators has been provided. Ba *et al.* (2018) also extended the aforementioned results to symmetrized forms of divergence measures and investigated in details the Tsallis and Renyi divergence measures as well as the Kullback-Leibler measures. Last the proofs of some results and the applicability to usual distribution functions have been addressed in Ba *et al.* (2019) .

We show the asymptotic normality and the strong consistency of this estimator and construct an asymptotically normal test for model selection using ϕ-divergence type statistics.

The rest of the paper is organized as follows. We study the bias reduced kernel density estimator and its asymptotic properties in Section 2 and we derive the nonparametric ϕ-divergence estimator and its asymptotic properties in Section 3. In Section 4, the test for model selection are proposed and the computational results are presented in Section 5. Finally the conclusion appears in Section 6.

2. Some important results on bias reduced kernel density estimator

Let $X_1, ..., X_n$ be iid random variables of unknown density, which we shall denote by f. Consider a probability density function K defined on $\mathbb{R}$ (the kernel) and a positive parameter h_n, the bandwidth. Assuming that the random variable of density K is centered with finite variance μ_2, the kernel density estimator proposed by Rosenblatt (1956) and Parzen (1962) of f is given by

$$\hat{f}_{n,h_n}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right). \tag{1}$$

The optimal performance of the kernel density estimator has been widely studied. For more details see Samiuddin and El-Sayyad (1990) and Samiuddin and El-Sayyad (1992). Xiaoran and Jingjing (2014) proposed an intuitive and feasible kernel density estimator which reduces the bias and the mean squared error (MSE) significantly compared to the ordinary kernel density estimator. It is defined by

$$\begin{aligned}\hat{f}^b_{n,h_n}(x) &= \hat{f}_{n,h_n}(x) - \frac{h_n^2}{2}\mu_2 \hat{f}^{(2)}_{n,h_n}(x)\\ &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) - \frac{\mu_2}{2nh_n}\sum_{i=1}^{n} K^{(2)}\left(\frac{x - X_i}{h_n}\right),\end{aligned}$$

where $\mu_2 := \int u^2 K(u)du$. Under the following conditions on f , K and h_n:

- $\int uK(u)du = 0$;
- f is fourth differentiable in a neighborhood of x;
- $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$;

they evaluated the bias, the variance of this estimator and came up with a convergence rate $O\left(n^{-6/7}\right)$. But the conditions used are not sufficient to get their results.

For example the passage $\frac{1}{h_n^2}\int K^{(2)}(u)f(x-uh_n)du = \int K(u)f^{(2)}(x-uh_n)du$ is not verified. The Epanechnikov kernel defined by $K(u) = \dfrac{3}{4}(1-u^2)1_{|u|\leq 1}$ is a counter example. Hence we add some regularity conditions on the kernel K to obtain the convergence rate $O\left(n^{-6/7}\right)$ for $\hat{f}^b_{n,h_n}$. Therefore the proofs of following propositions follow the proofs proposed by Xiaoran and Jingjing (2014) in Theorem 1.

Proposition 1. *Suppose that f is four times differentiable in a neighborhood of x and let K be the density of a centered random variable with finite second and third order moment denoted by μ_2 and μ_3 respectively, satisfying the following assumption*

$$A.1 : K(x) = K_1(x)1_A(x),\ A \subseteq \mathbb{R} \text{ such that } \lim_{\substack{x\to\inf A\\ x>\inf A}} K^{(i)}(x) = \lim_{\substack{x\to\sup A\\ x<\sup A}} K^{(i)}(x) = 0,\ \forall i = 0,1.$$

Then for all $n \in \mathbb{N}^*$ and $h_n > 0$ we have

$$Bias\left(\hat{f}^b_{n,h_n}(x)\right) = -\frac{h_n^3}{6}\mu_3 f^{(3)}(x) + O(h_n^4), \tag{2}$$

and

$$Bias\left(\hat{f}^b_{n,h_n}(x)\right) \longrightarrow 0 \text{ as } n\to\infty \text{ if } lim_{n\to\infty}h_n = 0.$$

Proposition 2. *Assume that f is four times differentiable in a neighborhood of x. Let K be the density of a centered random variable with finite second and third order moment denoted by μ_2 and μ_3 respectively satisfying the assumption $A.1$. If in addition, the condition*

$$A.2 : \int K^2(u)du < \infty;\ \int (K^{(2)}(u))^2 du < \infty;\ \int u(K^{(2)}(u))^2 du = 0$$

is satisfied. Then for all $n \in \mathbb{N}^*$ and $h_n > 0$ we have

$$Var\left(\hat{f}^b_{n,h_n}(x)\right) \leq \frac{1}{2nh_n}\mu_2^2 f(x)\int (K^{(2)}(u))^2 du + O((n)^{-1}). \tag{3}$$

Consequently the optimal MSE (Mean Squared Error) is of order $n^{-6/7}$. If K is a symmetric kernel, $\mu_3 = 0$; hence the optimal MSE for the bias reduced kernel density estimator is of order $n^{-8/9}$.

Write now the bias reduced kernel density estimator of f as follows

$$\begin{aligned}\hat{f}^b_{n,h_n}(x) &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) - \frac{\mu_2}{2nh_n}\sum_{i=1}^{n} K^{(2)}\left(\frac{x-X_i}{h_n}\right)\\ &= \frac{1}{nh_n}\sum_{i=1}^{n} \varphi\left(\frac{x-X_i}{h_n}\right)\end{aligned}$$

with $\varphi = K - \frac{\mu_2}{2}K^{(2)}$. The following result gives the asymptotic distribution of the bias reduced kernel density estimator.

Theorem 1. *Suppose that conditions on f and K in Proposition 1 hold and assume further that:*

1. *$f(x) > 0$ and in the neighborhood of x, $f^{(i)}$, $i = 0, 1, 2$ are bounded;*
2. *the function φ satisfies: $\int \varphi(u)du < \infty$, $\int \varphi^2(u)du < \infty$ and $\int |\varphi(u)|^3 du < +\infty$;*
3. *$nh_n \to +\infty$ and $\sqrt{n}h_n^{7/2} \to 0$ as $n \to \infty$.*

Then

$$\sqrt{nh_n}\left(\hat{f}^b_{n,h_n}(x) - f(x)\right) \longrightarrow \mathcal{N}\left(0, f(x)\int \varphi^2(u)du\right). \tag{4}$$

To prove this theorem, we need the Lyapunov central limit theorem (Mbuba *et al.* (1984)) and the dominated convergence theorem (Arzelà (1885)).

Now, we shall obtain the uniform in bandwidth consistency of $\hat{f}^b_{n,h_n}$ by considering the following conditions.

(H1) K is a density of a centered random variable with finite variance μ_2.

(H2) $\int \varphi(x)dx := \zeta < \infty$, $\|\varphi\|_\infty = \sup_{x\in\mathbb{R}} |\varphi(x)| := \gamma < +\infty$ and $\|\varphi\|_2 := \left(\int \varphi^2(u)du\right)^{1/2} < +\infty$.

