JIRSS (2019)
Vol. 18, No. 02, pp 173-197
DOI:10.29252/jirss.18.2.173

Shrinkage Estimators of the Probability Density Function and their Asymptotic Properties under Association

Marzieh Mahmoudi [1], Mohammad Arashi [1], and Ahmad Nezakati [1]

[1] Department of Statistics, Faculty of Mathematical Sciences, Shahrood University of Technology, Iran.

Received: 18/09/2018, Revision received: 24/05/2019, Published online: 29/08/2019

Abstract. In the present article, we develop the well-known preliminary test and Stein-type estimators for the probability density function under association. In this respect, we derive the asymptotic characteristics of the proposed estimators under a set of local alternatives. Some numerical studies are provided for supporting the findings. The result of this article improves the kernel estimate of the marginal probability density function of a strictly stationary sequence of associated random variables. For practical sake, the behavior of the proposed estimators is also analyzed using a real data set.

Keywords. Association, Asymptotic MSE, Kernel Estimate, Preliminary Test, Shrinkage.

MSC: 62G07; 62G20; 62G10; 62F12.

Corresponding Author: Mohammad Arashi (m_arashi_stat@yahoo.com)
M. Mahmoudi (marzieh_mahmoudi@yahoo.com)
Ahmad. Nezakati (anezakati591@gmail.com)

1 Introduction

As of 1967, when Esary et al. (1967) introduced the concept of positively associated (PA) random variables (r.v.s) in statistics, there has been a significant number of research papers on this subject matter. Negatively associated (NA) r.v.s was introduced by Joag-Dev and Proschan (1983). The negative association does occur in a number of important cases, but is not as popular as positive association. In a snapshot, one may say that positive association occurs often in certain reliability theory problems, as well as in some important models employed in statistical mechanics. The negative association also appears in some reliability theory problems, but less often than a positive association. For a selected review of the subject matter of association, interested readers may refer to Roussas (1999).

The definition of association is given below, and then a basic notation is introduced; this will allow the formulation of the basic assumptions in this paper.

Definition 1.1. For a finite index set I, the r.v.s $\{X_i, i \in I\}$ are said to be PA if for any real-valued coordinate-wise increasing functions G and H defined on R^I,

$$Cov[G(X_i, i \in I), H(X_j, j \in I)] \geq 0,$$

provided $E\left[G^2(X_i, i \in I)\right] < \infty$, $E\left[H^2(X_j, j \in I)\right] < \infty$.

These r.v.s are said to be NA, if for any disjoint non-empty subsets A and B of I, and any coordinate-wise increasing functions G and H with $G : \mathbb{R}^A \to \mathbb{R}$, $H : \mathbb{R}^B \to \mathbb{R}$, $E\left[G^2(X_i, i \in A)\right] < \infty$, and $E\left[H^2(X_j, j \in B)\right] < \infty$, the inequality $Cov[G(X_i, i \in A), H(X_j, j \in B)] \leq 0$ holds. If I is not finite, the r.v.s $\{X_i, i \in I\}$ are said to be PA or NA, if any finite sub-collection is a set of PA or NA r.v.s, respectively. When no distinction is necessary, PA and NA r.v.s will be referred to collectively as associated r.v.s. The underlying stochastic process consists of associated r.v.s, forming a strictly stationary sequence, having a finite second moment, and a one-dimensional marginal probability density function (PDF) $f(.)$.

Now, let the function K be a (known) bounded PDF (kernel), and $h = h_n$ is a sequence of positive bandwidths tending to 0, as $n \to \infty$. The ordinary kernel estimator of $f(.)$ is defined as

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \quad K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right). \tag{1.1}$$

The reader is referred to Wand and Jones (1995) for more details. The purpose of this study is to extend the preliminary-test and Stein-type estimators of Saleh (2006) for $f(\cdot)$ by making use of the kernel density estimator (1.1). We also refer to Arashi and Mahmoodi (2014), Saleh and Ghania (2016), and Saleh et al. (2018) to mention a few recent studies about the shrinkage estimate of the distribution function.

1.1 Regularity Conditions

For our purpose, we need the following assumptions:

(A1) :

- (i) The r.v.s $X_1, X_2, ...,$ form a strictly stationary sequence, and $f(.)$ is the one-dimensional marginal PDF (with respect to the Lebesgue measure).
- (ii) The X_i's are associated (either PA or NA).
- (iii) The X_i's have finite second moments, $EX_i^2 < \infty$, and $\sum_{j=1}^{\infty} |Cov(X_1, X_{j+1})| < \infty$.
- (iv) If $f_{X_1,X_j}(\cdot,\cdot)$ is the joint PDF of the r.v.s $X_1, X_j,\ j \geq 1$, then $|f_{X_1,X_j}(u,v) - f_{X_1}(u) f_{X_j}(v)| \leq C < \infty$, for some constant C and all $u, v \in \mathbb{R}$.

(A2) :

- (i) The kernel function K is such that:

$$K(u) \leq C, u \in \mathbb{R}; \quad \lim_{|u|\to\infty} (|u|K(u)) = 0.$$

- (ii) The derivative $(d/du)K(u) = K'(u)$ exists for all $u \in \mathbb{R}$ and is bounded $|K'(u)| \leq B$, for $u \in \mathbb{R}$.

(A3) : Assuming that $0 < p = p_n < n$, $0 < q = q_n < n$ are integers tending to ∞ along with n, and let $0 \leq k = k_n \xrightarrow[n\to\infty]{} \infty$ being defined by $k = [n/(p+q)]$ (where $[x]$ stands for the integer part of x), so that $k(p+q) \leq 1$ and $k(p+q)/n \xrightarrow[n\to\infty]{} 1$. Also, $h_n > 0$ being bandwidths. Then there is a determination of them for which:

- (i) $p_n k_n / n \to 1$ as $n \to \infty$.
- (ii) $p_n h_n \to 0$ as $n \to \infty$ and $p_n^2 / n h_n \to 0$ as $n \to \infty$.

(iii) $(1/h^3)\sum_{j=q_n}^{\infty}|Cov(X_1, X_{j+1}| \to 0$.

For more details see Roussas (2000).

1.2 Plan of the Paper

In Section 2, we will be proposing the preliminary-test and shrinkage estimators of $f(\cdot)$. Section 3 consists of the asymptotic properties of the proposed estimators for an associated random sample. As mentioned the null prior knowledge $f(x) = f_0(x)$ fails in improving the asymptotic properties of $\hat{f}(x)$ and hence asymptotic results are developed for a set of local alternatives. Section 4 is devoted to some numerical comparisons along with an analysis of a real data set. We conclude in Section 5.

2 Improved Estimators

Suppose that the pre-specified density $f_0(\cdot)$ is under suspicion of beibg the true population model. According to Fisher's receipt, the early first approach for modeling is to check whether the null hypothesis

$$H_0 : f(x) = f_0(x) \tag{2.1}$$

is rejected or not. For our purpose, upon the acceptance of H_0, the model $f_0(\cdot)$ is adopted, otherwise, we take the kernel density estimator $\hat{f}(\cdot)$ as our population model estimate. Hence, we need to develop a test statistic for testing (2.1) using the associated r.v.s. The following result is the key one.

Theorem 2.1 (Roussas, 2000). *Let the r.v.s $X_1, ..., X_n$ is satisfy the assumptions (A1)-(A3), and in addition: Either (a): (i) The derivative f' exists and is bounded: (ii) $\int |u|K(u)du < \infty$; (iii) $nh_n^3 \to 0$. Or (b) (i) The second-order derivative f'' exists and is bounded; (ii) $\int uK(u)du = 0$; (iii) $\int u^2K(u)du < \infty$; (iv) $nh_n^5 \to 0$. Then*

$$(nh)^{1/2}[\hat{f}(x) - f(x)] \xrightarrow{d} N(0, \sigma^2(x)), \quad \sigma^2(x) = f(x)\int K^2(u)du, \quad x \in C(f),$$

where $\hat{f}(x)$ is the kernel estimate of $f(x)$ defined in (1.1) and $C(f)$ is the continuity set of f.

