REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

Comparative Study on Probability Density Estimators of Sample Maximum and Data Transformation

Authors: Taku Moriyama ✉
– School of Data Science, Yokohama City University,
Yokohama, Japan
moriyama.tak.lu@yokohama-cu.ac.jp
– Department of Management of Social Systems and Civil Engineering, Tottori University,
Tottori, Japan

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- Comparative studies on estimators of the probability density function of sample maximum are conducted. This study presents a plug-in type of and a new block-maxima-based kernel density estimators as the alternatives of the parametric estimator fitting to the approximate generalized extreme value density function. Asymptotic properties of the density estimators are investigated, which shows that the optimal convergence rates depend on the extreme value index of the distribution. Furthermore, this study investigates the density estimators with data log-transformation. It is demonstrated that the log-transformation makes the estimators numerically stable in finite sample case. Finally, two illustrative examples are provided.

Keywords:

- *Extreme value; data transformation; kernel-type estimator; mean squared error; nonparametric estimation; sample maximum.*

AMS Subject Classification:

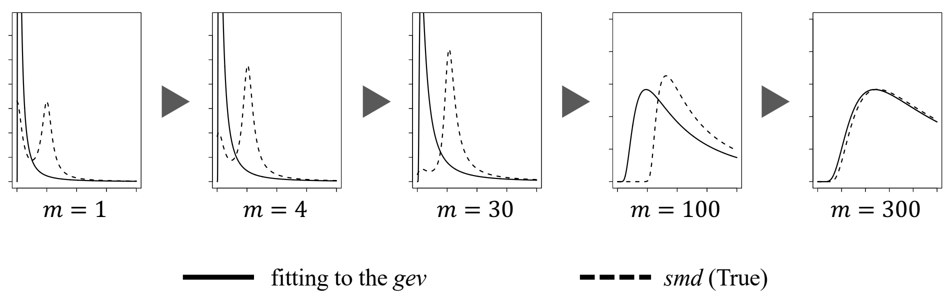
- 62G32, 62G07, 60G70.

✉ Corresponding author

1. INTRODUCTION

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables with a continuous distribution function F, where the density function f exists. This study considers nonparametric density estimation of sample maximum of future m observation $X_{n+1}, X_{n+2}, \cdots, X_{n+m}$. The sample maximum density (SMD) $f_{(m)} := mfF^{m-1}$ of the appropriately normalized random variables possibly converges to GEV. This study first surveys the accuracy of the estimator fitted to the density of the generalized extreme distribution (GEV). The accuracy of the fitting estimator fundamentally depends on the approximation to GEV, and the approximation accuracy can be very poor depending the case. Figure 1 shows an image of the convergence of a bimodal SMD the GEV, where the original density is a mixing density of two t-distributions. The fitting to the GEV with small m is neither bimodal nor good approximation to SMD (True).

Figure 1: Convergence of SMD of a mixing density of two t-distributions to the GEV.



Nonparametric density estimator can capture properties of the underlying distribution (e.g. multi-modality and skewness). We consider some nonparametric approaches for SMD estimation in this study. Nonparametric density estimation itself is useful for visualization, estimating modes, detecting multi-modality, and so on. Beirlant and Devroye (1999) [3] gave the lower bounds of the total variation and the supremum norm in a nonparametric setting. Using the same settings as Moriyama (2021) [19] investigating sample maximum distribution estimation (SMD) specified below, this study obtains the asymptotic mean squared error (MSE) of the estimators at the point x under the supposition $m \to \infty$, $(m/n) \to 0$ and $x \to \infty$ as $n \to \infty$. We investigate the properties of SMD estimators and conduct comparative studies. A comparison study on extreme quantile estimators are conducted by Banfi et al. (2022) [2], and the numerical accuracy of some estimators are reported.

Explicit forms are required in order to compare the asymptotic errors of different estimators, so the tail of F needs to be specified. As shown later, parameters up to the so-called second order have an effect on the first order error of the extreme-based SMD estimator. Let F belong to the Hall class (i)$\alpha > 0, \beta \geq 2^{-1}, A > 0, B \neq 0$ and

$$x^{\alpha+\beta}\{1 - F(x) - Ax^{-\alpha}(1 + Bx^{-\beta})\} \to 0 \quad \text{as} \quad x \to \infty$$

(see Hall and Welsh 1984 [12]), the Weibull class (ii) $\kappa > 0, C > 0$ and

$$\exp(Cx^{\kappa})\{1 - F(x) - \exp(-Cx^{\kappa})\} \to 0 \quad \text{as} \quad x \to \infty,$$

or the bounded class (iii) $\mu < -2, \sigma \leq -2^{-1}, D > 0, E \neq 0, x^* \in \mathbb{R}$ and

$$(x^* - x)^{\mu+\sigma}\{1 - F(x) - (x^* - x)^{-\mu}(D + E(x^* - x)^{-\sigma})\} \to 0 \quad \text{as} \quad x \uparrow x^*.$$

(e.g., Stupfler 2016 [23]). These three classes of distributions are widely utilized as representatives of the distributions belonging to the domain of attractions of the extreme value distribution (see Beirlant et al. 2004 [4]; Segers 2005 [20]).

This study proposes a plug-in type of kernel estimator for SMD defined as $\widehat{f}_{(m)} := m\widehat{f}\widehat{F}^{m-1}$, where $\widehat{f}$ and $\widehat{F}$ are the kernel density and kernel distribution estimators of f and F, respectively. Another candidate for the SMD estimator in this study is the new block-maxima-based (BM-based) kernel estimator $\bar{f}_{(m)}$. This BM-based estimator is the kernel density estimator of SMD itself, which is a function of the block maxima. As seen from the definition, the nonparametric estimators are consistent for wide class of distributions as long as m is fixed. The nonparametric estimators are thus a promising candidate for the SMD, at least for relatively small m.

The aim of this study is to investigate the difference between the 'parametric' and nonparametric approaches in SMD estimation, where the parametric' estimator (PE) is the estimator fitted to GEV. Sections 2, 3, and 4 of this paper describe the asymptotic properties of the fitting estimator, the plug-in kernel SMD estimator $\widehat{f}_{(m)}$ and the BM-based kernel SMD estimator $\bar{f}_{(m)}$, respectively. We then prove that the optimal convergence rates of the two nonparametric estimators are different and that the plug-in estimator $\widehat{f}_{(m)}$ outperforms the BM-based estimator as long as the condition of achieving the optimal convergence rate is satisfied. In Section 5, we compare the numerical accuracies of PE and the plug-in estimator. It is demonstrated that the extreme value index severely affects both the theoretical and numerical performances in this study.

The density estimators with data log-transformation are discussed in Section 6, where the log-transformation changes the non-negative extreme indices of the distributions to zero. Section 7 provides these numerical properties, and we see that the log-transformation makes the estimators numerically much stable. Two illustrative examples of application are provided in Section 8. The proofs are given in the Supplementary file.

2. Parametric density estimation of sample maximum

It follows from the Fisher-Tippett-Gnedenko theorem that

$$F^m(x) - G_\gamma\left(\frac{x - b_m}{a_m}\right) \to 0 \quad \text{as} \quad m \to \infty \quad for \quad 1 + \frac{x - b_m}{a_m} > 0,$$

where G_γ is the generalized extreme value distribution,

$$\gamma := \begin{cases} \alpha^{-1} & \text{for} \quad \text{(i)} \\ 0 & \text{for} \quad \text{(ii)} \\ \mu^{-1} & \text{for} \quad \text{(iii)}, \end{cases}$$

$$a_m := \begin{cases} \gamma(Am)^\gamma & \text{for} \quad \text{(i)} \\ \kappa^{-1}C^{-1/\kappa}(\ln m)^{-\theta} & \text{for} \quad \text{(ii)} \\ -\gamma(Dm)^\gamma & \text{for} \quad \text{(iii)}, \end{cases} \qquad b_m := \begin{cases} (Am)^\gamma & \text{for} \quad \text{(i)} \\ (C^{-1}\ln m)^{1/\kappa} & \text{for} \quad \text{(ii)} \\ x^* - (Dm)^\gamma & \text{for} \quad \text{(iii)}, \end{cases}$$

and $\theta := 1 - (1/\kappa)$. The setting in this study follows that the extreme value index $\gamma > -1/2$. This study employs the maximum likelihood based on the BM method in the parameter $\boldsymbol{\gamma}_m := (\gamma, a_m, b_m)$ estimation, where the block size is k and $^\exists N \in \mathbb{N}$ s.t. $n = N \times k$. k can be different from m but needs to satisfy the following assumptions.

Assumption 2.1. One of $(M_n \vee K_n) \to 0$, $(M_n \wedge K_n) \to \infty$, or $^\exists \delta > 0$ s.t. $M_n \to \delta$ and $K_n \to \delta$ holds, where

$$M_n := \begin{cases} Amx^{-\alpha} & \text{for (i)} \\ m\exp(-Cx^{\kappa}) & \text{for (ii)} \\ Dm(x^* - x)^{-\mu} & \text{for (iii),} \end{cases} \qquad K_n := \begin{cases} Akx^{-\alpha} & \text{for (i)} \\ k^{\kappa}\exp(-\kappa C^{1/\kappa}(\ln k)^{\theta} x) & \text{for (ii)} \\ Dk(x^* - x)^{-\mu} & \text{for (iii).} \end{cases}$$

M_n and K_n have the following meanings $f_{(m)}(x) \sim \alpha x^{-1} M_n \exp(-M_n)$ and $g_{\boldsymbol{\gamma}_k}(x) \sim \alpha x^{-1} K_n \exp(-K_n)$ respectively. Assumption 2.1 means the block size k is asymptotically same as m. The following proposition on the convergence of a bias term of the fitting estimator PE requires the assumption.

Proposition 2.1. *Under Assumption 2.1,*

$$\widetilde{\tau}_n := f_{(m)}(x) - g_{\boldsymbol{\gamma}_k}(x) \to 0,$$

where

$$\begin{aligned} g_{\boldsymbol{\gamma}_k}(x) :=& \frac{1}{a_k} g_{\gamma}\left(\frac{x - b_k}{a_k}\right), \\ g_{\gamma}(x) :=& \begin{cases} \{1 + \gamma x\}^{-(1/\gamma)-1} \exp(-\{1 + \gamma x\}^{-1/\gamma}), & 1 + \gamma x > 0, \quad \textit{for} \quad \text{(i), (iii)} \\ \exp(-x)\exp(-\exp(-x)), & x \in \mathbb{R}, \quad \textit{for} \quad \text{(ii).} \end{cases} \end{aligned}$$

Set the MLE $\widehat{\boldsymbol{\gamma}}_k$. The following Assumptions 2.2 and 2.3 ensures the convergence of $\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k$ and $g_{\widehat{\boldsymbol{\gamma}}_k}(x) - g_{\boldsymbol{\gamma}_k}(x)$; however, we note that both Assumptions 2.1 and 2.2 do not hold for the Weibull class.