(H3) Consider the class of functions:

$$\Phi = \{t \mapsto \varphi((x-t)/h_n) : h_n > 0, \quad x \in \mathbb{R}\}.$$

For $\varepsilon > 0$, let $N(\varepsilon, \Phi) = \sup_Q N(\gamma\varepsilon, \Phi, d_{2Q})$ where the supremum is taken over all probability measures Q on $(\mathbb{R}, \mathcal{B}), d_{2Q}$ is the $L_2(Q)$-metric and $N(\gamma\varepsilon, \Phi, d_{2Q})$ is the minimal number of balls of radius $\gamma\varepsilon$ needed to cover Φ.
For some $C > 0$ and $\nu > 0$, $N(\varepsilon, \Phi) \leq C\varepsilon^{-\nu}$, $0 < \varepsilon < 1$.
(H4) Φ is a pointwise measurable class, that is, there exists a countable subclass

Φ_0 of Φ such that we can find for any function $\phi \in \Phi$ a sequence of functions ϕ_m in Φ_0 for which $\phi_m(y) \to \phi(y), \quad y \in \mathbb{R}$.

(H5) f is four-times differentiable in neighbourhood of x.

Remark 1. The condition (H3) holds whenever φ is a function of bounded variation (Nolan and Pollard (1987)) and the condition (H4) is satisfied whenever φ is right continuous as shown by Van der Vaart and Wellner (1996). For instance the Gaussian, biweight and triweight kernels satisfy the conditions (H1)-(H4) above.

Theorem 2. *Assuming (H1-H5) are satisfied. For each pair of sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$; and choosing a suitable bandwidth sequence $h_n \to 0$ and $0 < a_n < b_n \leq 1$ we have with probability 1,*

$$\limsup_{n\to\infty} \sup_{a_n \leq h_n \leq b_n} \frac{\sqrt{nh_n}\left\|\hat{f}^b_{n,h_n} - \mathbb{E}\hat{f}^b_{n,h_n}\right\|_\infty}{\sqrt{\log\left(1/h_n\right) \vee \log\log n}} =: \omega < \infty. \tag{5}$$

The proof of this theorem follows along the lines of the proof of theorem 1 obtained by Einmahl and Mason (2005).

Remark 2. Theorem 2 implies for any sequences $0 < a_n < b_n \leq 1$, satisfying $na_n/\log(n) \to \infty$, that

$$\sup_{a_n \leq h_n \leq b_n} \left\|\hat{f}^b_{n,h_n} - \mathbb{E}\hat{f}^b_{n,h_n}\right\|_\infty = O\left(\sqrt{\frac{\log(1/a_n) \vee \log\log n}{na_n}}\right); \tag{6}$$

with probability 1. This in turn implies

$$\lim_{n\to\infty} \sup_{a_n \leq h_n \leq b_n} \left\|\hat{f}^b_{n,h_n} - \mathbb{E}\hat{f}^b_{n,h_n}\right\|_\infty = 0 \quad a.s.$$

Theorem 3. *Let f be Lipschitz function on $\mathbb{R}$. Assume that the conditions (H.1) and (H.5) are satisfied; and the derivatives of order j of f are bounded, $\forall j = 2, 3, 4$. For each pair of sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$; choosing a suitable bandwidth sequence $h_n \to 0$ and for $0 < a_n < b_n \leq 1$ together with $b_n \to 0$ as $n \to \infty$, we have with probability 1*

$$\sup_{a_n \leq h_n \leq b_n} \left\|\mathbb{E}\hat{f}^b_{n,h_n} - f\right\|_\infty = O\left(b_n\right).$$

The proof of theorem 3 is a combination of lemma 1 (Einmahl and Mason (2005)) and of proposition 1.5 (Tsybakov (2008)).

3. Nonparametric ϕ-divergence estimator

Let $X_1, ..., X_n$ be a random sample of unknown density function f defined on $\mathbb{R}$ and let g be a candidate model. Here we consider g as a strictly positive density. The discrepancy between f and g can be measured by a ϕ-divergence proposed by Csiszár (1963) and also independently defined by Ali and Silvey (1966). Let $\phi(.)$ be a convex function over $(0, \infty)$ such that $\phi(1) = 0$. The ϕ-divergence is associated to $\phi(.)$ is:

$$\mathcal{D}_\phi(f, g) = \int \phi\left(\frac{f(x)}{g(x)}\right) g(x)dx. \tag{7}$$

Many common divergences are special cases of ϕ-divergence like the Kullback-Leibler divergence (KLD) for $\phi(t) = t\ln(t)$, the Hellinger divergence for $\phi(t) = (\sqrt{t}-1)^2$, $t^2 - 1$ for $\mathcal{X}^2-$ divergence and $\frac{t^\alpha-\alpha t+(\alpha-1)}{\alpha(\alpha-1)}$ for α-divergence.

Since f is unknown, $\mathcal{D}_\phi(f, g)$ has to be estimated. Bouzebda and Elhattab (2011) proved the uniform in bandwidth consistency of kernel-type estimator of Shannon's entropy.

This work was based on Rosenblatt kernel density estimator. In this paper, we propose a nonparametric estimator of $\mathcal{D}_\phi(f, g)$ based on bias reduced kernel density estimator and defined by

$$\widehat{\mathcal{D}}_\phi(f, g) = \int_{A_n} \phi\left(\frac{\hat{f}^b_{n,h_n}(x)}{g(x)}\right) g(x)dx, \tag{8}$$

where $\hat{f}^b_{n,h_n}$ is the bias reduced kernel density estimator and $A_n = \{x \in \mathbb{R}; \hat{f}^b_{n,h_n}(x) \geq \varepsilon_n\}$ with ε_n a sequence of positive constants such that $\varepsilon_n \to 0$ as $n \to \infty$. Similarly to lemma 1 (Dhaker *et al.* (2017)), we obtain the asymptotic normality of this estimator.

Theorem 4. *Let $\mathcal{D}_\phi(f, g)$ be the ϕ-divergence between f and g; and $\widehat{\mathcal{D}}_\phi(f, g)$ its estimator. Suppose that $\phi \in C^1([0, \infty))$ and there exists a measurable and integrable function G such that forall $x \in A_n$, $\left|\phi'\left(\frac{f(x)}{g(x)}\right)\right| < G(x)$. If in addition the conditions of theorem 1 are satisfied, then*

$$\sqrt{nh_n}\left(\widehat{\mathcal{D}}_\phi(f, g) - \mathcal{D}_\phi(f, g)\right) \longrightarrow \mathcal{N}\left(0, \left(\int_{A_n} \sigma_*(x)\phi'\left(\frac{f(x)}{g(x)}\right)dx\right)^2\right),$$

with $\sigma_*^2(x) = f(x)\int \varphi^2(u)du$.

We further establish the strong uniform in bandwidth consistency of nonparametric ϕ-divergence estimator.