To develop a test statistic for testing (the null hypothesis), we apply the result of Theorem 2.1 and using the pivotal quantity (see Shao, 2003), suggest using the following test statistic for testing H_0.

$$\mathcal{L}_n(x) = nh\sigma^{-2}(x)\left(\hat{f}(x) - f_0(x)\right)^2. \tag{2.2}$$

It can be readily deduced from Theorem 2.1 that $\mathcal{L}_n(x)$ asymptotically follows a central chi-square distribution with one degree of freedom (d.f.) under H_0 as $n \to \infty$, $h \to 0$ and $nh \to \infty$.

If the null hypothesis H_0 is accepted, we choose $f_0(\cdot)$ as the estimate of $f(\cdot)$. Now we suggest choosing the ordinary kernel estimator if H_0 is rejected. Hence, one can join these two extremes and form a combined estimator using the indicator function. This is known in the literature as the preliminary testing or testimation. We refer to Saleh (2006) for an extensive overview on this topic. Hence, combining the estimators $f_0(\cdot)$ and $\hat{f}(\cdot)$, under the preliminary testing approach, gives the preliminary test estimator (PTE) of $f(.)$ as

$$\hat{f}^{PT}(x) = \hat{f}(x) - \left(\hat{f}(x) - f_0(x)\right) I(\mathcal{L}_n(x) < \chi_1^2(\alpha)), \tag{2.3}$$

where $I(A)$ is the indicator function of the set A and $\chi^2(\alpha)$ is the α-level critical value. In this case, one estimates $f(x)$ by $f_0(x)$ if the test $\mathcal{L}_n(x)$ accepts H_0 at the level of significance α; otherwise, $\hat{f}(x)$ is used.

Note that the PTE is discontinuous, leading to extreme choices for the estimators. Moreover, it is highly dependent on the level of significance α. To make a smooth transition of the Equation (2.3), making use of the test statistic (2.2), we suggest using the Stein-type shrinkage kernel density estimator (SSKDE) of $f(x)$ as

$$\hat{f}^{S}(x) = \hat{f}(x) - d\left(\hat{f}(x) - f_0(x)\right)\mathcal{L}_n^{-\frac{1}{2}}(x), \tag{2.4}$$

where $d > 0$ is the shrinkage factor. It is easy to show that

$$\hat{f}^{S}(x) = f_0(x) + \left[1 - d\mathcal{L}_n^{-\frac{1}{2}}(x)\right]\left(\hat{f}(x) - f_0(x)\right).$$

Hence, for the values $\mathcal{L}_n^{\frac{1}{2}}(x) < d$, the shrinkage factor $\left[1 - d\mathcal{L}_n^{-\frac{1}{2}}(x)\right]$ becomes negative and eliminates the shrinkage effect. As a remedy, one can consider the positive part in proposing a smooth shrinkage estimator. For this purpose, we propose the positive-rule Stein-type estimator (PRSE), given by

$$\begin{aligned}\hat{f}^{S+}(x) &= f_0(x) - \left[1 - d\mathcal{L}_n^{-\frac{1}{2}}(x)\right] I(\mathcal{L}_n(x) > d)\left(\hat{f}(x) - f_0(x)\right) \\ &= \hat{f}^{S}(x) - (\hat{f}^{S}(x) - f_0(x)) I(\mathcal{L}_n(x) < d). \end{aligned} \tag{2.5}$$

All the aforementioned estimators belong to the class of shrinkage estimators defined by

$$\hat{f}^{\text{Shrinkage}}(x) = \hat{f}(x) - \left(\hat{f}(x) - f_0(x)\right) g(\mathcal{L}(x)),$$

where $g(\cdot)$ is a measurable function. In this shrinkage structure, the ordinary kernel estimator will be shrunken to the prior information $f_0(\cdot)$. Apparently, taking $g(x) = I(\mathcal{L}_n(x) < \chi_1^2(\alpha))$ and $g(x) = d\mathcal{L}_n^{-\frac{1}{2}}(x)$, gives the PTE and SSKDE, respectively. After some algebra, it can be shown that the $g(\cdot)$ function for the PRSE has the form $g(x) = d\mathcal{L}_n^{-\frac{1}{2}}(x) I(\mathcal{L}_n(x) \geq d) + I(\mathcal{L}_n(x) < d)$.

3 Asymptotic Properties

In this section, we derive the asymptotic distributional mean square error (ADMSE) expressions for the PTE, SSKDE, and PRSE.

In a similar fashion as in Lemma 2 of Arashi and Mahmoodi (2014), it can be shown that under the fixed alternative hypothesis, $A_\delta : f(x) = f_0(x) + \delta$, the estimators $\hat{f}^{PT}(x)$, $\hat{f}^{S}(x)$ and $\hat{f}^{S+}(x)$ are asymptotically distributed as $\hat{f}(x)$. For more clarity, we represent their result here only for the SSKDE, to save space.

Lemma 3.1. *Under the fixed alternatives $A_\delta : f(x) = f_0(x) + \delta$, as $n \to \infty$, $h \to 0$ such that $nh \to \infty$,*

$$(nh)^{\frac{1}{2}}\left(\hat{f}_n^{S}(x) - f(x)\right) = (nh)^{\frac{1}{2}}\left(\hat{f}_n(x) - f(x)\right) + o_p(1).$$

Proof. Under A_δ and using (2.2), we have

$$\begin{aligned}
\lim_{n\to\infty,h\to 0} E\left[n\left(\hat{f}_n^S(x)-\hat{f}_n(x)\right)^2\right] &= d^2 \lim_{n\to\infty} E\left[n\left(\hat{f}_n(x)-\hat{f}_0(x)\right)^2 \text{Ł}_n^{-1}(x)\right] \\
&= d^2\rho^2(x) \lim_{n\to\infty,h\to 0} E\left[h^{-1}\right] \\
&\to 0 \qquad (3.1)
\end{aligned}$$

Thus we have $(nh)^{\frac{1}{2}}\left(\hat{f}_n^S(x)-f(x)\right) = (nh)^{\frac{1}{2}}\left(\hat{f}_n(x)-f(x)\right)+o_p(1)$. □

Lemma 3.1 clearly shows that the asymptotic characteristics of $\hat{f}_n^S(x)$ and $\hat{f}_n(x)$ are the same. This fact is true for the PTE and PRSE as well. Hence, studying the asymptotic behavior of $\hat{f}^{PT}(x)$, $\hat{f}^S(x)$, and $\hat{f}^{S+}(x)$ is worthless under the set of fixed alternatives because of asymptotic distributional similarity to $\hat{f}(x)$.

To combat this problem, consider a sequence of local alternatives $K_{(n)}$ defined by (see Saleh, 2006)

$$\{K_{(n)}\} : f_{(n)}(x) \equiv f(x) = f_0(x) + (nh)^{-1/2}\delta, \qquad (3.2)$$

where δ is a fixed positive number. In this case, the test statistics $\mathcal{L}_n(x)$ is approximately distributed as a non-central chi-square distribution with one degree of freedom (df) and non-centrality parameter

$$\frac{\Delta^2}{2} = \frac{\delta^2}{2\sigma^2}. \qquad (3.3)$$

Now, under the alternative hypothesis $K_{(n)}$ given by (3.2) and the assumed regularity conditions in Section 1.1, we may observe that

$$\begin{aligned}
(nh)^{1/2}\left[\hat{f}^{PT}(x)-f_{(n)}(x)\right] &= (nh)^{1/2}\left[(\hat{f}(x)-f_{(n)}(x))-(\hat{f}(x)-f_0(x))I(\mathcal{L}_n(x)<\chi_1^2(\alpha))\right] \\
&= \sigma(x)\left[Z-(Z+\Delta)I((Z+\Delta)^2<\chi_1^2(\alpha))\right]+o_p(1),
\end{aligned}$$

and also

$$\begin{aligned}
(nh)^{1/2}\left[\hat{f}^S(x)-f_{(n)}(x)\right] &= (nh)^{1/2}\left[(\hat{f}(x)-f_{(n)}(x))-d\mathcal{L}_n^{-1/2}(x)\left(\hat{f}(x)-f_0(x)\right)\right] \\
&= (nh)^{1/2}\left[\hat{f}(x)-f_{(n)}(x)\right]-d\sigma\frac{\hat{f}(x)-f_0(x)}{|\hat{f}(x)-f_0(x)|}
\end{aligned}$$