Assumption 2.2. $^\exists \lambda \in \mathbb{R}$ s.t. $\lambda_n \to \lambda$, where

$$\lambda_n := \begin{cases} km^{-2\beta} & \text{for} \quad \text{(i)} \\ k(\ln m)^{-2} & \text{for} \quad \text{(ii)} \\ km^{2\sigma} & \text{for} \quad \text{(iii).} \end{cases}$$

Regarding the MSE of PE $g_{\widehat{\boldsymbol{\gamma}}_k}(x)$, the following theorem holds.

Theorem 2.1. *Under Assumptions 2.1 and 2.2,*

$$\mathbb{E}[(f_{(m)}(x) - g_{\widehat{\boldsymbol{\gamma}}_k}(x))^2] \sim (\widetilde{\tau}_n - N^{-1/2} k^{-\gamma} \lambda_n \widetilde{\boldsymbol{\eta}}_n^{\mathsf{T}} I_0^{-1} \boldsymbol{b})^2 + N^{-1} k^{-2\gamma} (\widetilde{\boldsymbol{\eta}}_n^{\mathsf{T}} I_0^{-1} \widetilde{\boldsymbol{\eta}}_n),$$

where $\widetilde{\boldsymbol{\eta}}_n := (\widetilde{s}_n, \widetilde{t}_n, \widetilde{u}_n)^{\mathsf{T}}$,

$$\begin{aligned} \widetilde{s}_n &:= K_n^{1+\gamma} \exp(-K_n) \left\{(1 - K_n)(1 - K_n^{\gamma} + \gamma \ln K_n) + \gamma\,(1 - K_n^{\gamma})\right\}, \\ \widetilde{t}_n &:= K_n^{1+\gamma} \exp(-K_n)(K_n - 1)\left(\frac{K_n^{\gamma} - 1}{\gamma}\right), \\ \widetilde{u}_n &:= K_n^{1+\gamma} \exp(-K_n) K_n^{\gamma} (1 + \gamma - K_n). \end{aligned}$$

b *and the Fisher information matrix I_0 are given in Dombry and Ferreira (2019) [11]. Furthermore,*

$$N^{1/2}k^{\gamma}(\widetilde{\boldsymbol{\eta}}_n^{\mathsf{T}} I_0^{-1}\widetilde{\boldsymbol{\eta}}_n)^{-1/2}\{f_{(m)}(x) - g_{\widehat{\boldsymbol{\gamma}}_k}(x) - (\widetilde{\tau}_n - N^{-1/2}k^{-\gamma}\lambda_n\widetilde{\boldsymbol{\eta}}_n^{\mathsf{T}} I_0^{-1}\boldsymbol{b})\}$$

converges in distribution to the standard normal distribution, if the MSE converges to zero.

The symbol $\sim$ means $p_n \sim q_n \iff p_n/q_n = 1 + o(1)$. This paper utilizes the same symbols as Moriyama (2021) [19] for clarity. We add a tilde to the symbols but are actually different. The following corollary on the convergence rate follows from Theorem 2.1. Proposition 2.1 and Theorem 2.1 are proved in the Supplementary file, and Corollary 2.1 is a direct consequence of Theorem 2.1.

Corollary 2.1. *Under the assumptions of Theorem 2.1, $(f_{(m)}(x) - g_{\widehat{\boldsymbol{\gamma}}_k}(x))$ converges with the rate the larger of $f_{(m)}(x) - g_{\boldsymbol{\gamma}_k}(x)$ and*

$$N^{-1/2}k^{-\gamma}K_n^{1+\gamma}\exp(-K_n) \times \begin{cases} \lambda_n(K_n^{\gamma} + \ln K_n) + \ln K_n & \textit{for} \quad (M_n \vee K_n) \to 0 \\ \lambda_n + 1 & \textit{for} \quad M_n \to \delta,\ K_n \to \delta \\ \lambda_n K_n(K_n^{\gamma} + \ln K_n) + K_n^{1+\gamma} & \textit{for} \quad (M_n \wedge K_n) \to \infty. \end{cases}$$

The difference of the SMD estimator and the GEV is decomposed into the three bias terms λ_n, $g_{\boldsymbol{\gamma}_m}(x) - g_{\boldsymbol{\gamma}_k}(x)$ and $f_{(m)}(x) - g_{\boldsymbol{\gamma}_m}(x)$. λ_n corresponds to the asymptotic bias of $\widehat{\boldsymbol{\gamma}}_k$, which depends on the second order parameter. The bias $g_{\boldsymbol{\gamma}_m}(x) - g_{\boldsymbol{\gamma}_k}(x)$ comes from the asymptotic difference between m and k. Although a large k yields a large variance of $\widehat{\boldsymbol{\gamma}}_k$, the precise approximation of $g_{\boldsymbol{\gamma}_m}(x)$ to $f_{(m)}(x)$ (i.e., $\widetilde{\tau}_n$) requires a large m. As the extreme value index tends to be zero, i.e., $\gamma \to 0$, the convergence rate of the approximation of $g_{\boldsymbol{\gamma}_m}(x)$ to $f_{(m)}(x)$ becomes slow. In the Weibull cases, the rate is slower than any polynomial. To sum up, the convergence rate is sensitive to k through m and the first order parameter γ. The convergence rates of the whole of PE in some example cases are given later in Table 1 in Section 5.

3. Nonparametric plug-in type of SMD estimation

Applying the plug-in rule, we obtain the following kernel-type nonparametric estimator (NE1):

$$\widehat{f}_{(m)}(x; h_1, h_2) := m\widehat{f}(x; h_1)\widehat{F}^{m-1}(x; h_2),$$

where $\widehat{f}$ and $\widehat{F}$ are the kernel estimators given by

$$\widehat{f}(x; h_1) = \frac{1}{nh_1}\sum_{i=1}^{n} w\left(\frac{x - X_i}{h_1}\right), \quad \widehat{F}(x; h_2) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x - X_i}{h_2}\right).$$

Assumption 3.1. w is a symmetric and bounded density function, W is the cumulative distribution function, and h_j is bandwidth that satisfies $h_j \to 0$ for $j = 1, 2$. For the bounded class (iii), the support of w be bounded and $h_j(x^* - x)^{-1} \to 0$ for $j = 1, 2$ (it is necessary to avoid the so-called boundary bias problem).

Assumption 3.2. f is twice continuously differentiable at x.

The Hall class and the Weibull class satisfy Assumption 3.2, which requires $\mu \leq -2$ or both $\mu = -1$ and $\sigma \leq -1$ of the bounded class.

Assumption 3.3.

$$\int z^2 w(z)\mathrm{d}z < \infty, \quad \int w^2(z)\mathrm{d}z < \infty, \quad \int zW(z)w(z)\mathrm{d}z < \infty.$$

Using the asymptotic normalities of the estimators, we have the following theorem.

Theorem 3.1. *Given Assumptions 3.1–3.3, then*

$$\mathbb{E}[(f_{(m)}(x) - \widehat{f}_{(m)}(x; h_1, h_2))^2] \sim b_{n,1}^2 + v_{n,1},$$

where

$$\begin{aligned}
b_{n,1} :=& \frac{h_1^2}{2} \exp(-M_n) M_n^{1+3\gamma} m^{-3\gamma} \psi_n \int z^2 w(z)\mathrm{d}z + \frac{h_2^2}{2} \exp(-M_n) M_n m \xi_n f(x) \int z^2 w(z)\mathrm{d}z, \\
v_{n,1} :=& \frac{m^2}{nh_1} \exp(-2M_n) f(x) \int w^2(z)\mathrm{d}z + \frac{m^3}{n} \exp(-2M_n) M_n f^2(x), \\
\psi_n :=& \begin{cases} \alpha(\alpha+1)(\alpha+2)A^{-3\gamma} & \text{for} \quad \text{(i)} \\ \kappa^3 C^3 x^{3\kappa-3} & \text{for} \quad \text{(ii)} \\ -\mu(\mu+1)(\mu+2)D^{-3\gamma} & \text{for} \quad \text{(iii)}, \end{cases} \\
\xi_n :=& \begin{cases} \alpha(\alpha+1)x^{-2} & \text{for} \quad \text{(i)} \\ \kappa^2 C^2 x^{2\kappa-2} & \text{for} \quad \text{(ii)} \\ \mu(\mu+1)(x^* - x)^{-2} & \text{for} \quad \text{(iii)}, \end{cases} \\
\omega_n :=& \begin{cases} A^{-1}\alpha x^{\alpha-1} & \text{for} \quad \text{(i)} \\ \kappa C x^{\kappa-1} \exp(C x^\kappa) & \text{for} \quad \text{(ii)} \\ -D^{-1}\mu(x^* - x)^{\mu-1} & \text{for} \quad \text{(iii)}. \end{cases}
\end{aligned}$$

Furthermore, $v_{n,1}^{-1/2}\{f_{(m)}(x) - \widehat{f}_{(m)}(x; h_1, h_2) + b_{n,1}\}$ *converges in distribution to the standard normal distribution, if the MSE converges to zero.*

The proof of Theorem 3.1 is provided in the Supplementary file, which shows there exists $p_{m,n}(x)$ and $q_{m,n}(x)$ s.t. $f_{(m)}(x) - \widehat{f}_{(m)}(x; h_1, h_2)$ is asymptotically given by

$$p_{m,n}(x)(f(x) - \widehat{f}(x; h_1)) + q_{m,n}(x)(F(x) - \widehat{F}(x; h_2)).$$

Hence, asymptotically optimal bandwidths of NE1 are those of $\widehat{f}(x; h_1)$ and $\widehat{F}(x; h_2)$.