Theorem 5. *Let f be Lipschitz function on $\mathbb{R}$ and let ϕ be a convex function over $(0,\infty)$ and satisfying $\phi(1)=0$. Assume that the conditions (H.1-H.5) are satisfied; and the derivatives of order j of f are bounded, $\forall j=2,3,4$. For each pair of sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$; choosing a suitable bandwidth sequence $h_n\to 0$ and $0<a_n<b_n\leq 1$ together with $b_n\to 0$ and $na_n/\log(n)\to\infty$ as $n\to\infty$, we have with probability 1;*

$$\sup_{a_n\leq h_n\leq b_n}\left|\widehat{\mathcal{D}}_\phi(f,g)-\mathcal{D}_\phi(f,g)\right| = O\left(\sqrt{\frac{\log(1/a_n)\vee\log\log n}{na_n}}\vee b_n\right).$$

The proof of this theorem is a combination of two following lemmas. Define $\hat{\mathbb{E}}\widehat{\mathcal{D}}_\phi(f,g)$ by

$$\hat{\mathbb{E}}\widehat{\mathcal{D}}_\phi(f,g) := \int_{A_n}\phi\left(\frac{\mathbb{E}\hat{f}^b_{n,h_n}(x)}{g(x)}\right)g(x)dx.$$

Lemma 1. *Let ϕ be a convex function over $(0,\infty)$ and satisfying $\phi(1)=0$. Suppose that the conditions (H.1-H.5) hold. For each pair of sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$; choosing a suitable bandwidth sequence $h_n\to 0$ and $0<a_n<b_n\leq 1$ together with $na_n/\log(n)\to\infty$ as $n\to\infty$, we have with probability 1;*

$$\sup_{a_n\leq h_n\leq b_n}\left|\widehat{\mathcal{D}}_\phi(f,g)-\hat{\mathbb{E}}\widehat{\mathcal{D}}_\phi(f,g)\right| = O\left(\sqrt{\frac{\log(1/a_n)\vee\log\log n}{na_n}}\right).$$

Proof. Set $\Delta_{n1} := \widehat{\mathcal{D}}_\phi(f,g)-\hat{\mathbb{E}}\widehat{\mathcal{D}}_\phi(f,g)$. We have

$$\begin{aligned}|\Delta_{n1}| &= \left|\int_{A_n}\left[\phi\left(\frac{\hat{f}^b_{n,h_n}(x)}{g(x)}\right)-\phi\left(\frac{\mathbb{E}\hat{f}^b_{n,h_n}(x)}{g(x)}\right)\right]g(x)dx\right|\\ &\leq \int_{A_n}\left|\phi\left(\frac{\hat{f}^b_{n,h_n}(x)}{g(x)}\right)-\phi\left(\frac{\mathbb{E}\hat{f}^b_{n,h_n}(x)}{g(x)}\right)\right|g(x)dx.\end{aligned}$$

Since ϕ is a convex function, it is locally Lipschitz; so there exists a constant $k>0$ such that for n large enough (by Proposition 1), we have

$$\begin{aligned}|\Delta_{n1}| &\leq k\int_{A_n}\left|\frac{\hat{f}^b_{n,h_n}(x)}{g(x)}-\frac{\mathbb{E}\hat{f}^b_{n,h_n}(x)}{g(x)}\right|g(x)dx\\ &\leq k\int_{A_n}\left|\hat{f}^b_{n,h_n}(x)-\mathbb{E}\hat{f}^b_{n,h_n}(x)\right|dx.\end{aligned}$$

Since

$$\int_{A_n} dx \leq \int_{A_n} \frac{\hat{f}^b_{n,h_n}(x)}{\varepsilon_n} dx \leq \frac{1}{\varepsilon_n} \int_{\mathbb{R}} \hat{f}^b_{n,h_n}(x) dx \leq \frac{\zeta}{\varepsilon_n}, \tag{9}$$

$$\begin{aligned} |\Delta_{n1}| &\leq \frac{\zeta k}{\varepsilon_n} \sup_{x \in A_n} \left| \hat{f}^b_{n,h_n}(x) - \mathbb{E}\hat{f}^b_{n,h_n}(x) \right| \\ &\leq \frac{\zeta k}{\varepsilon_n} \sup_{x \in \mathbb{R}} \left| \hat{f}^b_{n,h_n}(x) - \mathbb{E}\hat{f}^b_{n,h_n}(x) \right|. \end{aligned}$$

Hence by Remark 2,

$$\sup_{a_n \leq h_n \leq b_n} |\Delta_{n1}| = O\left(\sqrt{\frac{\log(1/a_n) \vee \log\log n}{n a_n}} \right). \blacksquare$$

Lemma 2. *Let f be Lipschitz function on $\mathbb{R}$ and let ϕ be a convex function over $(0, \infty)$ and satisfying $\phi(1) = 0$. Assume that the conditions (H.1) and (H.5) are satisfied; and the derivatives of order j of f are bounded, $\forall j = 2, 3, 4$. For each pair of sequences $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$; choosing a suitable bandwidth sequence $h_n \to 0$ and $0 < a_n < b_n \leq 1$ together with $b_n \to 0$ as $n \to \infty$, we have with probability 1;*

$$\sup_{a_n \leq h_n \leq b_n} \left| \hat{\mathbb{E}}\widehat{\mathcal{D}}_\phi(f,g) - \mathcal{D}_\phi(f,g) \right| = O(b_n).$$

Proof. Set $\Delta_{n2} := \hat{\mathbb{E}}\widehat{\mathcal{D}}_\phi(f,g) - \mathcal{D}_\phi(f,g)$. We have

$$\begin{aligned} |\Delta_{n2}| &= \left| \int_{A_n} \left[\phi\left(\frac{\mathbb{E}\hat{f}^b_{n,h_n}(x)}{g(x)} \right) - \phi\left(\frac{f(x)}{g(x)} \right) \right] g(x) dx \right| \\ &\leq \int_{A_n} \left| \phi\left(\frac{\mathbb{E}\hat{f}^b_{n,h_n}(x)}{g(x)} \right) - \phi\left(\frac{f(x)}{g(x)} \right) \right| g(x) dx. \end{aligned}$$

By the convexity of ϕ, for n large enough, there exists $\alpha > 0$ such that:

$$\begin{aligned} \left| \hat{\mathbb{E}}\widehat{\mathcal{D}}_\phi(f,g) - \mathcal{D}_\phi(f,g) \right| &\leq \alpha \int_{A_n} \left| \frac{\mathbb{E}\hat{f}^b_{n,h_n}(x)}{g(x)} - \frac{f(x)}{g(x)} \right| g(x) dx \\ &\leq \alpha \int_{A_n} \left| \mathbb{E}\hat{f}^b_{n,h_n}(x) - f(x) \right| dx. \end{aligned}$$

Using (9), we get

$$\begin{aligned}|\Delta_{n2}| &\leq \frac{\zeta\alpha}{\varepsilon_n}\sup_{x\in A_n}\left|\mathbb{E}\hat{f}^b_{n,h_n}(x)-f(x)\right| \\ &\leq \frac{\zeta\alpha}{\varepsilon_n}\left\|\mathbb{E}\hat{f}^b_{n,h_n}(x)-f(x)\right\|_\infty .\end{aligned}$$

By Theorem 3,

$$\sup_{a_n\leq h_n\leq b_n}|\Delta_{n2}| = O(b_n). \blacksquare$$

4. Test for model selection

Having two candidate models g_1 and g_2, we propose to choose the model which is close to the true probability density f using a ϕ-divergence. Consider the following model selection test
$H_0 : \mathcal{D}_\phi(f, g_1) = \mathcal{D}_\phi(f, g_2)$ means that the two models g_1 and g_2 are equivalent,
$H_1 : \mathcal{D}_\phi(f, g_1) \neq \mathcal{D}_\phi(f, g_2)$ means that g_1 is not equivalent to g_2.
The statistic of this test is:

$$\Lambda_n := \frac{\sqrt{nh_n}}{\hat{\xi}}\left[\widehat{\mathcal{D}}_\phi(f,g_1)-\widehat{\mathcal{D}}_\phi(f,g_2)\right] \tag{10}$$

where $\hat{\xi}$ is an estimator of $\xi = \int_{A_n}\sigma_*(x)\left[\phi'\left(\frac{f(x)}{g_1(x)}\right)-\phi'\left(\frac{f(x)}{g_2(x)}\right)\right]dx$ obtained by replacing f by $\hat{f}^b_{n,h_n}$. Next, we give the asymptotic distribution of Λ_n.