$$= \quad \sigma(x)\left[Z - d\frac{Z+\Delta}{|Z+\Delta|}\right] + o_p(1).$$

Similarly, under $K_{(n)}$ and the assumed regularity conditions, it yields

$$\begin{aligned}(nh)^{1/2}\left[\hat{f}^{S+}(x) - f_{(n)}(x)\right] &= (nh)^{1/2}\left[(\hat{f}^{S}(x) - f_{(n)}(x)) - (\hat{f}^{S}(x) - f_0(x))I(\mathcal{L}_n(x) < d)\right] \\ &= \sigma(x)\left[Z - d\frac{Z+\Delta}{|Z+\Delta|}\right. \\ &\quad \left. -((Z+\Delta) - d\frac{Z+\Delta}{|Z+\Delta|})I((Z+\Delta)^2 < d)\right] \\ &= \sigma(x)\left[Z - d\frac{Z+\Delta}{|Z+\Delta|}\right. \\ &\quad \left. -(Z+\Delta)(1 - d|Z+\Delta|^{-1})I((Z+\Delta)^2 < d))\right] + o_p(1),\end{aligned}$$

where

$$Z = (nh)^{1/2}\sigma^{-1}(x)[\hat{f}(x) - f_{(n)}(x)] \xrightarrow{\mathcal{D}} N(0,1), \quad as \quad n \to \infty.$$

In the following, we derive closed-form expressions for the asymptotic distributional bias (ADB) and ADMSE. For any estimator $\hat{f}(x)$ of $f_{(n)}(x)$, ADB and ADMSE are respectively defined as

$$\begin{aligned}\text{ADB}(\hat{f}(x)) &= \lim_{nh\to\infty} E\left[(nh)^{\frac{1}{2}}\left(\hat{f}(x) - f_{(n)}(x)\right)\right], \\ \text{ADMSE}(\hat{f}(x)) &= \lim_{nh\to\infty} E\left[nh\left(\hat{f}(x) - f_{(n)}(x)\right)^2\right].\end{aligned}$$

Lemma 3.2 (Saleh, 2006). *If $Z \sim N(\Delta, 1)$ and $\varphi(.)$ be a Borel measurable function, then*

(i) $E\left(Z\varphi\left(Z^2\right)\right) = \Delta\, E\left(\varphi\left(\chi_3^2\left(\Delta^2\right)\right)\right)$,

(ii) $E\left(Z^2\varphi\left(Z^2\right)\right) = E\left(\varphi\left(\chi_3^2\left(\Delta^2\right)\right)\right) + \Delta^2 E\left(\varphi\left(\chi_5^2\left(\Delta^2\right)\right)\right)$,

(iii) $E\,|Z| = \sqrt{\frac{2}{\pi}}e^{-\frac{\Delta^2}{2}} + \Delta\left\{2\Phi\left(\Delta\right) - 1\right\}$,

(iv) $E\left[\frac{Z}{|Z|}\right] = 1 - 2\Phi(-\Delta)$,

where $\chi^2_\gamma(\Delta^2)$ stands for the non-central chi-square r.v. with γ degrees of freedom and non-central parameter Δ^2 and $\Phi(.)$ is the cumulative distribution function (CDF) of the standard normal distribution.

Theorem 3.1. *Under the local alternatives $K_{(n)}$ in* (3.2) *and the assumptions (A1)-(A3), ADB and ADMSE of $\hat{f}^{PT}(x)$ are respectively given by*

$$\begin{aligned} ADB(\hat{f}^{PT}(x)) &= -\delta H_3(\chi^2_1(\alpha);\Delta^2), \\ ADMSE(\hat{f}^{PT}(x)) &= \sigma^2(x)\left[1 - H_3(\chi^2_1(\alpha);\Delta^2) + \Delta^2[(2H_3(\chi^2_1(\alpha);\Delta^2) - H_5(\chi^2_1(\alpha);\Delta^2)]\right], \end{aligned}$$

where $H_\gamma(\cdot;\Delta^2)$ is the CDF of a non-central chi-squared distribution with γ degrees of freedom and the non-central parameter Δ^2 given by (3.3)*.*

Proof. First, note that under the local alternatives $K_{(n)}$ we conclude

$$Z = (nh)^{1/2}\sigma^{-1}(x)\left[\hat{f}(x) - f_{(n)}(x)\right] \xrightarrow{\mathcal{D}} N(0,1), \quad as \quad n \to \infty.$$

Thus, using Lemma 3.2, we obtain

$$\begin{aligned} \text{ADB}(\hat{f}^{PT}(x)) &= \lim_{nh\to\infty} E\left[(nh)^{\frac{1}{2}}(\hat{f}^{PT}(x) - f_{(n)}(x))\right] \\ &= E\sigma(x)\left[Z - (Z+\Delta)I((Z+\Delta)^2 < \chi^2_1(\alpha))\right] \\ &= -\delta H_3(\chi^2_1(\alpha);\Delta^2). \end{aligned}$$

For the ADMSE, we have

$$\begin{aligned} \text{ADMSE}(\hat{f}^{PT}(x)) &= \lim_{nh\to\infty} E\left[nh(\hat{f}^{PT}(x) - f_{(n)}(x))^2\right] \\ &= E\left[\sigma(x)[Z - (Z+\Delta)I((Z+\Delta)^2 < \chi^2_1(\alpha))]\right]^2 \\ &= \sigma^2(x)\left[1 - (Z+\Delta)^2 I((Z+\Delta)^2 < \chi^2_1(\alpha)) + 2\Delta(Z+\Delta)I((Z+\Delta)^2 < \chi^2_1(\alpha))\right]. \end{aligned}$$

The result follows from Lemma 3.2. □

Theorem 3.2. *Under the local alternatives $K_{(n)}$ and the assumptions (A1)-(A3), ADB and ADMSE of $\hat{f}^S(x)$ are respectively given by*

$$\begin{aligned} ADB(\hat{f}^S(x)) &= d\sigma(x)\,[2\Phi(\Delta) - 1]\,. \\ ADMSE(\hat{f}^S(x)) &= \sigma^2(x)\left[1 + d^2 - 2d\left(\frac{2}{\pi}\right)^{\frac{1}{2}} e^{-\frac{\Delta^2}{2}}\right]. \end{aligned}$$

Proof. The following fact can be readily obtained under the local alternatives $K_{(n)}$

$$(nh)^{\frac{1}{2}}(\hat{f}(x) - f_0(x)) \xrightarrow{\mathcal{D}} N(\delta, \sigma^2(x)).$$

Thus, we have $\text{ADB}(\hat{f}(x)) = 0$. Then,

$$\begin{aligned} \text{ADB}(\hat{f}^S(x)) &= \lim_{nh\to\infty} E\left[(nh)^{\frac{1}{2}}\left(\hat{f}^S(x) - f_{(n)}(x)\right)\right] \\ &= \text{ADB}(\hat{f}(x)) - d\sigma(x) \lim_{nh\to\infty} E\left(\frac{\hat{f}(x) - f_0(x)}{|\hat{f}(x) - f_0(x)|}\right) \\ &= -d\sigma(x)E\left[\frac{Z + \Delta}{|Z + \Delta|}\right]. \end{aligned}$$

The result follows using Lemma 3.2.