Corollary 3.1. *An asymptotically optimal value of* (h_1, h_2) *is given by*

$$\left(M_n^{-2-6\gamma} m^{2+6\gamma} \psi_n^{-2} f(x) n^{-1} \left(\int z^2 w(z)\mathrm{d}z \right)^{-2} \int w^2(z)\mathrm{d}z \right)^{1/5},$$

$$\left(2\xi_n^{-2}\omega_n n^{-1}\left(\int z^2 w(z)\mathrm{d}z\right)^{-2}\int zW(z)w(z)\mathrm{d}z\right)^{1/3},$$

which are asymptotically identical to those of $\widehat{f}(x;h_1)$ and $\widehat{F}(x;h_2)$ if both of the optimal values converge 0. Under the assumption of Theorem 3.1, $f_{(m)}(x) - \widehat{f}_{(m)}(x;h_1,h_2)$ with the optimal bandwidths is asymptotically non-degenerate normal with the following asymptotic mean:

$$\nu_1\exp(-M_n)M_n^{(1+3\gamma)/5}m^{(4-3\gamma)/5}\psi_n^{1/5}f^{2/5}(x)n^{-2/5} + \nu_2\exp(-M_n)M_n m\xi_n^{-1/3}f(x)\omega_n^{2/3}n^{-2/3},$$

where

$$\nu_1 := \frac{1}{2}\left(\int z^2 w(z)\mathrm{d}z\right)^{1/5}\left(\int w^2(z)\mathrm{d}z\right)^{2/5},$$
$$\nu_2 := \left(2\int z^2 w(z)\mathrm{d}z\right)^{-1/3}\left(\int zW(z)w(z)\mathrm{d}z\right)^{2/3}.$$

Corollary 3.2. *The asymptotically optimal bandwidths for $M_n = O(1)$ are of the order*

$$\left(\frac{m}{n}\right)^{1/5} \times \begin{cases} m^{\gamma} & \textit{for} \quad \text{(i), (ii)} \\ \{C^{-1}\ln m\}^{-\theta} & \textit{for} \quad \text{(ii),} \end{cases}$$

and

$$\left(\frac{m}{n}\right)^{1/3} \times \begin{cases} m^{\gamma} & \textit{for} \quad \text{(i), (iii)} \\ \{C^{-1}\ln m\}^{-\theta} & \textit{for} \quad \text{(ii).} \end{cases}$$

$\widehat{f}_{(m)}(x;h_1,h_2)$ for $M_n = O(1)$ with the optimal bandwidths has the following asymptotic bias of the order:

$$\left\{\left(\frac{m}{n}\right)^{2/5} + \left(\frac{m}{n}\right)^{2/3}\right\} \times \begin{cases} m^{-\gamma} & \textit{for} \quad \text{(i), (iii)} \\ (\ln m)^{\theta} & \textit{for} \quad \text{(ii).} \end{cases}$$

Corollary 3.2 states that the requirement for the bounded class $h_j(x^* - x)^{-1} \to 0$ for $j = 1, 2$ means $m = o(n)$ for $M_n = O(1)$. Corollary 3.2 is directly obtained from Theorem 3.1 and Corollary 3.1.

Ideally, the density estimator itself should be a density function. NE1 satisfies non-negativity but its integral does not equal those if the bandwidth values are different, i.e. $h_1 \neq h_2$. Under the assumption of $h_1 = h_2$ we cannot obtain the algebraic solution.

The nonparametric density estimator presented in the next section has only one hyper-parameter and always satisfies the requirement that the integral equals one.

4. Nonparametric block-maxima-based estimation of SMD

Suppose that the block size is m and $^{\exists}\bar{n} \in \mathbb{N}$ s.t. $n = \bar{n} \times m$. We present here a BM-based kernel-type of nonparametric estimator (NE2),

$$\bar{f}_{(m)}(x;h) = \frac{1}{\bar{n}h}\sum_{i=1}^{\bar{n}} w\left(\frac{x - Y_i}{h}\right),$$

where $Y_j := \max\{X_{m(j-1)+1}, X_{m(j-1)+2}, \cdots, X_{mj}\}$ $(j = 1, \cdots, \bar{n})$.

Assumption 4.1. The kernel function w is a symmetric and bounded density function. h is a bandwidth that satisfies $h \to 0$. The support of w be bounded and that $h(x^* - x)^{-1} \to 0$ for (iii) (it is necessary to avoid the so-called boundary bias problem).

The following theorem on the asymptotic normality of the naive nonparametric density estimator holds.

Theorem 4.1. *Given Assumptions 3.2, 3.3 and 4.1, if $x^{\kappa-1}h \to 0$,*

$$\mathbb{E}[(f_{(m)}(x) - \bar{f}_{(m)}(x;h))^2] \sim b_{n,2}^2 + v_{n,2},$$

where

$$\begin{aligned} b_{n,2} :=& \frac{mh^2}{2} f(x) \exp(-M_n)\phi_n \int z^2 w(z)\,\mathrm{d}z, \\ v_{n,2} :=& (\bar{n}h)^{-1} f_{(m)}(x) \int w^2(z)\mathrm{d}z, \\ \phi_n :=& \begin{cases} m^2 & \text{for} \quad \text{(i), (iii)} \\ (\kappa C)^2 x^{2(\kappa-1)} - 2m\kappa C x^{\kappa-1} + m^2 & \text{for} \quad \text{(ii).} \end{cases} \end{aligned}$$

Furthermore, $v_{n,2}^{-1/2}\{f_{(m)}(x) - \bar{f}_{(m)}(x;h) + b_{n,2}\}$ converges in distribution to the standard normal distribution, if the MSE converges to zero.

Corollary 4.1. *The asymptotically optimal value of h is given by*

$$\left(n^{-1}\{f(x)\}^{-1}\phi_n^{-2}\left(\int z^2 w(z)\mathrm{d}z\right)^{-2}\int w^2(z)\mathrm{d}z\right)^{1/5}$$

if both of the optimal values converge 0. Under the assumption of Theorem 4.1, $f_{(m)}(x) - \bar{f}_{(m)}(x;h)$ with the optimal bandwidths is asymptotically non-degenerate normal with the following asymptotic mean:

$$\nu_1 n^{-2/5} m \{f(x)\}^{3/5} \phi_n^{1/5}.$$

Corollary 4.2. *The asymptotically optimal bandwidth for $M_n = O(1)$ is of the order*

$$n^{-1/5} m^{-3/5} \times \begin{cases} m^{\gamma/5} & \text{for} \quad \text{(i), (iii)} \\ (\ln m)^{-1/5} & \text{for} \quad \text{(ii).} \end{cases}$$

Under the assumption of Theorem 4.1, $\bar{f}_{(m)}(x;h)$ has the asymptotic bias of the order

$$\left(\frac{m^2}{n}\right)^{2/5} \begin{cases} m^{-(3/5)\gamma} & \text{for} \quad \text{(i), (iii)} \\ (\ln m)^{-(3/5)\gamma} & \text{for} \quad \text{(ii).} \end{cases}$$

Corollary 4.2 states that the requirement for the bounded class $h(x^* - x)^{-1} \to 0$ means $m = o(n^{-1/(3+4\gamma)})$ for $M_n = O(1)$.

Theorem 4.1 is proved in the Supplementary file. Corollaries 4.1 and 4.2 are the direct consequences.

5. Comparative study on SMD estimators

In this section, we compare the performances of three estimators: PE, NE1, and NE2. The first comparison is theoretical and the second is numerical. $M_n \equiv K_n \equiv \delta > 0$ is assumed throughout this section. Each of the MSEs of NE1 and NE2 with their respective optimal bandwidths converge on the order of $m^{-2\gamma}(m/n)^{4/5}$ and $m^{-(6/5)\gamma}(m^2/n)^{4/5}$, respectively. The following theorem is immediately evident from Corollaries 3.2 and 4.2.

Theorem 5.1. *Suppose that each of the optimal bandwidths converge to zero. Then, the convergence rate of NE1 $\widehat{f}_{(m)}(x; h_1, h_2)$ with the optimal bandwidth is faster than NE2 $\bar{f}_{(m)}(x; h)$ with the optimal bandwidth for $\gamma > -1$.*

The MSE of PE converges with the rate $m^{-2\gamma} \times (N^{-1}m^{2-4\beta} + m^{-2\gamma\beta} + m^{-2} + N^{-1})$ for the Hall class. For the Weibull class, PE does not satisfy both Assumptions 2.1 and 2.2 with any block size k and is inconsistent. For the bounded class, the convergence rate is $m^{-2\gamma} \times (N^{-1}m^{2+4\sigma} + m^{-2\gamma\sigma} + m^{-2} + N^{-1})$. Assumption 2.3 requires $\gamma > -3/2$, $\gamma > -1/2$, and $\gamma > -1/6$ for $m = n^{1/4}$, $m = n^{1/2}$, and $m = n^{3/4}$, respectively.

The parameters of the distributions and the convergence rates of MSE of the estimators without terms slower than any polynomial are summarized in Table 1, where the hyphen indicates that the distribution breaks the assumption of the theorem. All the target values $f_{(m)}(x)$ tend toward zero with the exception of the bounded class with $\mu = -1$ while $F^m(x) \to \exp(-\delta) > 0$ in every case. The convergence rates divided by the order of $f_{(m)}(x)$ are given in Table 2 (e.g. $(f_{(m)}(x))^{-2} \times \mathbb{E}[(f_{(m)}(x) - g_{\widehat{\gamma}_k}(x))^2]$ for PE). The hyphen in Table 2 changed from a value in Table 1 means that the density estimator becomes inconsistent by dividing $f_{(m)}(x)$. All the convergence rates of NE2 are slower than or the same as those of NE1 in Table 2, as Theorem 5.1 stated.

Comparing NE1 with PE, we see that the rates of PE become especially slow and that NE1 becomes better as γ gets small. For $m = n^{1/4}$, all the rates of NE1 are faster than PE, but the differences become small as m gets large, which basically coincides with the results of the SMD estimation reported in Moriyama (2021) [19]. For $m = n^{3/4}$, where the convergence rates of the parametric estimator and the nonparametric estimator of SMD are almost the same and $n^{-1/4}$, the convergence rates of the PE and NE1 of SMD are almost always different. Which one is better depends on γ.

We next investigate the numerical properties of the estimators. NE1 seems theoretically better in many cases, but in the theoretically comparative study, the convergence rate of the bandwidth estimator is not taken into account. In this numerical study, we also examine the effect of bandwidth selection. By simulating the mean integrated squared error (MISE) of PE $g_{\widehat{\gamma}_k}$,

$$L_m^{-1} \int_{Q_m(0.1)}^{Q_m(0.9)} \left(g_{\widehat{\gamma}_k}(x) - f_{(m)}(x)\right)^2 \mathrm{d}x,$$

and that of $\widehat{f}_{(m)}$, we investigated the numerical accuracy in small-sample cases, where $L_m := Q_m(0.9) - Q_m(0.1)$ and $Q_m(q)$ denotes the qth quantile of the SMD. Tables 3–4 show the mean values and the standard deviation (sd) of the obtained MISE values after 1000 simulations. The underlying distributions F were Pareto distributions, T distributions, Burr distributions, Fréchet distributions, Weibull distributions, and inverse Burr distributions. The forecast periods were $m = n^{1/4}$, $m = n^{1/2}$, and $m = n^{3/4}$. All kernel functions were Epanechnikov for the bounded class and Gaussian for the other classes.

Both of the bandwidths h_1 and h_2 were estimated by the cross-validation approach or the plug-in approach, respectively denoted as 'CV' and 'PI' in the tables. CV means the unbiased cross-validation estimator (e.g., Silverman 1986 [22]) and Bowman et al (1998) [7]'s cross-validation estimator. Sheather and Jones (1991) [21]'s method and Altman and Leger (1995) [1]'s method were utilized as the plug-in approaches. The sample sizes were $(n =)2^8$ in Table 3 and $(n =)2^{12}$ in Table 4 .