Theorem 6. *(Asymptotic distribution of the Λ_n-statistic).*
Suppose that the conditions of theorems 1 and 4 are satisfied. Under the null hypothesis H_0, $\Lambda_n \longrightarrow \mathcal{N}(0,1)$.

Proof. Under the theorem 4, we have

$$\begin{aligned}\widehat{\mathcal{D}}_\phi(f,g_1) &= \mathcal{D}_\phi(f,g_1)+\int_{A_n}\left(\frac{\hat{f}^b_{n,h_n}(x)}{g_1(x)}-\frac{f(x)}{g_1(x)}\right)\phi'\left(\frac{f(x)}{g_1(x)}\right)g_1(x)dx + \\ &+\int_{A_n} o\left(\left\|\frac{\hat{f}^b_{n,h_n}}{g_1}-\frac{f}{g_1}\right\|\right)g_1(x)dx,\end{aligned}$$

$$\begin{aligned}\widehat{\mathcal{D}}_\phi(f,g_2) &= \mathcal{D}_\phi(f,g_2)+\int_{A_n}\left(\frac{\hat{f}^b_{n,h_n}(x)}{g_2(x)}-\frac{f(x)}{g_2(x)}\right)\phi'\left(\frac{f(x)}{g_2(x)}\right)g_2(x)dx + \\ &+\int_{A_n} o\left(\left\|\frac{\hat{f}^b_{n,h_n}}{g_2}-\frac{f}{g_2}\right\|\right)g_2(x)dx.\end{aligned}$$

Then

$$\widehat{\mathcal{D}}_\phi(f,g_1)-\widehat{\mathcal{D}}_\phi(f,g_2) = \mathcal{D}_\phi(f,g_1)-\mathcal{D}_\phi(f,g_2)+\int_{A_n}\left(\frac{\hat{f}^b_{n,h_n}(x)}{g_1(x)}-\frac{f(x)}{g_1(x)}\right)\phi'\left(\frac{f(x)}{g_1(x)}\right)g_1(x)dx + \\ -\int_{A_n}\left(\frac{\hat{f}^b_{n,h_n}(x)}{g_2(x)}-\frac{f(x)}{g_2(x)}\right)\phi'\left(\frac{f(x)}{g_2(x)}\right)g_2(x)dx.$$

Under $H_0,\ \mathcal{D}_\phi(f,g_1)=\mathcal{D}_\phi(f,g_2)$, we therefore have

$$\widehat{\mathcal{D}}_\phi(f,g_1)-\widehat{\mathcal{D}}_\phi(f,g_2) = \int_{A_n}\left(\frac{\hat{f}^b_{n,h_n}(x)}{g_1(x)}-\frac{f(x)}{g_1(x)}\right)\phi'\left(\frac{f(x)}{g_1(x)}\right)g_1(x)dx + \\ -\int_{A_n}\left(\frac{\hat{f}^b_{n,h_n}(x)}{g_2(x)}-\frac{f(x)}{g_2(x)}\right)\phi'\left(\frac{f(x)}{g_2(x)}\right)g_2(x)dx, \tag{11}$$

and

$$\sqrt{nh_n}\left[\widehat{\mathcal{D}}_\phi(f,g_1)-\widehat{\mathcal{D}}_\phi(f,g_2)\right]=\int_{A_n}\sqrt{nh_n}\left(\hat{f}^b_{n,h_n}(x)-f(x)\right)\left[\phi'\left(\frac{f(x)}{g_1(x)}\right)-\phi'\left(\frac{f(x)}{g_2(x)}\right)\right]dx \tag{12}$$

Finally, applying the theorem 1 in the relation (12), we immediately obtain $\Lambda_n \longrightarrow \mathcal{N}(0,1)$. ■

5. Computational results

5.1. Performance of ϕ-divergence estimator

In this subsection, we investigate the behaviour of the ϕ-divergence estimator by considering $\phi(t) = t\ln(t)$ (Kullback-Leibler divergence) and $\phi(t) = \left(\sqrt{t}-1\right)^2$ (Hellinger divergence). The reference model used is the mixture of gamma distribution $Gamma(4.02, 0.05)$ of density $f_{GA}(x) = \frac{0.05^{(4.02)}}{\Gamma(4.02)}x^{(4.02-1)}e^{-0.05x}1_{x\geq 0}$ and log-normal distribution $Log-normal(4.15, 0.52)$ of density $f_{LN}(x) = \frac{1}{0.52\sqrt{2\pi x}}e^{-\frac{1}{2(0.52)^2}(\ln(x)-4.15)^2}1_{x\geq 0}$. Therefore the reference distribution of density $f_{Mix}(x) = \pi f_{GA}(x)+(1-\pi)f_{LN}(x)$ is defined by

$$Mix(x) := \pi\textit{Gamma}(4.02, 0.05) + (1-\pi)\textit{Log-normal}(4.15, 0.52) \tag{13}$$

where $\pi \in (0,1)$ is specific to each set of experiments. We choose two values of π which are $\pi = 0.5$ and $\pi = 0.75$. Note that the value $\pi = 0.5$ is the value for which the Gamma and Log-normal distributions are approximately at equal distance to the reference distribution and for $\pi = 0.75$, the reference distribution is gamma but slightly contaminated by the log-normal distribution. Further the log-normal distribution as the candidate model is considered $Log-normal(4.15, 0.52)$. The Kullback-Leibler divergence and the Hellinger divergence are respectively

$$D_{KL} \equiv \mathcal{D}_{KL}(f_{Mix}, f_{LN}) = \int f_{Mix}(x) \ln\left(\frac{f_{Mix}(x)}{f_{LN}(x)}\right) dx$$
$$D_{H} \equiv \mathcal{D}_{H}(f_{Mix}, f_{LN}) = \int \left(\sqrt{f_{Mix}(x)} - \sqrt{f_{LN}(x)}\right)^2 dx.$$

Dhaker *et al.* (2016) showed that the Kullback-Leibler and Hellinger divergences estimators based on the kernel density estimator are strongly consistent estimators. We consider the Kullback-Leibler and the Hellinger divergences based on the bias reduced kernel density estimator ($\widehat{D}_{KL1}$ and $\widehat{D}_{H1}$) and we further consider the case of Kullback-Leibler and Hellinger divergences based on the kernel density estimator ($\widehat{D}_{KL2}$ and $\widehat{D}_{H2}$).