For the ADMSE, we have

$$\begin{aligned} \text{ADMSE}(\hat{f}^S(x)) &= \lim_{nh\to\infty} E\left[nh\left(\hat{f}^S(x) - f_{(n)}(x)\right)^2\right] \\ &= \text{ADMSE}(\hat{f}(x)) + d^2 \lim_{nh\to\infty} E\left[nh\left(\hat{f}(x) - f_0(x)\right)^2 \mathcal{L}_n^{-1}(x)\right] \\ &\quad -2d\sigma(x) \lim_{nh\to\infty} E\left[(nh)^{\frac{1}{2}}\left(\hat{f}(x) - f_n(x)\right)\left(\frac{\hat{f}_n(x) - f_0(x)}{|\hat{f}(x) - f_0(x)|}\right)\right] \\ &= \text{ADMSE}(\hat{f}(x)) + d^2\sigma^2(x) \\ &\quad -2d\sigma^2(x)\left(E(|Z + \Delta|) - \Delta E\left[\frac{Z + \Delta}{|Z + \Delta|}\right]\right) \\ &= \sigma^2(x)\left[1 + d^2 - 2d\left(\frac{2}{\pi}\right)^{\frac{1}{2}} e^{-\frac{\Delta^2}{2}}\right]. \end{aligned}$$

The result follows from Lemma 3.2 and the fact that $\text{ADMSE}(\hat{f}(x)) = \sigma^2(x)$. □

Theorem 3.3. *Under the local alternatives* $K_{(n)}$ *in* (3.2) *and the assumptions (A1) and (A2), ADB and ADMSE of* $\hat{f}^{S+}(x)$ *are respectively given by*

$$\begin{aligned} ADB(\hat{f}^{S+}(x)) &= \sigma[d(\Phi(\sqrt{d}-\Delta)-\Phi(\sqrt{d}+\Delta))-\Delta H_3(d;\Delta^2)], \\ ADMSE(\hat{f}^{S+}(x)) &= \sigma^2(x)\Big[1+d^2(1-H_1(d;\Delta^2))+H_3(d;\Delta^2)+\Delta^2 H_5(d;\Delta^2) \\ &\quad -\sqrt{\frac{2}{\pi}}(\Phi(\sqrt{d}-\Delta)-\Phi(-\sqrt{d}-\Delta) \\ &\quad +e^{-\frac{(\sqrt{d}+\Delta)^2}{2}}(d-\sqrt{d}+2\Delta)+e^{-\frac{(\sqrt{d}-\Delta)^2}{2}}(d-\sqrt{d}))\Big]. \end{aligned}$$

Proof.

$$\begin{aligned} \mathrm{ADB}(\hat{f}^{S+}(x)) &= \lim_{nh\to\infty} E[(nh)^{\frac{1}{2}}(\hat{f}^{S+}(x)-f_{(n)}(x))] \\ &= E\sigma(x)\left[Z-d\frac{Z+\Delta}{|Z+\Delta|}-(Z+\Delta)(1-d|Z+\Delta|^{-1})I((Z+\Delta)^2<d))\right] \\ &= \sigma(x)\left[-d(2\Phi(\Delta)-1)-\Delta H_3(d;\Delta^2)+d(2\Phi(\Delta)+\Phi(\sqrt{d}-\Delta)-\Phi(\sqrt{d}+\Delta)-1)\right] \\ &= \sigma(x)[d(\Phi(\sqrt{d}-\Delta)-\Phi(\sqrt{d}+\Delta))-\Delta H_3(d;\Delta^2)]. \end{aligned}$$

For the ADMSE, we have

$$\begin{aligned} \mathrm{ADMSE}(\hat{f}_n^{S+}(x)) &= \lim_{nh\to\infty} E[nh(\hat{f}_n^{S+}(x)-f_{(n)}(x))^2] \\ &= E\left[\sigma(x)\left(Z-d\frac{Z+\Delta}{|Z+\Delta|}-(Z+\Delta)(1-d|Z+\Delta|^{-1})I((Z+\Delta)^2<d))\right)\right]^2 \\ &= \sigma^2(x)E\left[Z-d\frac{Z+\Delta}{|Z+\Delta|}(1-I((Z+\Delta)^2<d))-(Z+\Delta)I((Z+\Delta)^2<d))\right]^2 \\ &= \sigma^2(x)E\Big[Z^2+d^2(1-I((Z+\Delta)^2<d))^2+(Z+\Delta)^2 I((Z+\Delta)^2<d) \\ &\quad -2dZ\frac{Z+\Delta}{|Z+\Delta|}(1-I((Z+\Delta)^2<d))-2Z(Z+\Delta)I((Z+\Delta)^2<d) \\ &\quad +2d(Z+\Delta)\frac{Z+\Delta}{|Z+\Delta|}(1-I((Z+\Delta)^2<d))I((Z+\Delta)^2<d)\Big] \end{aligned}$$

$$
\begin{aligned}
&= \sigma^2(x)E\Big[1 + d^2(1 - H_1(d;\Delta^2)) + H_3(d;\Delta^2) + \Delta^2 H_5(d;\Delta^2) \\
&\quad -d\sqrt{\frac{2}{\pi}}(e^{-\frac{(\sqrt{d}+\Delta)^2}{2}} + e^{-\frac{(\sqrt{d}-\Delta)^2}{2}}) - \sqrt{\frac{2}{\pi}}(\Delta(e^{-\frac{(\sqrt{d}+\Delta)^2}{2}} - e^{-\frac{(\sqrt{d}-\Delta)^2}{2}}) \\
&\quad - \sqrt{\frac{2}{\pi}}(\Phi(\sqrt{d}-\Delta) - \Phi(-\sqrt{d}-\Delta) - (\sqrt{d}-\Delta)e^{-\frac{(\sqrt{d}-\Delta)^2}{2}} - (\sqrt{d}+\Delta)e^{-\frac{(\sqrt{d}+\Delta)^2}{2}}))\Big].
\end{aligned}
$$

After some algebraic manipulation, we have

$$
\begin{aligned}
\mathrm{ADMSE}(\hat{f}_n^{S+}(x)) &= \sigma^2(x)\Big[1 + d^2(1 - H_1(d;\Delta^2)) + H_3(d;\Delta^2) + \Delta^2 H_5(d;\Delta^2) \\
&\quad - \sqrt{\frac{2}{\pi}}(\Phi(\sqrt{d}-\Delta) - \Phi(-\sqrt{d}-\Delta) \\
&\quad + e^{-\frac{(\sqrt{d}+\Delta)^2}{2}}(d - \sqrt{d} + 2\Delta) + e^{-\frac{(\sqrt{d}-\Delta)^2}{2}}(d - \sqrt{d}))\Big].
\end{aligned}
$$

□

Based on the results of Theorems 3.1 - 3.3 and also the fact that ADMSE($\hat{f}(x)$) = $\sigma^2(x)$, using Theorem 2.1, the asymptotic efficiency of the estimators $\hat{f}^{PT}(x)$,$\hat{f}^{S}(x)$ and $\hat{f}^{S+}(x)$ relative to $\hat{f}(x)$ can be computed as

$$
\begin{aligned}
\mathrm{ARE}[\hat{f}^{PT}(x);\hat{f}(x)] &= \frac{\mathrm{ADMSE}(\hat{f}(x))}{\mathrm{ADMSE}(\hat{f}^{PT}(x))} \\
&= [1 - H_3(\chi_1^2(\alpha);\Delta^2) + \Delta^2[(2H_3(\chi_1^2(\alpha);\Delta^2) - H_5(\chi_1^2(\alpha);\Delta^2)]]^{-1},
\end{aligned} \tag{3.4}
$$

and

$$
\mathrm{ARE}[\hat{f}^{S}(x);\hat{f}(x)] = \frac{\mathrm{ADMSE}(\hat{f}(x))}{\mathrm{ADMSE}(\hat{f}^{S}(x))} = [1 + g(\Delta)]^{-1}, \tag{3.5}
$$

respectively, where

$$
g(\Delta) = d^2 - 2d\left(\frac{2}{\pi}\right)^{\frac{1}{2}} e^{-\frac{\Delta^2}{2}}.
$$

And also,

$$\begin{aligned}\text{ARE}[\hat{f^{S+}}(x);\hat{f}(x)] &= \Big[1 + d^2(1 - H_1(d;\Delta^2)) + H_3(d;\Delta^2) + \Delta^2 H_5(d;\Delta^2)\\ &\quad - \sqrt{\frac{2}{\pi}}(\Phi(\sqrt{d} - \Delta) - \Phi(-\sqrt{d} - \Delta)\\ &\quad + e^{-\frac{(\sqrt{d}+\Delta)^2}{2}}(d - \sqrt{d} + 2\Delta) + e^{-\frac{(\sqrt{d}-\Delta)^2}{2}}(d - \sqrt{d}))\Big]^{-1}. \qquad (3.6)\end{aligned}$$

Using Theorem 3.2, the value of d that minimizes ADMSE($\hat{f^S}(x)$) is equal to

$$d^* = \sqrt{\frac{2}{\pi}}\exp\left\{-\frac{\Delta^2}{2}\right\}.$$

Thus, substituting d by d^*, the proposed criterion depends only on Δ^2.