Comparing CV with PI, we can see that CV performs significantly better for $1 \leq \gamma \leq 2$ while PI is much better for $\gamma \geq 4$. In the other cases, the bandwidth estimators are numerically comparable. Comparing NE1 with PE, we find that the MISE values of PE were smaller in most of the cases with $m = n^{1/4}$ and that PE worked numerically better than what we expected on the whole. For $\gamma \fallingdotseq 0$, those of NE1 were smaller as m got large. However, NE1 was not necessary better than PE in finite sample case, which is slightly different from the results of cumulative distribution estimation (see Moriyama 2021 [19]). For the light-tailed distributions with $\gamma < 0$, the MISE values of NE1 were comparable with those of PE, although so-called boundary bias in NE1 for bounded distributions with $\gamma \geq -1$ was observed. Due to this boundary bias, the results for the distributions with large extreme indices (e.g., $\gamma \geq -1/3$) were different from the SMD estimation results reported by Moriyama (2021) [19]. Comparing the cases $(n =)2^{12}$ with $(n =)2^8$, we can see that PE substantially reduces the MISE values, which are smaller than those of NE1.

Table 1: Polynomial convergence rates of MSE of the estimators.

Pareto			$m = n^{1/4}$			$m = n^{1/2}$			$m = n^{3/4}$		
ℓ	α	β	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$1/2$	$1/2$	1	$-1/2$	–	-1	-1	–	$-6/5$	$-3/2$	–	$-7/5$
1	1	1	$-1/2$	–	$-7/10$	-1	–	$-3/5$	$-3/2$	–	$-1/2$
3	3	1	$-1/3$	$-23/30$	$-1/2$	$-2/3$	–	$-1/5$	$-3/4$	–	–
10	10	1	$-1/10$	$-13/20$	$-43/100$	$-1/5$	$-1/2$	$-3/50$	$-3/10$	–	–

T			$m = n^{1/4}$			$m = n^{1/2}$			$m = n^{3/4}$		
ℓ	α	β	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$1/2$	$1/2$	2	$-3/2$	–	-1	$-5/2$	–	$-6/5$	$-13/4$	–	$-7/5$
1	1	2	$-1/2$	–	$-7/10$	$-3/2$	–	$-3/5$	$-7/4$	–	$-1/2$
3	3	2	$-1/3$	$-23/30$	$-1/2$	$-2/3$	–	$-1/5$	$-3/4$	–	–
10	10	2	$-1/10$	$-13/20$	$-43/100$	$-1/5$	$-1/2$	$-3/50$	$-3/10$	–	–

Burr			$m = n^{1/4}$			$m = n^{1/2}$			$m = n^{3/4}$		
c, ℓ	α	β	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$1/2, 1/2$	$1/4$	$1/2$	$-5/2$	–	$-8/5$	$-9/2$	–	$-12/5$	$-25/4$	–	$-16/5$
$1, 1/2$	$1/2$	1	$-1/2$	–	-1	-1	–	$-6/5$	$-3/2$	–	$-7/5$
$3, 1/2$	$3/2$	3	$-2/3$	–	$-3/5$	$-7/6$	–	$-2/5$	$-5/4$	–	$-1/5$
$1/2, 1$	$1/2$	$1/2$	$-3/2$	–	-1	$-3/2$	–	$-6/5$	$-13/4$	–	$-7/5$
$1, 1$	1	1	$-1/2$	–	$-7/10$	-1	–	$-3/5$	$-3/2$	–	$-1/2$
$3, 1$	3	3	$-1/3$	$-23/30$	$-1/2$	$-2/3$	–	$-1/5$	$-3/4$	–	–
$1/2, 3$	$3/2$	$1/2$	$-1/2$	–	$-3/5$	-1	–	$-2/5$	$-5/4$	–	$-1/5$
$1, 3$	3	1	$-1/3$	$-23/30$	$-1/2$	$-2/3$	–	$-1/5$	$-3/4$	–	–
$3, 3$	9	3	$-1/9$	$-59/90$	$-13/30$	$-2/9$	$-23/45$	$-1/15$	$-1/3$	–	–

Fréchet			$m = n^{1/4}$			$m = n^{1/2}$			$m = n^{3/4}$		
γ	α	β	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
5	$1/5$	$1/5$	–	–	$-19/10$	–	–	–	–	–	–
2	$1/2$	$1/2$	$-3/2$	–	-1	$-3/2$	–	$-6/5$	$-13/4$	–	$-7/5$
1	1	1	$-1/2$	–	$-7/10$	-1	–	$-3/5$	$-3/2$	–	$-1/2$
$1/2$	2	1	$-1/2$	$-17/20$	$-11/20$	-1	–	$-3/10$	-1	–	$-1/20$
$1/4$	4	1	$-1/4$	$-29/40$	$-19/40$	$-1/2$	–	$-3/20$	$-5/8$	–	–

Weibull			$m = n^{1/4}$			$m = n^{1/2}$			$m = n^{3/4}$		
κ	γ	ρ	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$1/2$	0	0	–	$-3/5$	$-2/5$	–	$-2/5$	–	–	$-1/5$	–
1	0	0	–	$-3/5$	$-2/5$	–	$-2/5$	–	–	$-1/5$	–
3	0	0	–	$-3/5$	$-2/5$	–	$-2/5$	–	–	$-1/5$	–
10	0	0	–	$-3/5$	$-2/5$	–	$-2/5$	–	–	$-1/5$	–

inv.Burr			$m = n^{1/4}$			$m = n^{1/2}$			$m = n^{3/4}$		
c, ℓ	μ	σ	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$3, 2$	-6	-2	$-1/12$	$-31/60$	$-7/20$	$-1/6$	$-7/30$	–	–	–	–
$1, 2$	-2	-2	–	$-7/20$	$-1/4$	–	–	–	–	–	–
$1/2, 2$	-1	-2	–	$-1/10$	$-1/10$	–	–	–	–	–	–
$3, 1$	-3	-1	–	$-13/30$	$-3/10$	–	$-1/15$	–	–	–	–
$1, 1$	-1	-1	–	$-1/10$	$-1/10$	–	–	–	–	–	–
$1/2, 1$	$-1/2$	-1	–	–	–	–	–	–	–	–	–
$3, 1/3$	-1	$-1/3$	–	–	–	–	–	–	–	–	–
$1, 1/3$	$-1/3$	$-1/3$	–	–	–	–	–	–	–	–	–
$1/2, 1/3$	$-1/6$	$-1/3$	–	–	–	–	–	–	–	–	–

Table 2: Normalized polynomial convergence rates of MSE of the estimators.

Pareto			$m=n^{1/4}$			$m=n^{1/2}$			$m=n^{3/4}$		
ℓ	α	β	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$1/2$	$1/2$	1	–	–	–	–	–	–	–	–	–
1	1	1	–	–	$-1/5$	–	–	–	–	–	–
3	3	1	$-1/6$	$-3/5$	$-1/3$	$-1/3$	–	–	$-1/4$	–	–
10	10	1	$-1/20$	$-3/5$	$-19/50$	$-1/10$	$-2/5$	–	$-3/20$	–	–

T			$m=n^{1/4}$			$m=n^{1/2}$			$m=n^{3/4}$		
ℓ	α	β	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$1/2$	$1/2$	2	$-1/2$	–	–	$-1/2$	–	–	$-1/4$	–	–
1	1	1	–	–	$-1/5$	$-1/2$	–	–	$-1/4$	–	–
3	3	1	$-1/6$	$-3/5$	$-1/3$	$-1/3$	–	–	$-1/4$	–	–
10	10	1	$-1/20$	$-3/5$	$-19/50$	$-1/10$	$-2/5$	–	$-3/20$	–	–

Burr			$m=n^{1/4}$			$m=n^{1/2}$			$m=n^{3/4}$		
c,ℓ	α	β	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$1/2, 1/2$	$1/4$	$1/2$	$-1/2$	–	–	$-1/2$	–	–	$-1/4$	–	–
$1, 1/2$	$1/2$	1	–	–	–	–	–	–	–	–	–
$3, 1/2$	$3/2$	3	$-1/3$	–	$-4/15$	$-1/2$	–	–	$-1/4$	–	–
$1/2, 1$	$1/2$	$1/2$	$-1/2$	–	–	–	–	–	$-1/4$	–	–
$1, 1$	1	1	–	–	$-1/5$	–	–	–	–	–	–
$3, 1$	3	3	$-1/6$	$-3/5$	$-1/3$	$-1/3$	–	–	$-1/4$	–	–
$1/2, 3$	$3/2$	$1/2$	$-1/6$	–	$-4/15$	$-1/3$	–	–	$-1/4$	–	–
$1, 3$	3	1	$-1/6$	$-3/5$	$-1/3$	$-1/3$	–	–	$-1/4$	–	–
$3, 3$	9	3	$-1/18$	$-3/5$	$-17/45$	$-1/9$	$-2/5$	–	$-1/6$	–	–

Fréchet			$m=n^{1/4}$			$m=n^{1/2}$			$m=n^{3/4}$		
γ	α	β	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
5	$1/5$	$1/5$	–	–	–	–	–	–	–	–	–
2	$1/2$	$1/2$	$-1/2$	–	–	–	–	–	$-1/4$	–	–
1	1	1	–	–	$-1/5$	–	–	–	–	–	–
$1/2$	2	1	$-1/4$	$-3/5$	$-3/10$	$-1/2$	–	–	$-1/4$	–	–
$1/4$	4	1	$-1/8$	$-3/5$	$-7/20$	$-1/4$	–	–	$-1/4$	–	–

Weibull			$m=n^{1/4}$			$m=n^{1/2}$			$m=n^{3/4}$		
κ	γ	ρ	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$1/2$	0	0	–	$-3/5$	$-2/5$	–	$-2/5$	–	–	$-1/5$	–
1	0	0	–	$-3/5$	$-2/5$	–	$-2/5$	–	–	$-1/5$	–
3	0	0	–	$-3/5$	$-2/5$	–	$-2/5$	–	–	$-1/5$	–
10	0	0	–	$-3/5$	$-2/5$	–	$-2/5$	–	–	$-1/5$	–

inv.Burr			$m=n^{1/4}$			$m=n^{1/2}$			$m=n^{3/4}$		
c,ℓ	μ	σ	PE	NE1	NE2	PE	NE1	NE2	PE	NE1	NE2
$3, 2$	-6	-2	$-1/6$	$-3/5$	$-13/30$	$-1/3$	$-2/5$	–	–	$-1/5$	–
$1, 2$	-2	-2	–	$-3/5$	$-1/2$	–	$-2/5$	–	–	$-1/5$	–
$1/2, 2$	-1	-2	–	$-3/5$	$-3/5$	–	$-2/5$	–	–	$-1/5$	–
$3, 1$	-3	-1	$-1/6$	$-3/5$	$-7/15$	$-1/6$	$-2/5$	–	–	$-1/5$	–
$1, 1$	-1	-1	–	$-3/5$	$-3/5$	–	$-2/5$	–	–	$-1/5$	–
$1/2, 1$	$-1/2$	-1	–	–	–	–	–	–	–	–	–
$3, 1/3$	-1	$-1/3$	–	–	–	–	–	–	–	–	–
$1, 1/3$	$-1/3$	$-1/3$	–	–	–	–	–	–	–	–	–
$1/2, 1/3$	$-1/6$	$-1/3$	–	–	–	–	–	–	–	–	–

Table 3: Scaled MISE values and sd values of PE and NE1 where $n = 2^8$.