Therefore, the corresponding estimators are

$$\widehat{D}_{KL1} \equiv \widehat{\mathcal{D}}_{KL1}(f, f_{LN}) = \int_{A_n} \hat{f}^b_{n,h_n}(x) \ln\left(\frac{\hat{f}^b_{n,h_n}(x)}{f_{LN}(x)}\right) dx \tag{14}$$
$$\widehat{D}_{H1} \equiv \widehat{\mathcal{D}}_{H1}(f, f_{LN}) = \int_{A_n} \left(\sqrt{\hat{f}^b_{n,h_n}(x)} - \sqrt{f_{LN}(x)}\right)^2 dx \tag{15}$$

and

$$\widehat{D}_{KL2} \equiv \widehat{\mathcal{D}}_{KL2}(f, f_{LN}) = \int_{A_n} \hat{f}_{n,h_n}(x) \ln\left(\frac{\hat{f}_{n,h_n}(x)}{f_{LN}(x)}\right) dx \tag{16}$$
$$\widehat{D}_{H2} \equiv \widehat{\mathcal{D}}_{H2}(f, f_{LN}) = \int_{A_n} \left(\sqrt{\hat{f}_{n,h_n}(x)} - \sqrt{f_{LN}(x)}\right)^2 dx. \tag{17}$$

We generate 1000 samples of various sizes (20 to 2000) from the mixture defined by (13) and compute the estimates $\widehat{D}_{KL1}, \widehat{D}_{KL2}, \widehat{D}_{H1}$ and $\widehat{D}_{H2}$ using the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ since it has infinitely many (nonzero) derivatives. The optimal bandwidth h_n for the bias reduced kernel density estimator is obtained by cross-validation method as proposed by Rudemo (1982) and Bowman *et al.* (1984).

The numerical results (average values) and graphs are given below. The values in parenthesis are standard errors.

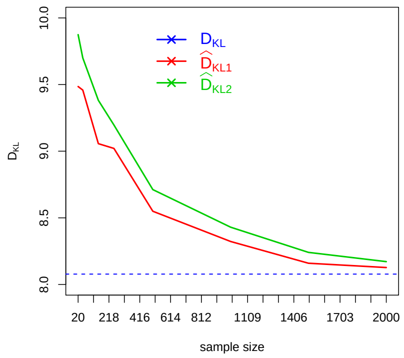
Kullback-Leibler estimates for various sample size and π.

For $\pi = 0.5$, $D_{KL} = 8.0778$.

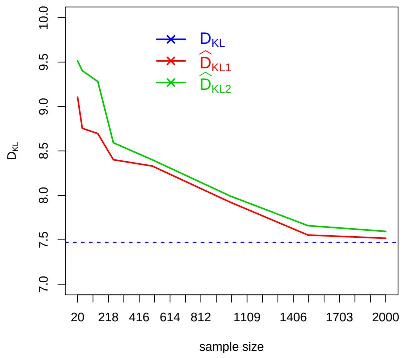
$n \rightarrow$	20	150	250	500	1000	1500	2000
$\widehat{D}_{KL1} \rightarrow$	9.4851 (1.7853)	9.0564 (0.4980)	9.0208 (0.3831)	8.2494 (0.1738)	8.1225 (0.1156)	8.1094 (0.0905)	8.0971 (0.0800)
$\widehat{D}_{KL2} \rightarrow$	9.8744 (1.7888)	9.3816 (1.1161)	9.1965 (0.8846)	8.4118 (0.3292)	8.3298 (0.2079)	8.2408 (0.1568)	8.1514 (0.1310)

For $\pi = 0.75$, $D_{KL} = 7.4725$.

$n \rightarrow$	20	150	250	500	1000	1500	2000
$\widehat{D}_{KL1} \rightarrow$	9.1055 (1.7480)	8.6951 (0.5194)	8.4011 (0.3099)	8.3301 (0.2199)	7.9239 (0.1096)	7.8536 (0.0876)	7.5175 (0.0686)
$\widehat{D}_{KL2} \rightarrow$	9.3142 (1.9332)	8.7812 (1.0715)	8.4920 (0.8049)	8.4003 (0.4886)	7.9931 (0.2093)	7.9592 (0.1612)	7.5946 (0.1086)



Graphs of $\widehat{D}_{KL1}$, $\widehat{D}_{KL2}$ and D_{KL}, for $\pi = 0.50$.



Graphs of $\widehat{D}_{KL1}$, $\widehat{D}_{KL2}$ and D_{KL}, for $\pi = 0.75$.

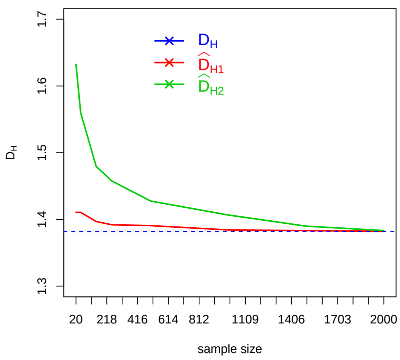
Hellinger divergence estimates for various sample size and π.

For $\pi = 0.5$, $D_H = 1.3818$.

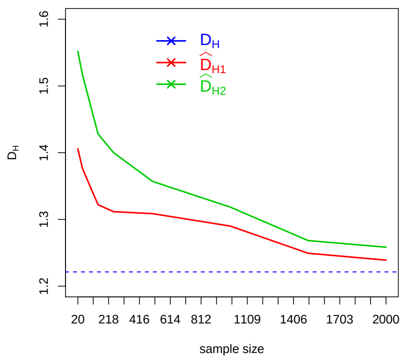
$n \rightarrow$	20	150	250	500	1000	1500	2000
$\widehat{D}_{H1} \rightarrow$	1.4106 (0.2190)	1.3968 (0.0892)	1.3920 (0.0744)	1.3907 (0.0511)	1.3842 (0.0333)	1.3832 (0.0266)	1.3822 (0.0245)
$\widehat{D}_{H2} \rightarrow$	1.6324 (0.8468)	1.4793 (0.3194)	1.4577 (0.2523)	1.4275 (0.1913)	1.4066 (0.1390)	1.3899 (0.0856)	1.3833 (0.0711)

For $\pi = 0.75$, $D_H = 1.2213$.

$n \rightarrow$	20	150	250	500	1000	1500	2000
$\widehat{D}_{H1} \rightarrow$	1.4058 (0.2891)	1.3220 (0.0822)	1.3116 (0.0601)	1.3087 (0.0439)	1.2900 (0.0276)	1.2693 (0.0188)	1.2390 (0.01455)
$\widehat{D}_{H2} \rightarrow$	1.5520 (0.6863)	1.4280 (0.2369)	1.4000 (0.2012)	1.3570 (0.1411)	1.3184 (0.1072)	1.2983 (0.0426)	1.2584 (0.0212)



Graphs of $\widehat{D}_{H1}$, $\widehat{D}_{H2}$ and D_H, for $\pi = 0.50$.



Graphs of $\widehat{D}_{H1}$,$\widehat{D}_{H2}$ and D_H for $\pi = 0.75$.