3.1 Comparisons

Now, we summarize some of the main important results proposed in the previous section in order to obtain the optimal estimator. Figure 1 represents the ARE of the PTE for some α values. It is clear that none of $\hat{f}(x)$, $\hat{f^{PT}}(x)$, $\hat{f^S}(x)$ nor $\hat{f^{S+}}(x)$ is the best in general.

Under the null hypothesis $H_0 : f(x) = f_0(x)$, we have $\Delta^2 = 0$, hence

$$\text{ARE}[\hat{f^S}(x);\hat{f}(x)] = \left(1 - \frac{2}{\pi}\right)^{-1} = 2.75 > 1. \qquad (3.7)$$

Therefore, the SSKDE performs better than the ordinary kernel estimator.

On the other hand

$$\text{ARE}[\hat{f^{PT}}(x);\hat{f}(x)] = [1 - H_3(\chi_1^2(\alpha);0)]^{-1} \geq 1, \qquad (3.8)$$

which depends on the level of significance α. Further, as $\Delta^2 \to \infty$, ARE[$\hat{f^S}(x);\hat{f}(x)$] = $[1 + \frac{2}{\pi}]^{-1} = 0.61$, while ARE[$\hat{f^{PT}}(x);\hat{f}(x)$] $\to$ 1. These facts imply that $\hat{f^S}(x)$ is superior to $\hat{f^{PT}}(x)$ when $f(x)$ is close to $f_0(x)$. On the other hand, the minimum guaranteed efficiency of $\hat{f^S}(x)$ relative to $\hat{f}(x)$ is 0.61, and that of $\hat{f^{PT}}(x)$ depending upon α. In general, ARE[$\hat{f^S}(x);\hat{f}(x)$] decreases from $\frac{\pi}{\pi-2}$ at $\Delta^2 = 0$ and crosses 1-line at $\Delta^2 =$

$\ln(4) = 1.38$ then, drops to the minimum value $\frac{\pi}{\pi+2} = 0.61$ at $\Delta^2 \to \infty$. The loss of asymptotic efficiency is $1 - \left\{1 + \frac{2}{\pi}\right\}^{-1} = 0.39$, while the gain in efficiency is 2.75. Thus, for $0 \leq \Delta^2 \leq 1.38$, $\hat{f}^S(x)$ is preferred to $\hat{f}(x)$; otherwise, $\hat{f}(x)$ performs better outside the interval.

By contrast, ARE$[\hat{f}^{PT}(x); \hat{f}(x)]$ has a maximum value $[1 - H_3(\chi_1^2(\alpha); 0)]^{-1}$ at $\Delta^2 = 0$, dropping to a value of one when $\Delta^2 = 1$. It continues to drop, reaching the minimum value of ADMSE, and then increases towards a value of one as $\Delta^2 \to \infty$. From this, one may conclude that the range of Δ^2 for which $\hat{f}^S(x)$ is better than $\hat{f}(x)$ is wider than the range produced by $\hat{f}^{PT}(x)$. Further, $\hat{f}^S(x)$ is independent of α, while the minimum of the ARE$[\hat{f}^{PT}(x); \hat{f}(x)]$ depends on the value of α. In general, $\hat{f}^S(x)$ does not dominate $\hat{f}^{PT}(x)$ uniformly except in the range $(0, \ln(4))$. Thus, considering the high asymptotic efficiency of $\hat{f}^S(x)$, and also the fact that it is independent of the size α, of the test, the estimate $\hat{f}^S(x)$ is preferable over $\hat{f}^{PT}(x)$ if $f_0(x)$ is close to $f(x)$.

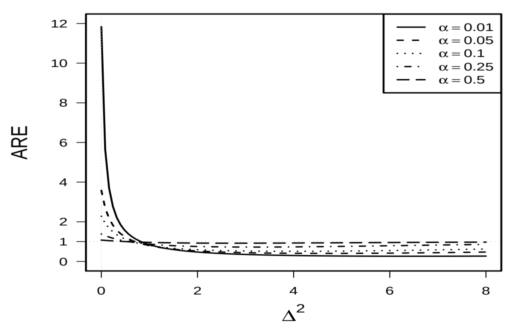

Figure 1: The ARE of the PTE given by (3.4) for some choices of α.

4 Numerical Comparisons

4.1 Simulation Results

Assuming the association property for the r.v.s, we compared the *asymptotic* behavior of the PRSE, PTE and SSKDE based on the closed-form expressions derived in Section 3. We complete the comparison between the proposed estimators for the non-asymptotic states (small/moderate sample sizes). So, a Monte Carlo simulation study is provided in the current section. The algorithm proposed by Cai and Roussas (1998) has been employed for generating the r.v.s having the PA/NA property. More specifically, if $\mathbf{X}_\rho := (X_1, \cdots X_n)$ follows an n-variate normal distribution with mean-vector $\mathbf{0}$ and the correlation-matrix $\mathbf{R} := [\rho_{ij}]$. Therefore, it has the PA property when $\mathbf{R}$'s off-diagonal elements are as $\rho_{ij} = \frac{\rho^{|i-j|}}{1-\rho^2}$, $i \neq j$, $i, j = 1, \cdots, n$, $\rho > 0$, and the NA property when $\rho_{ij} = -\frac{|\rho|^{|i-j|}}{1-\rho^2}$, $i \neq j$, $i, j = 1, \cdots, n$, $\rho < 0$.

Similar to Srihera and Stute (2011), the optimum bandwidth is set to $h = n^{-\frac{1}{5}}$. Five different choices are adopted for the hypothesized pre-specified density $f_0(\cdot)$, as the point mass at zero $f_0 = 0$, standard normal distribution $f_0 = N(0,1)$ (which is the right guess), student's t-distribution with $\nu = 4$ degrees of freedom $f_0 = t_4$, the uniform distribution on $(-\sqrt{3}, \sqrt{3})$, and the standard Cauchy distribution $C(0,1)$. These distributions are termed as null distributions in the sequel.

We compared all estimators with the ordinary kernel estimator (KDE) $\hat{f}(x)$. More specifically, let $\hat{f}_1(x)$ be an estimator for the density function $f(x)$. Then, the RE is defined as

$$\text{RE} := \text{RE}[\hat{f}_1(x), \hat{f}(x)] = \frac{\text{EMISE}(\hat{f}(x), f(x))}{\text{EMISE}(\hat{f}_1(x), f(x))},$$

where EMISE stands for the estimated mean integrated squared error given by

$$\text{EMISE}(\hat{f}_1(x), f(x)) = \frac{1}{M} \sum_{i=1}^{M} \int_{-\infty}^{\infty} \left(\hat{f}_1^{(i)}(x) - f(x)\right)^2 \, dx.$$

Here, M is the number of replications and $\hat{f}_1^{(i)}(x)$ denotes the ith observation of the estimator $\hat{f}_1(x)$.

We conducted a Monte Carlo simulation with $M = 5000$ and $n = 10(10)100, 150, 200$. The RE results obtained from the simulation are depicted in Figures 2-7. From these figures, we observe the following points:

- The REs are almost decreasing functions of the sample size n, even if the null distribution deviates from the true normal distribution. However, for the SSKDE, the corresponding RE of the best guesses is about constant.

- For appropriate f_0, the proposed improved estimators behave better than the KDE, in the RE sense. Otherwise, they have better performance just for small sample sizes.

- Considering that the data are drawn from a multivariate normal distribution, it is observed that the estimators borrowing $f_0 = N(0,1)$ as the initial guess, outperform the other competitors. Indeed, adopting the uniform, Cauchy, and a point mass at zero as the initial guesses, yield bad estimators for f. In other words, a reasonable guess for f_0 results in a better estimator for f.