Pareto	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
ℓ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
1/2	0.242	0.443	8.065	1.721	7.059	11.99	9.222	108.9	9.339	3.345	487.7	6491	51.01	383.8	108.6	296.2	22770	311400
1	0.037	0.041	0.353	0.393	0.771	0.422	0.510	5.696	1.474	1.156	7.401	12.05	33.36	696.2	20.49	27.17	122.8	500.6
3	0.019	0.021	0.133	0.057	0.088	0.036	0.116	0.162	0.799	0.344	0.514	0.178	54.79	1202	5.143	2.904	3.231	1.437
10	0.016	0.017	0.102	0.046	0.054	0.025	0.104	0.176	0.447	0.202	0.240	0.105	7.608	71.76	2.073	1.007	1.093	0.484

T	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
ℓ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
1/2	0.996	0.986	7.695	0.921	4.380	2.679	6.928	53.80	9.947	0.953	146.1	3708	41.92	205.4	31.32	61.64	28650	856500
1	0.068	0.067	0.279	0.323	0.314	0.267	0.367	0.744	0.909	0.446	2.751	7.268	42.09	695.3	8.641	9.806	45.99	296.1
3	0.016	0.014	0.021	0.020	0.018	0.014	0.101	0.213	0.095	0.075	0.088	0.065	12.73	215.9	0.507	0.336	0.496	0.322
10	0.011	0.012	0.015	0.015	0.012	0.010	0.088	0.137	0.049	0.044	0.042	0.036	7.153	72.64	0.193	0.177	0.162	0.144

Burr	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
c, ℓ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
1/2, 1/2	102.4	156.6	714.1	1.879	361.2	139.4	548.3	1212	942.8	14.97	942.3	0.909	1306	6052	9436	75610	1037	0.000
1, 1/2	0.217	0.416	8.115	1.693	7.056	11.85	5.507	22.83	9.281	3.120	385.1	5181	43.43	360.3	104.3	258.9	84390	2319000
3, 1/2	0.023	0.029	0.097	0.068	0.113	0.093	0.178	0.341	0.754	0.386	0.866	0.485	12.23	107.1	7.402	5.880	8.760	9.997
1/2, 1	0.411	0.674	11.52	1.685	10.20	27.86	7.554	46.29	10.36	4.495	1590	41270	44.30	288.4	117.4	394.4	11970	212600
1, 1	0.041	0.046	0.309	0.350	0.849	0.470	0.391	1.235	1.456	1.274	8.263	12.39	17.72	160.2	20.57	26.62	159.3	738.4
3, 1	0.014	0.015	0.034	0.025	0.031	0.019	0.112	0.190	0.187	0.121	0.176	0.090	7.005	39.81	1.117	0.676	1.072	0.509
1/2, 3	0.072	0.072	0.242	0.209	1.355	0.698	0.291	1.140	1.266	0.976	15.32	23.85	53.22	1205	13.95	14.99	178.7	748.7
1, 3	0.019	0.021	0.132	0.055	0.085	0.035	0.123	0.202	0.804	0.343	0.504	0.173	81.56	1379	5.059	2.907	3.255	1.361
3, 3	0.013	0.015	0.017	0.018	0.015	0.012	0.091	0.167	0.065	0.065	0.056	0.043	18.88	437.1	0.263	0.235	0.232	0.182

Fréchet	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
γ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
5	4449	4841	13700	1.368	5955	7608	11340	20670	13710	13690	29.60	108.4	13990	16200	74840	722000	13700	0.000
2	0.303	0.741	9.454	1.692	8.314	12.75	6.025	38.12	9.718	3.716	235.7	2446	61.86	412.0	114.2	316.4	74590	1840000
1	0.041	0.054	0.329	0.374	0.829	0.561	0.381	1.065	1.505	1.232	9.181	21.77	123.3	3281	20.28	27.15	246.2	2078
1/2	0.019	0.022	0.082	0.043	0.076	0.042	0.139	0.208	0.588	0.281	0.519	0.184	20.72	343.2	4.717	3.338	3.991	1.861
1/4	0.015	0.018	0.040	0.027	0.035	0.021	0.102	0.158	0.197	0.125	0.176	0.096	47.66	664.3	1.114	0.692	0.972	0.458

Weibull	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
κ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
1/2	0.044	0.038	0.159	0.067	0.299	0.085	0.108	0.169	0.978	0.438	1.828	0.702	54.80	1520	5.554	2.943	10.18	7.603
1	0.017	0.020	0.085	0.045	0.044	0.024	0.092	0.148	0.333	0.178	0.167	0.087	39.12	621.0	1.285	0.712	0.669	0.343
3	0.011	0.013	0.012	0.013	0.010	0.009	0.090	0.171	0.035	0.032	0.031	0.027	18.64	242.2	0.117	0.099	0.112	0.101
10	0.011	0.012	0.012	0.011	0.009	0.008	0.088	0.127	0.034	0.030	0.027	0.023	8.291	127.5	0.126	0.091	0.098	0.073

inv.Burr	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
c, ℓ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
3, 2	0.015	0.012	0.017	0.025	0.031	0.031	0.098	0.097	0.056	0.087	0.015	0.019	0.043	0.054	0.182	0.158	0.188	0.201
1, 2	0.015	0.011	0.019	0.022	0.032	0.023	0.144	0.100	0.078	0.092	0.015	0.023	0.039	0.077	0.258	0.204	0.121	0.126
1/2, 2	0.012	0.010	0.062	0.084	0.226	0.060	0.875	0.009	0.299	0.174	0.035	0.022	0.218	0.106	0.741	0.115	0.681	0.095
3, 1	0.175	0.152	0.533	0.229	0.572	0.160	0.684	0.021	0.672	0.084	0.363	0.225	0.682	0.024	0.689	0.005	0.674	0.000
1, 1	0.289	0.155	0.731	0.199	0.851	0.098	0.905	0.004	0.905	0.047	0.502	0.256	0.905	0.004	0.906	0.001	0.898	0.000
1/2, 1	1.524	0.184	2.400	0.192	3.027	0.022	3.260	0.000	3.043	0.014	1.921	0.471	3.260	0.000	3.260	0.000	3.340	0.000
3, 1/3	4.675	0.129	5.270	0.000	1.890	0.000	1.300	0.000	1.890	0.000	2.718	0.004	1.300	0.000	1.300	0.000	1.040	0.000
1, 1/3	56.10	0.203	56.60	0.000	30.70	0.000	26.50	0.000	30.70	0.000	37.40	0.000	26.50	0.000	26.50	0.000	25.40	0.000
1/2, 1/3	24500	0.000	24500	0.000	42700	0.000	54800	0.000	42700	0.000	33300	0.000	54800	0.000	54800	0.000	59600	0.000

Table 4: Scaled MISE values and sd values of PE and NE1 where $n = 2^{12}$.

Pareto	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
ℓ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
1/2	0.107	0.451	10.05	0.083	10.34	15.39	0.433	0.804	34.50	51.04	5735	173000	64.82	1061	17500	44480	10.83	0.000
1	0.004	0.004	1.682	0.043	3.820	3.328	0.042	0.070	5.073	23.29	208.8	842.2	8.390	141.0	49.05	206.5	11880	187200
3	0.002	0.002	0.034	0.009	0.043	0.041	0.017	0.020	0.515	0.179	0.653	0.621	0.964	5.703	8.932	6.583	9.987	12.90
10	0.002	0.002	0.034	0.007	0.019	0.005	0.013	0.016	0.311	0.073	0.183	0.048	0.414	1.135	3.073	1.143	1.782	0.582

T	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
ℓ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
1/2	0.153	0.314	9.952	0.020	9.941	12.32	0.328	0.633	10.74	0.000	791.8	24500	44.45	758.3	10.83	0.000	10.83	0.000
1	0.015	0.012	1.474	0.159	2.010	1.050	0.038	0.048	15.07	60.21	124.9	425.5	3.296	20.33	224.5	1409	2846	23870
3	0.003	0.002	0.015	0.009	0.005	0.005	0.015	0.017	0.181	0.100	0.064	0.073	3.085	71.44	2.735	2.013	0.945	0.927
10	0.002	0.002	0.003	0.002	0.003	0.002	0.013	0.014	0.023	0.017	0.019	0.014	3.835	70.94	0.207	0.186	0.156	0.127

Burr	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
c, ℓ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
1/2, 1/2	457.0	250.5	845.4	0.001	790.4	24.09	549.6	257.4	1037	0.000	1037	0.000	1183	3083	161600	281100	1070	0.000
1, 1/2	0.053	0.239	10.06	0.058	10.81	12.13	0.467	0.751	10.74	0.000	10.74	0.000	21.49	73.41	287.8	2015	10.83	0.000
3, 1/2	0.002	0.003	0.126	0.027	0.480	0.426	0.027	0.027	0.532	0.262	12.34	15.00	0.657	0.744	19.65	45.41	361.4	1623
1/2, 1	0.094	0.362	11.71	0.060	11.11	0.180	0.292	0.437	11.55	2.249	10.95	0.000	56.17	327.3	5921	32310	10.86	0.000
1, 1	0.004	0.005	1.702	0.033	3.707	2.544	0.033	0.044	1.912	0.216	316.2	1121	28.81	242.3	33.34	66.91	22.28	204.4
3, 1	0.002	0.002	0.008	0.004	0.009	0.007	0.016	0.016	0.124	0.043	0.131	0.119	1.263	8.059	1.842	0.618	1.983	0.715
1/2, 3	0.009	0.007	0.460	0.576	4.466	3.297	0.022	0.024	2.125	15.44	360.0	1408	0.948	1.459	21.50	30.83	3822	34410
1, 3	0.002	0.003	0.033	0.009	0.042	0.036	0.017	0.018	0.500	0.137	0.630	0.420	0.633	1.195	8.755	7.714	9.593	8.013
3, 3	0.002	0.001	0.003	0.002	0.003	0.002	0.012	0.012	0.027	0.021	0.024	0.016	0.459	0.749	0.253	0.170	0.227	0.121