The estimates of $\widehat{D}_{KL1}$, $\widehat{D}_{KL2}$, $\widehat{D}_{H1}$ and $\widehat{D}_{H2}$ decrease along with the sample size. The Kullback-Leibler statistic $\widehat{D}_{KL1}$ converge more rapidly than $\widehat{D}_{KL2}$ to the Kullback-Leibler divergence D_{KL}.

We also observe similar behavior for $\widehat{D}_{H1}$ and $\widehat{D}_{H2}$. We mention that the estimators $\widehat{\mathcal{D}}_{KL}(f,g)$ and $\widehat{\mathcal{D}}_{H}(f,g)$ were calculated using Monte Carlo method under a given distribution g.

5.2. Model selection

To illustrate well our model selection procedure, we consider Gamma and Log-normal distributions as our candidate models. Note that the Gamma and Log-normal distributions are the most popular distributions for analyzing lifetime data. The Data Generating Process (DGP) considered has density

$$Mix(x) \equiv l(\pi) = \pi \textit{Gamma}(4.02, 0.05) + (1-\pi)\textit{Log-normal}(4.15, 0.52)$$

where $\pi \in (0,1)$ is specific to each set of experiments. In each set, 1000 samples of various sizes varies from 20 to 1000 are drawn from this mixture. We choose different values of π which are $0.00, 0.25, 0.5, 0.75, 1.00$. Although our proposed model selection procedure does not require that the data generating process belongs to either of the candidate models. We consider the two limiting cases $\pi = 0.00$ and $= 1.00$ for they correspond to the correctly specified cases. For $\pi = 0.25$ and $\pi = 0.75$ both candidate models are misspecified but not at equal distance from the DGP. These cases correspond to a DGP which is Gamma or Log-normal distributions but slightly contaminated by the other distribution. The value $\pi = 0.5$ is the value for which the Gamma and Log-normal distributions are approximately at equal distance to the mixture $l(\pi)$ according to statistics $\widehat{\mathcal{D}}_{KL}(f, f_{LN})$ and $\widehat{\mathcal{D}}_{KL}(f, f_{GA})$ for Kullback-Leibler statistics; and $\widehat{\mathcal{D}}_{H}(f, f_{LN})$ and $\widehat{\mathcal{D}}_{H}(f, f_{GA})$ for Hellinger statistics, where $\widehat{\mathcal{D}}_{KL}(f, f_{LN})$ and $\widehat{\mathcal{D}}_{H}(f, f_{LN})$ are defined in (15) and $\widehat{\mathcal{D}}_{KL}(f, f_{GA}) = \int_{A_n} \hat{f}^b_{n,h_n}(x) \ln\left(\frac{\hat{f}^b_{n,h_n}(x)}{f_{GA}(x)}\right) dx$ and $\widehat{\mathcal{D}}_{H}(f, f_{GA}) = \int_{A_n}\left(\sqrt{\hat{f}^b_{n,h_n}(x)} - \sqrt{f_{GA}(x)}\right)^2 dx$.

Our model selection statistic are now given by

$$\Lambda_{n1} := \frac{\sqrt{nh_n}}{\hat{\xi}}\left[\widehat{\mathcal{D}}_{KL}(f, f_{LN}) - \widehat{\mathcal{D}}_{KL}(f, f_{GA})\right]$$

and

$$\Lambda_{n2} := \frac{\sqrt{nh_n}}{\hat{\xi}}\left[\widehat{\mathcal{D}}_{H}(f, f_{LN}) - \widehat{\mathcal{D}}_{H}(f, f_{GA})\right].$$

Here we consider the choice of kernel and of the practical optimal bandwidth as done in subsection 5.1. The correct model represents the "true" distribution of observations while the incorrect model represents an approximation of the true model. The results of our five sets of experiments are presented as follows.

Table 1. DGP=Gamma(4.02, 0.05)

n		20	150	250	500	600	1000
Λ_{n1}		0.9558	2.8411	3.6250	5.0711	5.6306	6.6091
		(0.0491)	(0.0169)	(0.0130)	(0.0093)	(0.0083)	(0.0012)
Λ_{n2}		0.2211	3.4511	4.8747	7.2027	8.2243	11.8156
		(0.0949)	(0.0182)	(0.0128)	(0.0086)	(0.0075)	(0.0052)
Model selection	Correct	16.3%	80.6%	96.4%	99.9%	100%	100%
based on Λ_{n1}	Indecisive	83.7%	19.4%	3.6%	0.1%	0.0%	0.0%
	Incorrect	0.0%	0.0%	0.0%	0.0 %	0.0%	0.0%
Model selection	Correct	4.7%	92.9%	99.2%	100%	100%	100%
based on Λ_{n2}	Indecisive	93.3%	7.1%	0.8%	0.1%	0.0%	0.0%
	Incorrect	2.0%	0.0%	0.0%	0.0 %	0.0%	0.0%

Table 2. DGP= Log-normal (4.15, 0.52)

n		20	150	250	500	600	1000
Λ_{n1}		-0.9555	-2.8248	-3.6424	-5.2139	-5.5865	-7.2653
		(0.0604)	(0.0211)	(0.0164)	(0.0114)	(0.0107)	(0.0065)
Λ_{n2}		-0.1340	-1.5514	-2.1026	-3.0582	-3.4307	-6.1096
		(0.1002)	(0.0263)	(0.0194)	(0.0130)	(0.0119)	(0.0083)
Model selection	Correct	15.7%	81.8%	95.3%	99.9%	100%	100%
based on Λ_{n1}	Indecisive	83.3%	18.2%	4.7%	0.1%	0.0%	0.0%
	Incorrect	1.0%	0.0%	0.0%	0.0 %	0.0%	0.0%
Model selection	Correct	4.4%	34.7%	59.0%	88.4%	92.4%	100%
based on Λ_{n2}	Indecisive	93.9%	65.1%	40.8%	11.6%	7.6%	0.0%
	Incorrect	1.7%	0.2%	0.2%	0.0 %	0.0%	0.0%

Table 3. DGP= 0.25 Gamma(4.02, 0.05) + 0.75 Log-normal (4.15, 0.52)

n		20	150	250	500	600	1000
Λ_{n1}		-1.0807	-2.8587	-3.6493	-4.3181	-4.8903	-7.2653
		(0.0467)	(0.0167)	(0.0130)	(0.0093)	(0.0084)	(0.0065)
Λ_{n2}		-0.2249	-2.1477	-3.0367	-3.0582	-3.4307	-6.1096
		(0.0928)	(0.0239)	(0.0169)	(0.0118)	(0.0104)	(0.0083)
Model selection	Favor Log-Norm	19.6%	82.5%	96.4%	99.9%	100%	1000%
based on Λ_{n1}	Equivalent	80.4%	17.5%	3.6%	0.1%	0.0%	0.0%
	Favor Gamma	1.0%	0.0%	0.0%	0.0 %	0.0%	0.0%
Model selection	Favor Log-Norm	4.1%	60.4%	87.4%	98.6%	99.6%	100%
based on Λ_{n2}	Equivalent	92.7%	39.4%	12.6%	1.4%	0.4%	0.0%
	Favor Gamma	3.2%	0.2%	0.0%	0.0 %	0.0%	0.0%