- Figure 6 shows the EMISE of the PRSE divided by SSKDE for the case $\rho = 0.3$. Based on this figure and some other visual evidence for other ρ values, that are not peresented here, it is seen that the PRSE outperforms the SSKDE.

Finally, for a brief study of convergence rates of estimators, we sketch the EMISE values for the PRSE, SSKDE as well as the KDE for $\rho = 0.3$ under the null, normal (good guess) and Cauchy (bad guess) distributions for sample sizes $n = 20(20)100, 150, 200, 500$. Figure 7 depicts the results. As it is clearly seen, the rate of convergence of the PRSE is better than the SSKDE, especially when the null distribution is close to the right one. Surprisingly, both estimators converge faster than KDE. This result is reversed for the case of bad guess and moderately/large sample sizes.

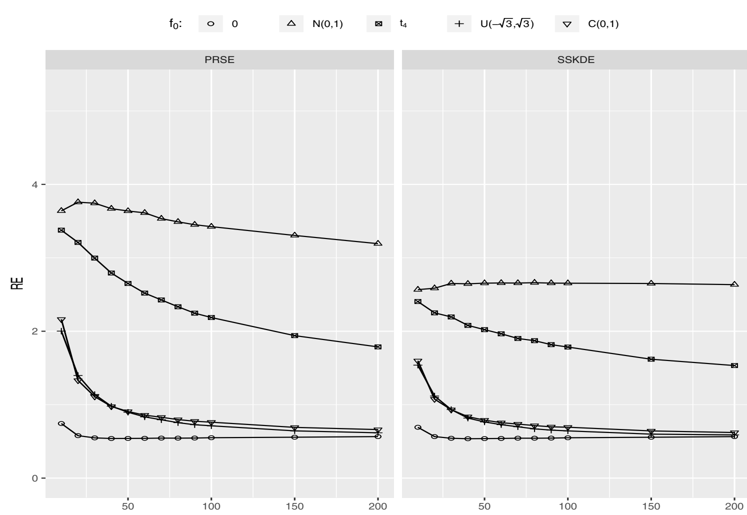

Figure 2: The REs of the improved estimators for $\rho = 0.3$.

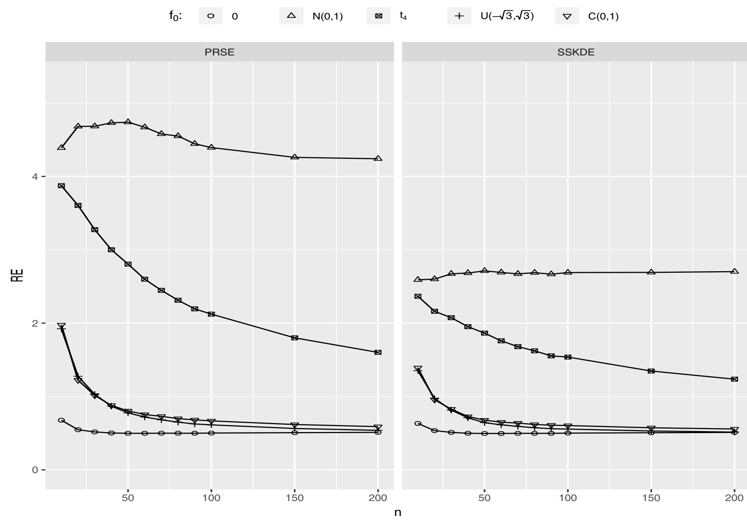

Figure 3: The REs of the improved estimators for $\rho = 0.1$.

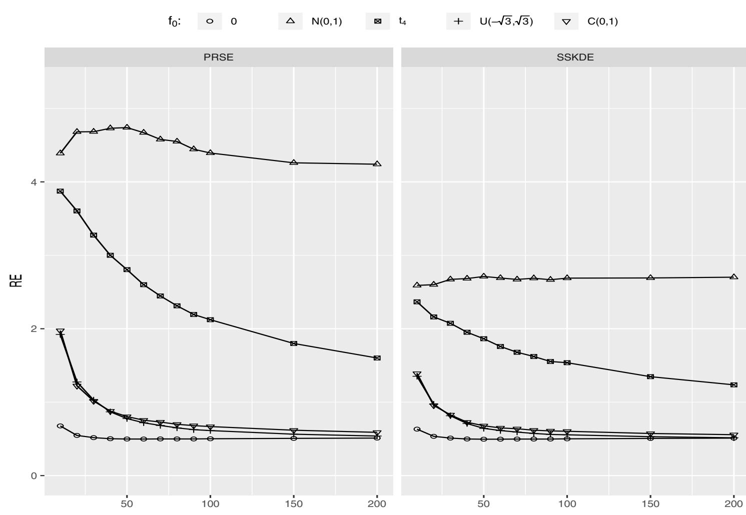

Figure 4: The REs of the improved estimators for $\rho = 0$.

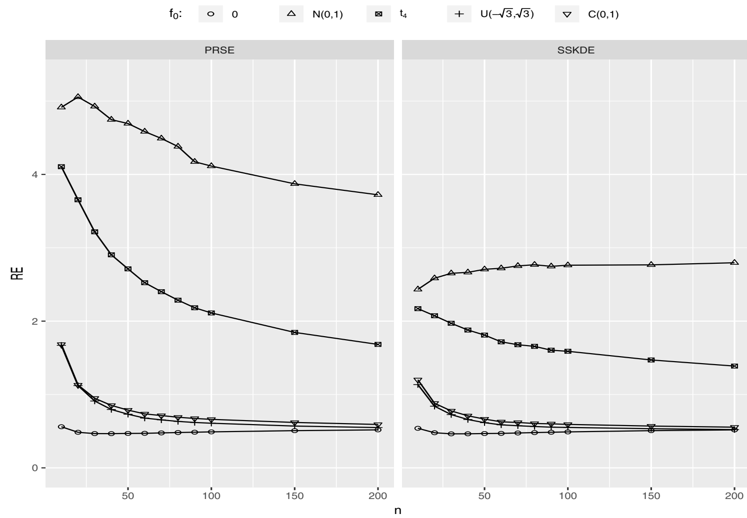

Figure 5: The REs of the improved estimators for $\rho = -0.3$.

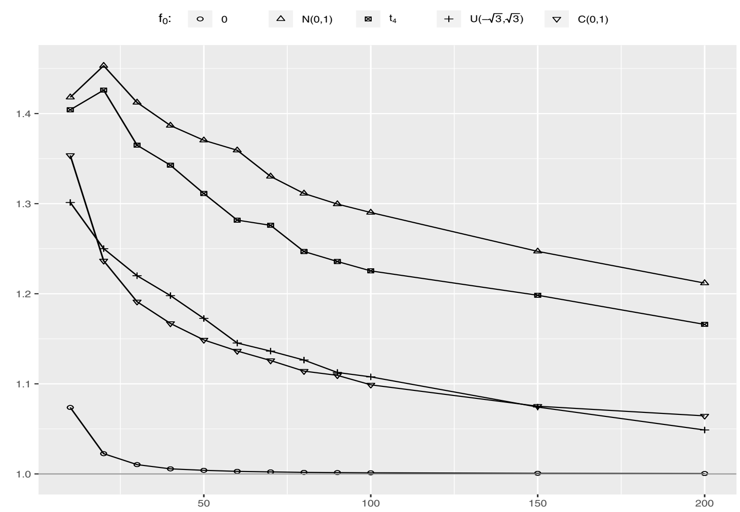

Figure 6: The EMISE of the PRSE divided by the EMISE of the SSKDE for the case $\rho = 0.3$.