Fréchet	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
γ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
5	11140	3586	13700	0.000	12880	631.1	10390	3020	13700	0.000	13700	0.000	31820	236400	698200	1287000	13700	0.000
2	0.108	0.511	10.80	0.075	12.45	66.10	0.389	0.636	407.6	447.7	10.84	0.060	71.17	1018	50850	96120	10.84	0.000
1	0.004	0.004	1.654	0.088	4.228	3.256	0.042	0.057	5.166	29.94	290.8	1724	4.465	32.78	138.6	1138	16710	307700
1/2	0.002	0.002	0.027	0.010	0.168	0.221	0.018	0.019	0.528	0.227	3.440	6.561	3.172	34.75	15.59	22.17	57.71	240.0
1/4	0.002	0.002	0.022	0.013	0.009	0.004	0.015	0.017	0.253	0.144	0.117	0.035	0.676	4.942	3.613	2.658	1.590	0.540

Weibull	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
κ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
1/2	0.007	0.004	0.039	0.010	0.303	0.131	0.015	0.016	0.526	0.148	4.235	3.119	0.756	6.083	6.578	3.672	47.28	97.62
1	0.002	0.002	0.031	0.007	0.015	0.005	0.012	0.013	0.242	0.060	0.115	0.035	0.433	1.330	1.932	0.586	0.924	0.326
3	0.002	0.001	0.002	0.002	0.002	0.002	0.011	0.012	0.014	0.010	0.014	0.011	0.490	3.103	0.092	0.073	0.090	0.081
10	0.002	0.002	0.002	0.002	0.002	0.002	0.011	0.013	0.019	0.014	0.018	0.012	0.465	2.997	0.126	0.108	0.134	0.101

inv.Burr	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
c, ℓ	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd	PE	sd	CV	sd	PI	sd
3, 2	0.006	0.003	0.012	0.019	0.012	0.019	0.011	0.011	0.035	0.050	0.035	0.050	0.298	0.370	0.165	0.186	0.165	0.186
1, 2	0.002	0.001	0.012	0.016	0.012	0.016	0.012	0.011	0.028	0.035	0.028	0.035	0.383	0.449	0.095	0.091	0.095	0.091
1/2, 2	0.015	0.061	0.013	0.008	0.013	0.008	0.017	0.022	0.221	0.048	0.221	0.048	0.379	0.441	0.876	0.009	0.876	0.009
3, 1	0.076	0.109	0.364	0.220	0.364	0.220	0.009	0.010	0.684	0.016	0.684	0.016	0.498	1.580	0.674	0.000	0.674	0.000
1, 1	0.226	0.246	0.493	0.272	0.493	0.272	0.023	0.035	0.904	0.006	0.904	0.006	0.467	0.845	0.898	0.000	0.898	0.000
1/2, 1	2.056	0.706	2.849	0.002	1.873	0.498	1.173	0.992	3.260	0.000	3.260	0.000	1.597	0.843	3.340	0.000	3.340	0.000
3, 1/3	2.457	0.161	2.670	0.000	2.719	0.003	0.141	0.258	1.300	0.000	1.300	0.000	0.423	1.253	1.040	0.000	1.040	0.000
1, 1/3	37.26	0.079	37.40	0.000	37.40	0.000	23.44	4.776	26.50	0.000	26.50	0.000	22.32	3.069	25.40	0.000	25.40	0.000
1/2, 1/3	33250	107.8	33300	0.000	33300	0.000	54800	0.000	54800	0.000	54800	0.000	59610	35.55	59600	0.000	59600	0.000

6. SMD estimation with data transformation

Ever since the Rosenblatt-Parzen kernel density estimator was invented, nonparametric density estimation has attracted the attention of both researchers and practitioners. Many researchers have tried to improve the accuracy of nonparametric density estimators. Most of them are on the choice of the bandwidth that significantly affects on the MSE. As in case of SMD estimation the kernel function does not change the convergence rates (as long as the sufficiently high-order moment exists); however, it is known that the choice possibly has effect on the numerical accuracy in finite sample case especially when the tail of the underlying distribution is heavy (e.g. Hall 1987 [13]; Maiboroda and Markovich 2004 [15]; Markovich 2007 [16]). Hall (1987) [13] proved that the KL divergence of the kernel density estimator and the performance of the likelihood cross-validation (for the bandwidth selection) depend on both the tails of the kernel and the underlying distribution.

Markovich (2018) [17] proposed the Weibull kernel to fit the tail part of the underlying distribution. Wang et al. (2020) [26] reported cross validation approaches in heavy tailed density estimation do not work well and proposed modified kernel functions. A reasonable kernel has a data-driven tail, that is the extreme value index of the kernel is fitted to that of the underlying distribution. The piecing-together approach (Markovich 2007 [16]; MacDonald et al. 2011 [14]), which fits the generalized Pareto distribution to the tail of the distribution, is related to the tail selection of the kernel.

Following the reviewer's comments, additional numerical experiments on the effect of kernel selection was conducted in this study. It was demonstrated that for heavy tailed distributions NE1 with equally heavy tailed kernel generally outperforms that of the Gaussian kernel used in Section 5. We omit the details; however, the experimental results imply that the numerical accuracy does not depend only on the level of heaviness (i.e. the extreme index) of the kernel. The experiments also clarified that the kernel being same as the underlying distribution is not necessarily best choice. More detailed studies on the choice of the kernel in SMD estimation may be important, but we postpone it to future work. This study aims at improving the nonparametric estimators without specifying the heaviness of the underlying distribution F.

Section 5 demonstrates that both of the theoretical and numerical accuracies of the nonparametric estimators of SMD heavily depends on the extreme value index $\gamma := \gamma_X$ of the underlying distribution (we don't add the subscript but note that γ is the extreme value index of F throughout the paper). When $\gamma \fallingdotseq 0$, NE functions well. However, the scaled MISE of NE explodes for large $|\gamma|$. Following these results, we consider the transformed nonparametric density estimation (see e.g. Wand et al. 1991 [27]; Marron and Ruppert 1994 [18]; Charpentier and Flachaire 2015 [9]; Béranger et al. 2019 [5]). Maiboroda and Markovich (2004) [15], Buch-Larsen et al. (2005) [8], Bolancé et al. (2015) [6], Béranger et al. (2019) [5] and Tokdar (2022) [24] intend to apply data transformation in order to change the distribution's tail to the same as that of the kernel.

By applying a data transformation $X \mapsto \varphi(X)$ to SMD estimation, the plug-in type nonparametric estimator with data transformation (TNE1):

$$\widehat{f}^{\varphi}_{(m)}(x; h_1, h_2) := m\widehat{f}^{\varphi}(x; h_1)(\widehat{F}^{\varphi})^{m-1}(x; h_2),$$

where $\widehat{f}^{\varphi}$ and $\widehat{F}^{\varphi}$ are the kernel estimators given by

$$\widehat{f}^{\varphi}(x;h_1) := \frac{\varphi'(x)}{nh_1}\sum_{i=1}^{n} w\left(\frac{\varphi(x)-\varphi(X_i)}{h_1}\right),$$
$$\widehat{F}^{\varphi}(x;h_2) := \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{\varphi(x)-\varphi(X_i)}{h_2}\right).$$

The BM-based kernel-type of nonparametric estimator with data transformation (TNE2) is

$$\bar{f}^{\varphi}_{(m)}(x;h) := \frac{\varphi'(x)}{\bar{n}h}\sum_{i=1}^{\bar{n}} w\left(\frac{\varphi(x)-\varphi(Y_i)}{h}\right).$$

This study considers the log-transformation $\varphi = \ln$, which is considered as a special version of the Box-Cox transformation. Then, the log-transformation $X \mapsto \ln(X)$ changes positive extreme value indices to zero as follows.

Corollary 6.1 (Wadsworth et al. 2010 [25]). *If F belongs to the Hall class, the extreme value index of the distribution of $\ln(X)$ is zero. If F belongs to the Weibull class or the bounded class, the extreme value index is same as γ.*

This is seen from the Hall class of distributions becoming the Weibull class of distributions as follows

$$\mathbb{P}[\ln X > x] \sim A\exp(-\alpha x) + AB\exp(-\alpha-\beta x),$$

which corresponds to the Weibull class with $C = \alpha$ and $\kappa = 1$.

As seen in Section 5, the nonparametric estimators NE1 and NE2 are numerically stable and work well for the Weibull class even when m is relatively large. Theoretically, the MSE is given by the following theorem.

Theorem 6.1. *Given Assumptions 3.1–3.3, then*

$$\mathbb{E}[(f_{(m)}(x) - \widehat{f}^{\ln}_{(m)}(x;h_1,h_2))^2] \sim b_{n,3}^2 + v_{n,3},$$

where

$$b_{n,3} := \frac{h_1^2}{2}\exp(-M_n)M_n^{1+3\gamma}m^{-3\gamma}x^2\psi_n'\int z^2w(z)\mathrm{d}z + \frac{h_2^2}{2}\exp(-M_n)M_n m\xi_n' x^2 f(x)\int z^2 w(z)\mathrm{d}z,$$
$$v_{n,3} := \frac{m^2}{nh_1}\exp(-2M_n)x^{-1}f(x)\int w^2(z)\mathrm{d}z + \frac{m^3}{n}\exp(-2M_n)M_n f^2(x),$$
$$\psi_n' := \begin{cases} \alpha\{1-3(\alpha+1)+(\alpha+1)(\alpha+2)\}A^{-3\gamma} & \text{for (i)} \\ \kappa^3C^3x^{3\kappa-3} & \text{for (ii)} \\ -\mu\{1-3(\mu+1)+(\mu+1)(\mu+2)\}D^{-3\gamma} & \text{for (iii),} \end{cases}$$
$$\xi_n' := \begin{cases} \alpha^2x^{-2} & \text{for (i)} \\ \kappa^2C^2x^{2\kappa-2} & \text{for (ii)} \\ \mu^2(x^*-x)^{-2} & \text{for (iii).} \end{cases}$$

Furthermore, $v_{n,3}^{-1/2}\{f_{(m)}(x) - \widehat{f}^{\ln}_{(m)}(x;h_1,h_2) + b_{n,3}\}$ converges in distribution to the standard normal distribution, if the MSE converges to zero.

Similarly as Corollary 3.1, asymptotically optimal bandwidths of TNE1 are those of $\widehat{f}^{\ln}(x;h_1)$ and $\widehat{F}^{\ln}(x;h_2)$.