Table 4. DGP= 0.5 Gamma (4.02, 0.05) + 0.5 Log-normal (4.15, 0.52)

n		20	150	250	500	600	1000
Λ_{n1}		0.6892	1.9597	2.6216	3.7084	4.0396	5.2276
		(0.0432)	(0.0162)	(0.0119)	(0.0085)	(0.0078)	(0.0060)
Λ_{n2}		0.2641	2.5695	3.5106	5.5109	5.6545	7.5232
		(0.0900)	(0.0221)	(0.0163)	(0.0103)	(0.0101)	(0.0075)
Model selection	Favor Log-Norm	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
based on Λ_{n1}	Equivalent	89.2%	51.7%	25%	3.8%	1.4%	0.0%
	Favor Gamma	10.7%	48.3%	75%	96.2 %	98.6%	100%
Model selection	Favor Log-Norm	4.0%	0.1%	0.0%	0.0%	0.0%	0.0%
based on Λ_{n2}	Equivalent	92.4%	24.9%	5.9%	0.3%	0.0%	0.0%
	Favor Gamma	3.6%	75.0%	94.1%	99.7 %	100%	100%

Table 5. DGP= 0.75 Gamma (4.02, 0.05)+ 0.25 Log-normal (4.15, 0.52)

n		20	150	250	500	600	1000
Λ_{n1}		0.7870	2.1240	2.6727	3.7705	4.1492	5.5733
		(0.0463)	(0.0161)	(0.0126)	(0.0089)	(0.0081)	(0.0060)
Λ_{n2}		0.2092	3.2673	4.4884	6.7044	7.1286	9.4481
		(0.0926)	(0.0187)	(0.0136)	(0.0091)	(0.0085)	(0.0064)
Model selection	Favor Log-Norm	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
based on Λ_{n1}	Equivalent	87.0%	44.4%	24.3%	3.5%	1.4%	0.0%
	Favor Gamma	12.8%	55.6%	75.7%	96.5%	98.6%	100%
Model selection	Favor Log-Norm	1.1%	0.0%	0.0%	0.0%	0.0%	0.0%
based on Λ_{n2}	Equivalent	94.0%	7.8%	1.3%	0.0%	0.0%	0.0%
	Favor Gamma	4.9%	92.2%	98.7%	100%	100%	100%

The first half of each table gives the average values of the statistics Λ_{n1} and Λ_{n2}. The values in parenthesis are standard errors. The second half of each table gives the probability of correct selection (PCS) which is in percentage the number of times our proposed model selection procedure based on Λ_{n1} or Λ_{n2} favors the Gamma model, the Log-normal model and indecisive. The tests are conducted at 5% nominal significance level. In the first two sets of experiments ($\pi = 0.00$ and $\pi = 1.00$) where one model is correctly specified **Table 1-2**, we use the labels *correct, incorrect* and *indecisive* when a choice is made. The first halves of **Tables 1-5** confirm our asymptotic results. The values of statistics Λ_{n1} and Λ_{n2} increase along with the sample size in **Tables 1, 4** and **5**; and decrease along with the sample size in **Tables 2** and **3** when the models are correctly specified and when the models are mis-specified.

Turning to the second halves of **Tables 1** and **2**, we note that the percentage of correct choice using model selection statistic steadily increases and ultimately converge to 100% expected in **Tables 1** with a short domination of the model selection based on Λ_{n1} over the Λ_{n2}.

This preceding comments can be applied to the second halves of **Tables 3**, **4** and **5**. In all tables, as sample size increases, the percentages of rejection (incorrect model) tends to zero (for Λ_{n2}) and still having the same value .i.e. 0.0% (for Λ_{n1}).

For $n = 500$, we plot the histogram of datasets and overlay the curves for Gamma and Log-normal distributions.

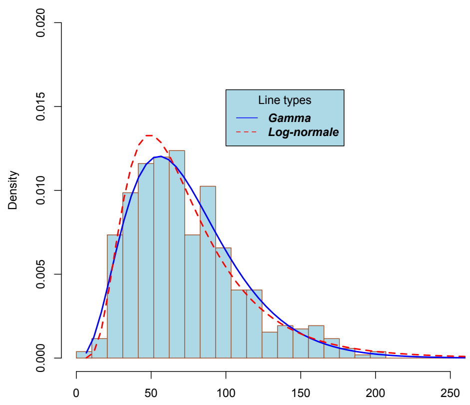


Fig. 1. DGP = Gamma(4.02, 0.05)

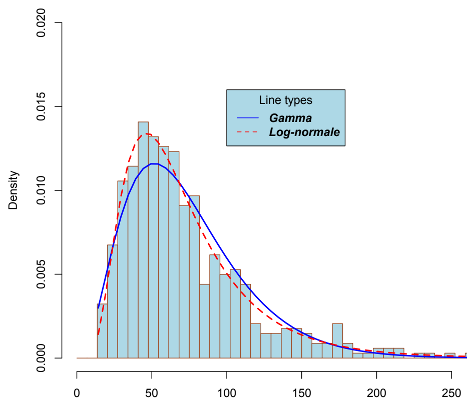


Fig. 2. DGP = Log-normal (4.15, 0.52).

As can be observed in **Figure 2** and **3** , the Log-normal distribution is distinguished from Gamma distribution and closely approximates the data sets. In **Figure 1** and **5** the Gamma distributions is much closer to the data sets. When $\pi = 0.5$ (**Figure 4**) the Gamma distribution and Log-normal distribution try to be closer to the data set. This follows from the fact that they are equidistant from the *DGP*.

6. Conclusion

We studied in this paper a bias reduced kernel density estimator and we derived a nonparametric ϕ-divergence estimator and their asymptotic properties such as the asymptotic normality limits and the strong consistencies. We further considered an informational criterion for model selection based on nonparametric ϕ-divergence estimator. This method allows us to take into account the stochastic nature of variations inherent in the values of the nonparametric ϕ-divergence estimator. Specifically, we have proposed some convenient asymptotically standard normal and hypothesis testing based on the ϕ-divergence which is a general case of Kullback-Leibler and Hellinger divergence estimators constructed in terms of the bias reduced kernel density estimator.

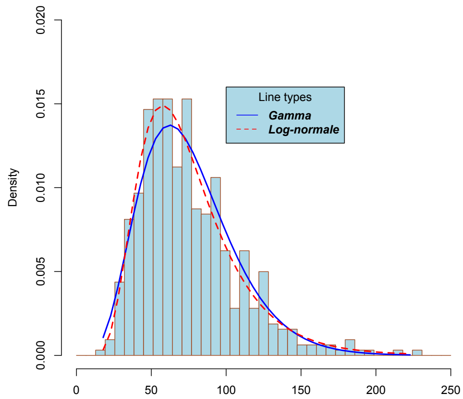


Fig. 3. DGP = 0.25 Gamma(4.02, 0.05) + 0.75 Log-normal (4.15, 0.52).