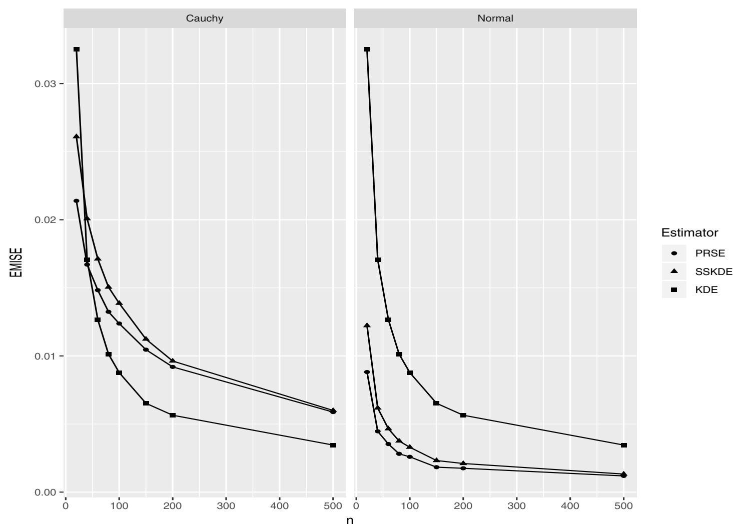

Figure 7: Comparisons of the EMISEs of the KDE with the PRSE and EMISE, under two initial guesses (null distributions) for $\rho = 0.3$.

4.2 Real Data Analysis

In this subsection, we analyze the performance of the proposed kernel density estimators using a real data set for illustrative purposes. The data set represents 300 monthly unemployed females between ages 16 and 19 in the United States from January 1961 to December 1985 in thousands (say Y). The trace plot, auto-correlation function (ACF) as well as the partial ACF (PACF) of the first-order difference of Y's (denoted by X) are displayed in Figure 8.

In order to come up with the correct model, we fitted several moving average (MA) and auto-regressive (AR) models according to the lags in Figure 8. The MA model with order 1, 15, 16 and 23 and an AR model with order 1, 2, 3 and 5 are the candidates for model fitting. For comparing the adequacy of the candidates, we used the corrected Akaike information criterion (AICC) for small sample sizes. If n and p denote the sample size and number of parameters, receptively, then the AICC is given by

$$AICC = AIC + \frac{2p(p+1)}{n-p-1}.$$

We also used the Hannan-Rissanen and Burg methods to fit the models. For more information about the AICC, Hannan-Rissanen, and Burg methods used for model

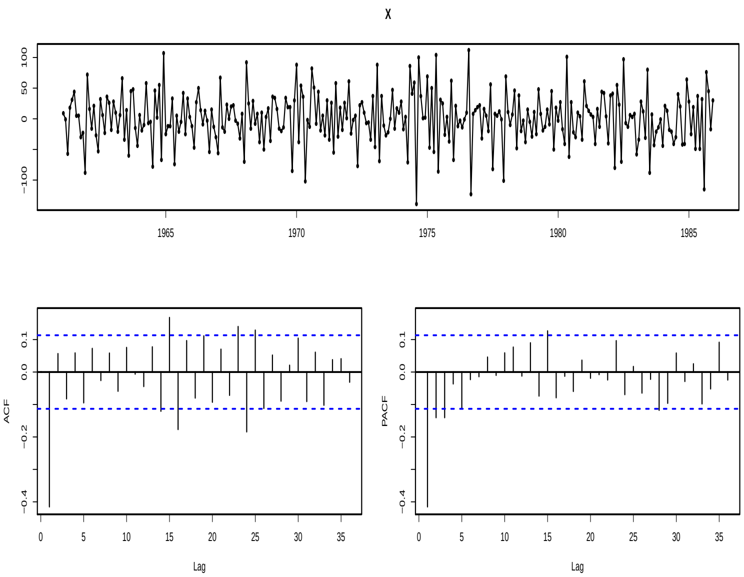

Figure 8: The time series plots.

Table 1: The Results of fitting the MA and AR models to X values using Hannan-Rissanen (HR) and Burg (B) methods along with the AICC values, for the real data.

Model	Method	AICC
MA(1)	HR	3015.92
MA(15)	HR	3039.26
MA(16)	HR	3036.31
MA(23)	HR	3062.71
AR(1)	B	3028.11
AR(2)	B	3024.26
AR(3)	B	3020.32
AR(5)	B	3020.23

fitting, see Brockwell and Davis (2002). The results of this fitting of models are reported in Table 1. According to the results of Table 1, the MA(1) has minimum AICC value and is selected as the underlying model. Further, the parameter estimate of the MA(1) is negative. Thus, as Wang et al. (2018) showed that the MA(1) model with a negative coefficient has NA property, we have a data set that satisfies the assumption of NA.

We estimate the density using the PRSE, SSKDE, and KDE. Indeed, here, the null distribution is not known. However, a simple visual check will indicate some appealing properties of our approach. Hence, as the next step, we test some candidate distributions as $f_0(\cdot)$ for this data set. Table 2 and Figure 9 summarize the results. Based on the results of Table 2 and Figure 9, it is observed that the Skew-t (ST) outperforms the other candidates. Thus, we use the ST, Logistic and Cauchy distributions as the good, moderate, and bad guesses, respectively, for the parent distribution. Next, for analyzing the behavior of the proposed estimators, we have employed a block bootstrap procedure to generate $B = 1000$ observations. The corresponding standard deviations of the PRSE, SSKDE, and KDE, as well as the boxplots, are given in Table 3 and Figure 10, respectively. From these results, we observe that a reasonable guess for the parent distribution of the data produces a more appropriate estimator. However, along with this point, it seems that the PRSE is more sensible in selecting the initial guess than the SSKDE. Another point that can be seen in Table 3 is that although the PRSE outperforms all of the other competitors even with a bad initial guess, the SSKDE works better than $\hat{f}(x)$. Hence, the results obtained here, confirm those of the simulation study.

Table 2: The values of Kolmogorov-Smirnov (KS) statistic, AIC and Bayesian information criterion (BIC) for some candidate distributions fitted on the real data.

	KS statistic	KS p-value	AIC	BIC
Normal	0.038	0.78	3084.50	3091.90
SN	0.032	0.92	3085.30	3096.40
ST	0.027	0.9	3084.81	3099.61
Logistic	0.031	0.94	3080.80	3088.17
Cauchy	0.067	0.13	3163.80	3171.20
Uniform	0.257	$< 2.2e\text{-}16$	...	...

Table 3: The bootstrap estimation of the standard deviations for the proposed estimators.

	PRSE	SSKDE	KDE
ST	0.00031	0.00049	0.00078
Logistic	0.00039	0.00052	0.00078
Cauchy	0.00080	0.00080	0.00078