Corollary 6.2. *An asymptotically optimal value of (h_1,h_2) is given by*

$$x^{-1}\left(M_n^{-2-6\gamma}m^{2+6\gamma}\psi_n'^{-2}f(x)n^{-1}\left(\int z^2w(z)\mathrm{d}z\right)^{-2}\int w^2(z)\mathrm{d}z\right)^{1/5},$$

$$x^{-1}\left(2\xi_n'^{-2}\omega_n n^{-1}\left(\int z^2w(z)\mathrm{d}z\right)^{-2}\int zW(z)w(z)\mathrm{d}z\right)^{1/3},$$

which are asymptotically identical to those of $\widehat{f}^{\ln}(x;h_1)$ and $\widehat{F}^{\ln}(x;h_2)$ if both of the optimal values converge 0. Under the assumption of Theorem 6.1, $f_{(m)}(x)-\widehat{f}^{\ln}_{(m)}(x;h_1,h_2)$ with the optimal bandwidths is asymptotically non-degenerate normal with the following asymptotic mean:

$$\nu_1\exp(-M_n)M_n^{(1+3\gamma)/5}m^{(4-3\gamma)/5}\psi_n'^{1/5}f^{2/5}(x)n^{-2/5}+\nu_2\exp(-M_n)M_n m\xi_n'^{-1/3}f(x)\omega_n^{2/3}n^{-2/3},$$

whose order is same as that of NE1 $\widehat{f}_{(m)}$.

Corollary 6.3. *The asymptotically optimal bandwidths for $M_n=O(1)$ are of the order*

$$\left(\frac{m}{n}\right)^{1/5}\times\begin{cases}1 & \textit{for}\quad \text{(i), (ii)}\\ \{C^{-1}\ln m\}^{-1} & \textit{for}\quad \text{(ii),}\end{cases}$$

and

$$\left(\frac{m}{n}\right)^{1/3}\times\begin{cases}1 & \textit{for}\quad \text{(i), (iii)}\\ \{C^{-1}\ln m\}^{-1} & \textit{for}\quad \text{(ii).}\end{cases}$$

The convergence rate of TNE1 $\widehat{f}^{\ln}_{(m)}$ is same as that of the non-transformed $\widehat{f}_{(m)}$. However, the order of the optimal bandwidth is different. The condition $h_j\to 0$ for $j=1,2$ is equivalent to $m=o(n)$ when $M_n=O(1)$, which is less restrictive for $\gamma>0$. The proof of Theorem 6.1 is provided in the Supplementary file. Corollaries 6.2 and 6.3 are directly obtained from Theorem 6.1.

Theorem 6.2. *Given Assumptions 3.2, 3.3 and 4.1, if $x^{\kappa-1}h\to 0$,*

$$\mathbb{E}[(f_{(m)}(x)-\bar{f}^{\ln}_{(m)}(x;h))^2]\sim x^4b_{n,2}^2+x^{-1}v_{n,2}.$$

Furthermore, $x^{1/2}v_{n,2}^{-1/2}\{f_{(m)}(x)-\bar{f}^{\ln}_{(m)}(x;h)+x^2b_{n,2}\}$ converges in distribution to the standard normal distribution, if the MSE converges to zero.

Corollary 6.4. *The asymptotically optimal value of h is given by*

$$x^{-1}\left(n^{-1}\{f(x)\}^{-1}\phi_n^{-2}\left(\int z^2w(z)\mathrm{d}z\right)^{-2}\int w^2(z)\mathrm{d}z\right)^{1/5}$$

if both of the optimal values converge 0. Under the assumption of Theorem 4.1, $f_{(m)}(x) - \bar{f}^{\ln}_{(m)}(x;h)$ *with the optimal bandwidths is asymptotically non-degenerate normal with the following asymptotic mean:*

$$\nu_1 n^{-2/5} m\{f(x)\}^{3/5}\phi_n^{1/5},$$

which is exactly same as that of NE2 $\bar{f}_{(m)}(x;h)$*.*

Corollary 6.5. *The asymptotically optimal bandwidth for* $M_n = O(1)$ *is of the order*

$$n^{-1/5}m^{-3/5} \times \begin{cases} m^{-4\gamma/5} & \textit{for} \quad \text{(i), (iii)} \\ (\ln m)^{-1/5} & \textit{for} \quad \text{(ii)}. \end{cases}$$

Corollary 6.5 shows that the optimal convergence rate of TNE2 $\bar{f}^{\ln}_{(m)}$ is same as $\bar{f}_{(m)}$. The asymptotically optimal bandwidth is different from that of NE2. $h \to 0$ in Assumption 4.1 is less restrictive for $\gamma > 0$ same as TNE1. Theorem 6.2 is proved in the Supplementary file. Corollaries 6.4 and 6.5 are the direct consequences.

Corollary 6.1 states the extreme value index of $\ln(X)$ do not need to be estimated as long as γ (the index of F) is non-negative. Motivated by the facts, applying the back transformation we can consider the following transformed type of PE (TPE)

$$\widetilde{g}^{\ln}_{\widehat{\gamma}_k}(\ln x) := \frac{1}{\widehat{a}^{\ln}_k x} \exp\left(-\frac{\ln x - \widehat{b}^{\ln}_k}{\widehat{a}^{\ln}_k}\right) \exp\left(-\exp\left(-\frac{\ln x - \widehat{b}^{\ln}_k}{\widehat{a}^{\ln}_k}\right)\right)$$

for distributions with $\gamma \geq 0$, where $\widehat{a}^{\ln}_k$ and $\widehat{b}^{\ln}_k$ are the estimators of the normalizing constants of $\max\{\ln(X_i)\}_{i=1}^n$. Specifically, $a^{\ln}_k = b_k^{-1} a_k$ and $b^{\ln}_k = \ln b_k$ (Wadsworth et al. 2010 [25]). However, the distributions with $\gamma = 0$ never satisfy all the assumptions of PE at the same time as described in Section 5.

For distributions with $\gamma < 0$ pointwise estimation by PE at a point $x \sim (x^* - z_n^{-1/\mu})$ changes that of by the following type of estimator

$$g^{\ln}_{\widehat{\gamma}_k}(\ln x) := \frac{1}{\widehat{a}^{\ln}_k x} g_{\widehat{\gamma}^{\ln}_k}\left(\frac{\ln x - \widehat{b}^{\ln}_k}{\widehat{a}^{\ln}_k}\right)$$

at the point $\ln x \sim (\ln x^* - z_n^{-1/\mu})$ by the transformation, which means the asymptotics of do not change since the extreme value index is same as γ. $\widehat{\gamma}^{\ln}_k$ is the estimator based on $\{\ln(X_i)\}_{i=1}^n$ of the extreme value index of $\{\ln(X)\}$, that is same as γ in this case.

The above means TPE is only applicable for the bounded class under the assumption of negativity of γ, but the log-transformation does not improve theoretical accuracy. In short, there are no theoretical reasons to employ TPE instead of PE in any case.

7. Numerical study on transformed SMD estimators

This section surveys the numerical performances of the estimators TPE and TNE1. Suppose $M_n \equiv \delta > 0$. Then, Corollary 6.5 states that the asymptotic convergence rates of the transformed nonparametric estimator TNE1 correspond to those of TNE1 (shown in Table 1 and Table 2). On the other hand, TPE for the distributions with $\gamma \geq 0$ cannot be consistent.

The scaled MISE values were simulated in Pareto, T, Burr, Fréchet, Weibull, and inverse Burr cases. To apply the log-transformation the distributions are truncated to the range $(0, \infty)$ or $(0, \ln x^*)$. Other technical details are same as Section 5. In Table 5 the MISE value in bold means being smaller than that of the MISE value in Table 3. That means the log-transformation improves the estimator.

Table 5 shows the log-transformation improves the estimators in many cases on the whole. The exceptions are mainly TPE and the cases where m is relatively small. Though the MSE values increase in the cases, the increments are relatively not large. For heavy-tailed data the log-transformation especially improves the accuracy of the estimators. To sum up, the log-transformation makes TNE1 numerically stable and much more accurate.

Table 5: Scaled MISE values and sd values of TPE and TNE1 where $n = 2^8$.

Pareto	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
ℓ	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd
1/2	4.013	0.714	**0.668**	0.456	**0.383**	0.310	**7.332**	2.885	**2.754**	2.182	**1.630**	1.568	**8.739**	3.653	**12.95**	15.47	**7.500**	9.460
1	0.070	0.034	**0.160**	0.093	**0.085**	0.054	**0.097**	0.135	**0.611**	0.407	**0.334**	0.252	**0.705**	0.649	**2.562**	1.963	**1.356**	1.045
3	0.465	0.013	**0.092**	0.050	**0.047**	0.027	0.431	0.038	**0.344**	0.184	**0.181**	0.097	**0.579**	0.167	**1.378**	0.825	**0.721**	0.428
10	0.643	0.006	**0.085**	0.043	**0.045**	0.024	0.620	0.015	**0.317**	0.170	**0.169**	0.093	**0.658**	0.049	**1.280**	0.699	**0.695**	0.394

T	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
ℓ	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd
1/2	30.44	3.044	11.82	2.253	11.75	2.203	33.62	8.386	11.91	4.714	**11.65**	4.558	**22.87**	17.51	**16.17**	12.67	**15.65**	12.23
1	1.978	0.145	2.009	0.303	2.006	0.300	1.559	0.265	2.113	0.651	**2.056**	0.620	**1.027**	0.658	**2.687**	1.660	**2.576**	1.553
3	0.324	0.027	0.848	0.104	0.844	0.098	0.130	0.021	0.916	0.209	0.880	0.189	**0.065**	0.042	1.263	0.599	1.151	0.469
10	0.182	0.018	0.688	0.080	0.685	0.073	**0.047**	0.008	0.674	0.123	0.640	0.107	**0.017**	0.011	0.895	0.354	0.777	0.245

Burr	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
c, ℓ	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd
1/2, 1/2	4974	1196	**17.78**	19.19	**15.91**	17.39	6531	3155	**82.01**	115.8	**76.20**	113.3	5276	4927	**426.2**	857.8	**397.2**	804.2
1, 1/2	6.648	1.423	**0.215**	0.211	**0.194**	0.180	7.942	2.976	**0.826**	0.939	**0.761**	0.847	**8.633**	3.691	**3.608**	5.146	**3.362**	4.835
3, 1/2	0.185	0.021	**0.029**	0.025	**0.026**	0.022	0.214	0.080	**0.104**	0.088	**0.096**	0.081	**0.554**	0.362	**0.439**	0.437	**0.408**	0.397
1/2, 1	24.43	6.227	**0.307**	0.327	**0.268**	0.269	11.41	4.900	**0.726**	0.793	**0.664**	0.723	**9.402**	4.217	**3.177**	4.600	**2.938**	4.275
1, 1	0.067	0.038	**0.040**	0.039	**0.033**	0.032	**0.079**	0.113	**0.125**	0.141	**0.108**	0.126	**0.680**	0.646	**0.518**	0.640	**0.458**	0.560
3, 1	0.326	0.014	**0.017**	0.017	**0.015**	0.013	0.400	0.043	**0.057**	0.052	**0.050**	0.044	**0.576**	0.171	**0.230**	0.239	**0.203**	0.207
1/2, 3	0.717	0.262	**0.039**	0.038	**0.033**	0.031	**0.075**	0.088	**0.075**	0.076	**0.065**	0.060	**0.498**	0.438	**0.264**	0.263	**0.218**	0.201
1, 3	0.072	0.014	**0.016**	0.017	**0.014**	0.013	0.230	0.051	**0.045**	0.042	**0.039**	0.037	**0.503**	0.205	**0.184**	0.183	**0.146**	0.133
3, 3	0.407	0.010	**0.013**	0.012	**0.011**	0.009	0.503	0.023	**0.040**	0.033	**0.034**	0.029	**0.612**	0.068	**0.157**	0.124	**0.123**	0.097