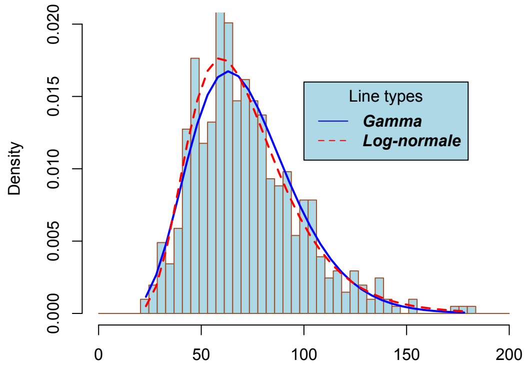


Fig. 4. DGP = 0.5 Gamma(4.02, 0.05) + 0.5 Log-normal (4.15, 0.52).

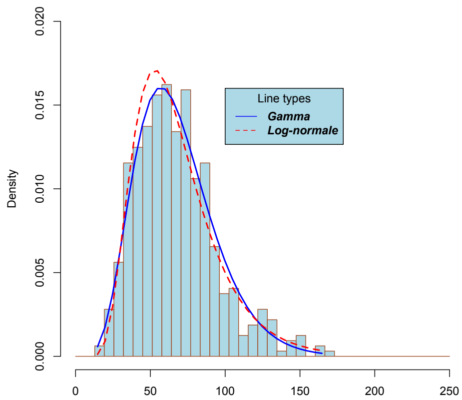


Fig. 5. DGP = 0.75 Gamma(4.02, 0.05) + 0.25 Log-normal (4.15, 0.52).

We test whether the candidate models are equally close to the true distribution against the alternative hypothesis that one model is closer than the other where closeness of a model is measured according to the discrepancy implicit in the ϕ-divergence type statistics used. This model selection procedure is especially suitable for testing the null hypothesis that some competing models are equally close to the observed data and performs very well even in small sample problems.

Acknowledgements

This research was supported, in part, by grants from CEA-SMA Centre d'Excellence Africain en Sciences
Mathématiques et Applications IMSP, B.P. 613 Porto-Novo, Benin. The authors are grateful for helpful comments from Carlos OGOUYANDJOU and highly thank the unknown referees for valuable comments and suggestions and the editorial board for assistance in improving the paper.

References

M. Rosenblatt (1956). On Estimation of a Probability Density Function and the Mode. *The Annals of Mathematical Statistics*, Vol 33, pp. 1065-1076. doi.org/10.1214/aoms/1177704472.

E. Parzen (1962). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, Vol 27, pp. 832-837. doi.org/10.1214/aoms/1177728190.

J. Dony and M. Einmahl (2006). Weighted uniform consistency of kernel density estimators with general bandwidth sequences. *Electronic Journal of Probability*, Vol 11(33), pp. 844-859.

S. Bouzebda and I. Elhattab (2011). Uniform in Bandwidth consistency of Kernel-Type Estimators of Shannon's Entropy. *Electronic Journal of Statistics*, Vol 5, pp. 440-459. doi: 10.1214/11-EJS614.

U. Einmahl and D. M. Mason (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Annals of Statistics*, Vol 33(3), pp. 1380-1403. doi.10.1214/009053605000000129.

P. Ngom., H. Dhaker, E, Deme and P. Mendy (2016). Kernel-Type Estimators of Divergence Measures and Its Strong Uniform Consistency. *American Journal of Theoretical and Applied Statistics*, Vol 5(1), pp. 13-22. doi: 10.11648/j.ajtas.20160501.13.

M. Rudemo (1982). Empirical choice of histograms and kernel density estimators. *Scandinavia, Journal of Statistics*, Vol 9, pp. 65-78.

A. W. Bowman., P. Hall and D. M. Titterington (1984). Cross-validation in nonparametric estimation of probabilities and probability densities. *Biometrika*, Vol 71, pp. 65-78.

X. Xiaoran and W. Jingjing (2014). Some Improvement on Convergence Rates of Kernel Density Estimator. *Published Online June 2014 in SciRes.*, Vol 5, pp. 1684-1696. doi.org/10.4236/am.2014.511161.

I. Csiszár (1963) Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica.*, Vol 2, pp. 299-318.

S. M. Ali and S. D. Silvey (1966) A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society B*, Vol 28, pp. 131-142.

S. Kullback and R.A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics*, Vol 22(1), pp. 79-86. doi. 10.1214/aoms/1177729694. MR 39968.

R. J. Bearn (1977). Minimum Hellinger Distance Estimates for Parametric Models. *Annals of Mathematical Statistics*, Vol 5(3), pp. 445-463.

L. Pardo (2006). Statistical Inference based on Divergence Measures. *Chapman, Hall/CRC, Taylor and Francis Group.*,6000 Broken Sound Parkway NW, Suite 300.

A. Basu., I.Harris., N. Hjort and M. Jones (1998). Robust and efficient estimation by minimizing a density power divergence.), with *Biometrika*, Vol 85, pp. 549-559.

M. Samiuddin and G.M. El-Sayyad (1990). On Nonparametric Kernel Density Estimates. *The Annals of Statistics*, Vol 77, pp. 865-874. doi.org/10.1093/biomet/77.4.865.

M. Samiuddin and G.M. El-Sayyad (1992). A New Kernel Density Estimate. *Journal of Nonparametric Statistics*, Vol 3, pp. 1-11. doi.org/10.1080/10485259308832568.

M. Mbuba., W.Samson and D. Marangu (1984). Central Limit Theorem and Its Applications in Determining Shoe Sizes of University Students. *Asian Journal of Probability and Statistics*, Vol 3(1), pp. 1-9.

C. Arzelà (1885). Sulla integrazioneper serie, Atti Acc. Lincei Rend. *Rome*, Vol 4, pp. 532-537.

D. Nolan and D. Pollard (1987). U-processes: Rate of Convergence. *The Annals of Statistics*, Vol 15(2), pp. 780-799.

D. Van and J. Wellner (1996). Weak Convergence and Empirical. Processes: With Applications to Statistics. *Springer, New York*, 1996.

A. Tsybakov (2008). Introduction to Nonparametric Estimation. *Springer Series in Statistics*, ISBN: 978-0-387-79051-0, doi. 10.1007/978-0-387-79052-7.

H. Dhaker., P. Ngom., P. Mendy and E. Dème (2017). Generalized divergence criteria for model selection between random walk and AR(1) model. *Journal of Statistics: Advances in Theory and Applications*, Vol 17(2), pp.83-109. doi.org/10.18642/jsata_-7100121830.

Amadou Diadié BA,Gane Samb LO, Diam Ba (2017). Divergence Measures Estimation and Its Asymptotic Normality Theory Using Wavelets Empirical Processes I (2018). *Journal of Statistical Theory and Applications*, Vol. 17, No. 1 (March 2018), pp. 158–171.

Amadou Diadié BA,Gane Samb LO, Diam Ba (2018). Divergence Measures Estimation and Its Asymptotic Normality Theory Using Wavelets Empirical Processes II (2018). *Afrika Statistika*, pp. 1667-1681. DOI : http://dx.doi.org/10.16929/as/1667.127.

Amadou Diadié BA,Gane Samb LO, Diam Ba (2019). Divergence Measures Estimation and Its Asymptotic Normality Theory Using Wavelets Empirical Processes III. *Journal of Statistical Theory and Applications*, Volume 18 (2), pp. 113- 122. Doi: https://doi.org/10.2991/jsta.d.190514.002. (June 2009).