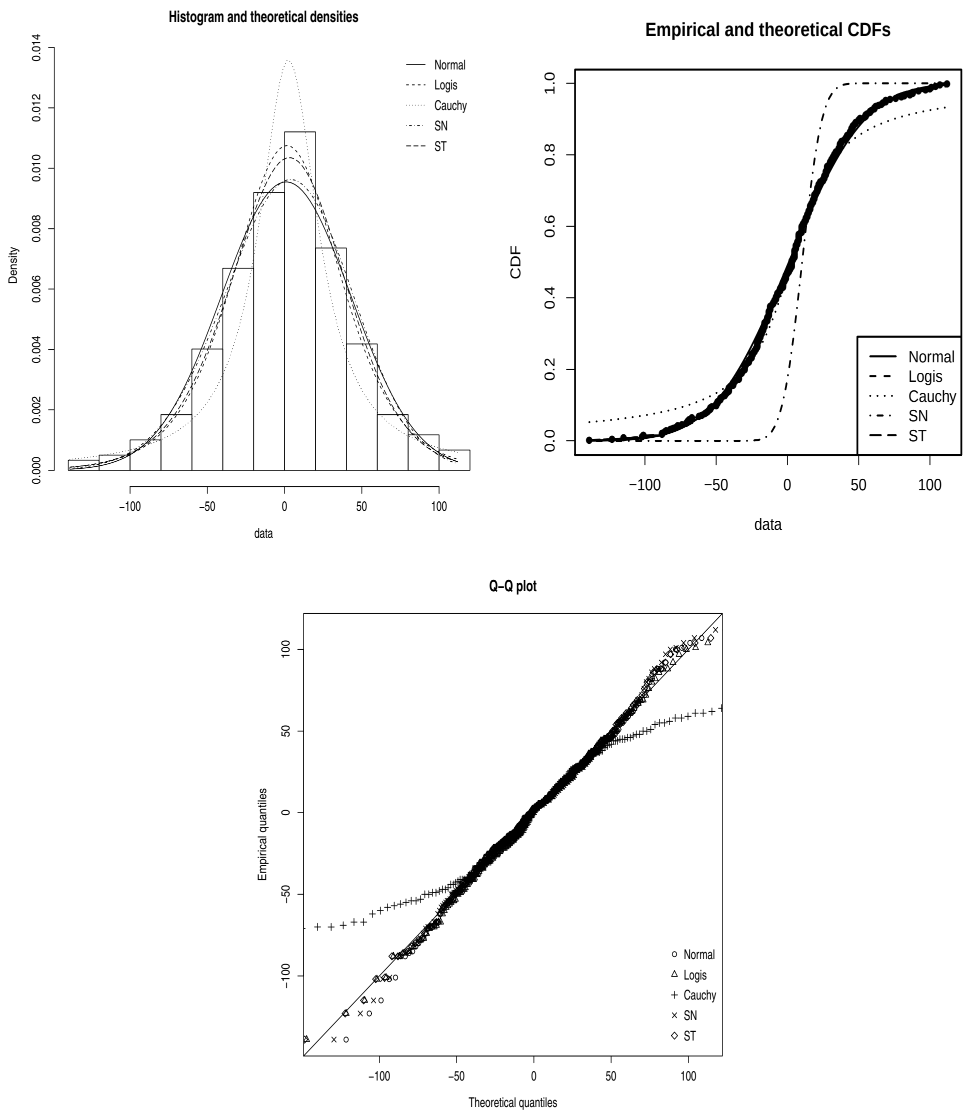

Figure 9: Histogram, empirical cumulative distribution and Q-Q plot of the real data versus the fitted distributions.

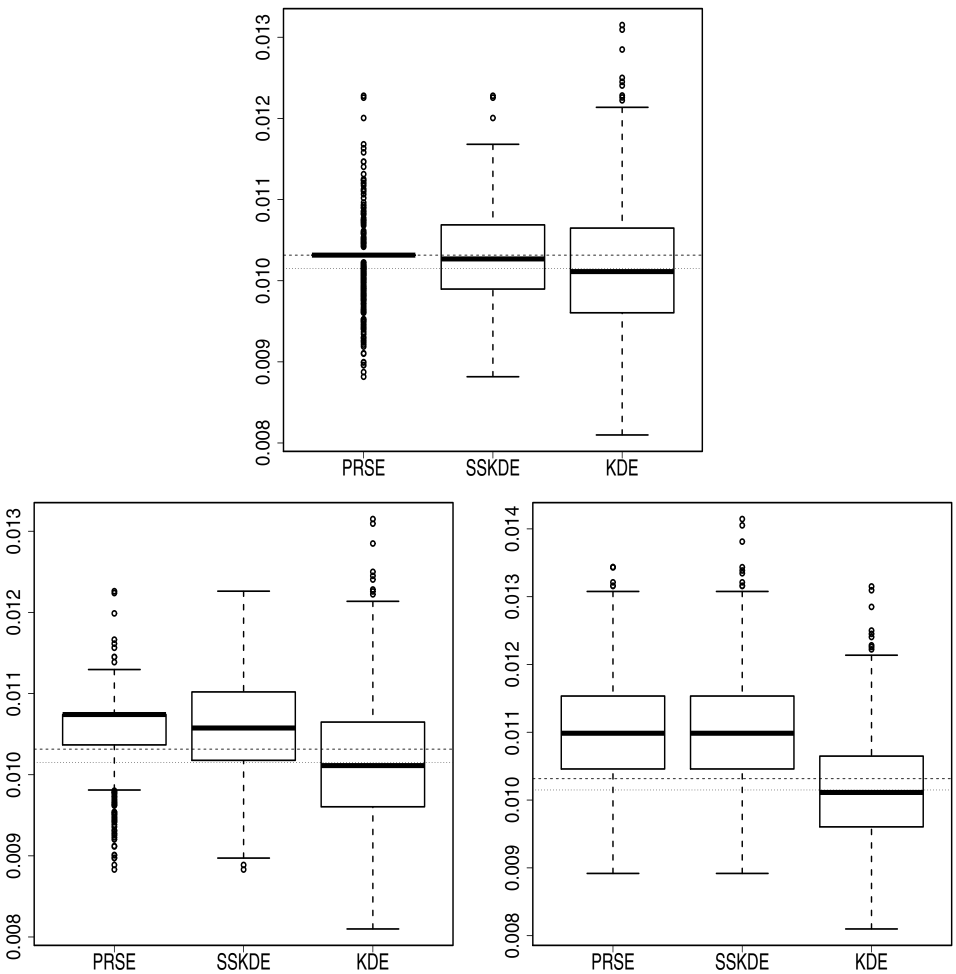

Figure 10: The boxplots of the estimators assuming ST (top), Logistic (bottom left) and Cauchy (bottom right) as the null distributions. The dashed and dotted lines indicate the distribution value lines indicate the distribution value at $x = 0$ based on the ST distribution and KDE itself.

5 Conclusion

In this article, we proposed the preliminary test and Stein-type shrinkage estimators of $f(x)$ for the associated r.v.s, and derived their exact asymptotic distributional characteristics. The result of this article improved the KDE of the marginal probability density function of a strictly stationary sequence of associated random variables via the preliminary test and Stein-type estimators. The results of simulation studies suggested that the improved estimators always perform better than the KDE when we choose an appropriate null distribution ($f_0(\cdot)$). They even have reasonable performances under the bad guesses for the smaller sample sizes. Moreover, results showed that the rate of convergence of the PRSE is better than that of the SSKDE.

Acknowledgments

We would like to thank two anonymous reviewers for their constructive comments which significantly improved the presentation and led to put many details.

References

Arashi, M. and Mahmoudi, M. (2014) A note on shrinkage kernel density estimation, International Statistical Institute Regional Statistical Conference (ISI-RSC), Kuala Lumpur, Malaysia.

Brockwell, P. J. and Davis, R. A. (2002), *Introduction to Time Series and Forecasting*, Springer, New York.

Cai, Z. W. and Roussas, G. G. (1998), Kaplan–Meier estimator under association, *Journal of Multivariate Analysis*, **67**, 318–348.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967), Association of random variables with applications, *Annals of Mathematical Statistics*, **38**, 1466–1474.

Joag–Dev, K. and Proschan, F. (1983), Negative association of random variables with applications, *Annals of Statistics*, **11**, 286–295.

Roussas, G. G. (1999), Positive and negative dependence with some statistical application, In: Ghosh, S. (Ed.), *Asymptotics, Nonparametrics and Time Series*. Marcel Dekker, New York, pp. 757–788.

Roussas, G. G. (2000), Asymptotic normality of the kernel estimate of a probability density function under association, *Statistics and Probability Letters*, **50**, 1–12.

Saleh, A. K. Md. E. (2006), *Theory of Preliminary Test and Stein-type Estimation with Applications*, John Wiley, New York.

Saleh, A. K. Md. E. and Ghania, A. (2016), New estimators of distribution functions, *Communication in Statistics–Theory and Methods*, **45**, 3145–3157.

Saleh, A. K. Md. E., Kibria B. M. G., and George, F. (2018), Simultaneous Estimation of Several CDF's: Homogeneity Constraint, *Communication in Statistics- Theory and Methods*, **47**, 2813–2826.

Shao, J. (2003), *Mathematical Statistics*, 2nd Ed. Springer, New York.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

Srihera, R. and Stute, W. (2011) Kernel adjusted density estimation, *Statistics and Probability Letters*, **81**, 571–579.

Wand, M. P. and Jones, M. C. (1995), *Kernel Smoothing*, Chapman & Hall, London.

Wang, X., Wu, Y., and Hu, SH. (2018), Strong and weak consistency of LS estimators in the EV regression model with negatively superadditive–dependent errors, *AStA Advances in Statistical Analysis*, **102(1)**, 41–65.