Fréchet	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
γ	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd
5	253600	23340	16110	3595	16000	3407	341200	101000	18350	9540	18030	9273	246000	188000	25310	27120	24300	25270
2	45.47	2.126	13.65	2.458	13.58	2.415	42.02	7.146	13.20	4.971	12.92	4.817	**26.01**	18.43	**17.09**	13.50	**16.53**	12.75
1	2.730	0.113	2.136	0.315	2.131	0.308	1.845	0.255	2.178	0.635	2.125	0.608	**1.098**	0.739	**2.713**	1.542	**2.605**	1.461
1/2	0.700	0.030	1.057	0.136	1.055	0.131	0.343	0.051	1.126	0.277	1.091	0.255	**0.175**	0.119	**1.523**	0.731	**1.413**	0.611
1/4	0.361	0.016	0.800	0.096	0.799	0.090	0.137	0.022	0.842	0.175	0.812	0.154	**0.062**	0.040	1.149	0.505	1.042	0.398

Weibull	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
κ	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd
1/2	0.106	0.064	**0.020**	0.019	**0.016**	0.014	0.131	0.059	**0.039**	0.034	**0.035**	0.032	**0.443**	0.194	**0.123**	0.088	**0.101**	0.074
1	0.151	0.014	**0.013**	0.012	**0.011**	0.009	0.339	0.034	**0.031**	0.027	**0.027**	0.025	**0.532**	0.103	**0.116**	0.084	**0.090**	0.066
3	0.454	0.007	0.012	0.011	0.010	0.009	0.545	0.013	**0.032**	0.028	**0.028**	0.025	**0.619**	0.036	0.118	0.084	**0.092**	0.066
10	0.606	0.003	0.012	0.011	0.010	0.008	0.633	0.004	**0.031**	0.025	**0.026**	0.023	**0.657**	0.011	**0.121**	0.081	**0.091**	0.063

inv.Burr	$m = n^{1/4}$						$m = n^{1/2}$						$m = n^{3/4}$					
c, ℓ	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd	TPE	sd	CV	sd	PI	sd
3, 2	0.719	0.000	0.019	0.022	**0.016**	0.017	0.678	0.000	**0.049**	0.051	0.060	0.061	0.672	0.000	**0.181**	0.130	0.244	0.119
1, 2	0.722	0.000	0.019	0.022	**0.016**	0.015	0.691	0.000	**0.053**	0.054	0.065	0.059	0.685	0.000	**0.248**	0.139	0.322	0.103
1/2, 2	0.836	0.000	**0.035**	0.028	**0.034**	0.023	**0.868**	0.000	**0.194**	0.100	0.281	0.079	0.888	0.000	**0.684**	0.134	0.801	0.044
3, 1	0.870	0.000	**0.324**	0.234	**0.180**	0.126	0.727	0.000	**0.541**	0.193	0.619	0.115	0.690	0.000	**0.661**	0.077	0.688	0.006
1, 1	1.071	0.000	**0.566**	0.263	**0.313**	0.143	**0.934**	0.000	**0.834**	0.125	0.872	0.071	0.906	0.000	**0.903**	0.012	0.906	0.000
1/2, 1	2.657	0.000	**2.244**	0.261	**1.642**	0.162	**3.048**	0.000	**3.030**	0.028	3.031	0.019	3.262	0.000	3.262	0.000	**3.262**	0.000
3, 1/3	**4.041**	0.000	**3.867**	0.050	3.232	0.194	1.770	0.000	**1.769**	0.001	**1.767**	0.001	**1.263**	0.000	**1.263**	0.000	1.263	0.000
1, 1/3	**51.58**	0.000	**51.58**	0.004	50.99	0.189	30.13	0.000	**30.13**	0.000	**30.13**	0.000	**26.42**	0.000	**26.42**	0.000	26.42	0.000
1/2, 1/3	**23800**	0.000	**23800**	0.000	**23800**	0.129	**4009**	0.000	**4009**	0.000	**4009**	0.000	**5.562**	0.000	**5.562**	0.000	**5.562**	0.000

8. Illustrative examples

In this section, we give two illustrative examples of applying SMD estimation. The solid and dashed lines in Figures 2 and 3 respectively show the probability density functions estimated by the nonparametric approach with cross-validated bandwidth (NE1) and by the fitting to the GEV (PE). Figure 2 shows the Potomac River peak stream flow (cfs) data for water years 1901–2000 (Oct–Sep) at Point of Rocks, Maryland. Figure 3 shows the latest $n = 2100$ losses in Danish fire insurance over one million Danish kroner (DKK) collected from 1980 to 1990. The datasets are available in the extRemes and evir packages in the R software environment, respectively. The range of the horizontal axis in Figure 2 is 0–480000, where 480000 is the maximal value observed in 1936. The minimum value was 27800 observed in 1969. The range in Figure 3 is 0–(263.25×1.5), where 263.25 is the maximum and the minimum is 1.

Figure 2: The estimated probability ($\times 10^6$) density functions of Annual peak flow ($\times 10^{-5}$) data of the Potomac River at Point of Rocks, MD, USA, 1895–2000 by the log-transformed approach (solid line), the naive kernel approach (dotted line) and by the fitting to the GEV (dashed line).

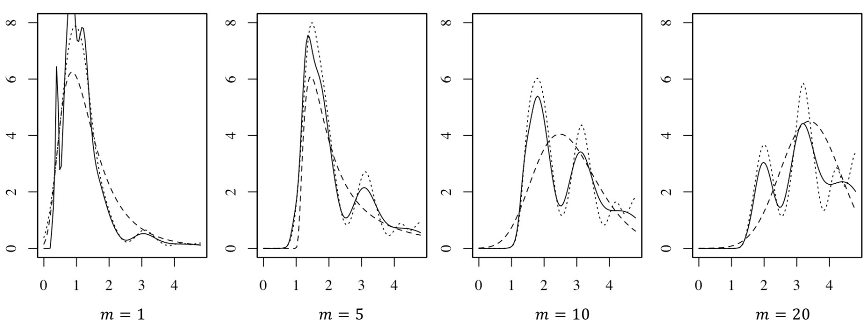
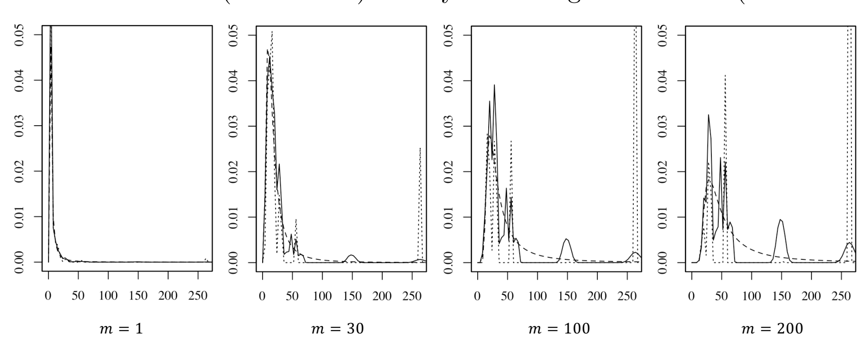


Figure 3: The estimated probability density functions of 2156 Danish data consisting of losses over one million Danish kroner (DKK) by the log-transformed approach (solid line), the naive kernel approach (dotted line) and by the fitting to the GEV (dashed line).



Figures 2 and 3 suggest the existence of multimodality. Moriyama (2021) [19] reported that, when comparing the estimated probabilities by the nonparametric approach and by the fitting to the GEV, the estimated risk is not one-sided. This may stems partly from the multimodality of the estimated density function. Note that the nonparametrically estimated density is sensitive to bandwidth. Many studies support the heavy-tailness in the Danish datasets, and so the statistical evidence of multimodality is not strong. Since the mixture of distributions that belong to the maximum domain of attractions converges to a unimodal GEV (see Bolancé et al. 2015 [6]), there are two possible reasons for the multimodality. One is the slowness of the convergence of the limitation to the GEV. The other is that the underlying distribution does not belong to the maximum domain of attractions. Future work should examine how to construct a method for testing the properties of underlying distributions including multimodality.

ACKNOWLEDGMENTS

The author appreciates the referee's valuable comments that helped me to improve this manuscript. This work was supported by the Japan Society for the Promotion of Science (JSPS) *JSPS KAKENHI Grant Numbers JP19K20223, JP23K16850*.

REFERENCES

[1] Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, **46**, 195–214.

[2] Banfi, F.; Cazzaniga, G. and De Michele, C. (2022). Nonparametric extrapolation of extreme quantiles: a comparison study, *Stochastic Environmental Research and Risk Assessment*, **36**(6), 1579–1596.

[3] Beirlant, J. and Devroye, L. (1999). On the impossibility of estimating densities in the extreme tail, *Statistics and Probability Letters*, **43**, 57–64.

[4] Beirlant, J.; Goegebeur, Y.; Teugels, J. and Segers, J. (2004). *Statistics of Extremes*. John Wiley and Sons, Ltd.

[5] Béranger, B.; Duong, T.; Perkins-Kirkpatrick, S. E. and Sisson, S. A. (2019). Tail density estimation for exploratory data analysis using kernel methods, *Journal of Nonparametric Statistics*, **31**, 1, 144–174.

[6] Bolancé, C.; Bahraoui, Z. and Alemany, R. (2015). Estimating extreme value cumulative distribution functions using bias-corrected Kernel approaches. *XREAP2015-01*

[7] Bowman, A. W.; Hall, P. and Prvan, T. (1998). Cross-validation for the smoothing of distribution functions. *Biometrika*, **85**, 799–808.

[8] Buch-Larsen, T.; Nielsen, J. P.; Guillén, M. and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, **39**(6), 503–516.

[9] Charpentier, A. and Flachaire, E. (2015). Log-Transform Kernel Density Estimation of Income Distribution, *L'Actualité Économique*, **91**, 141–159.

[10] Cohen, J. P. (1988). Fitting extreme value distributions to maxima, *Sankhyā: The Indian Journal of Statistics, Series A*, 74–97.

[11] Dombry, C. and Ferreira, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli*, **25**, 3, 1690–1723.

[12] Hall, P. and Welsh, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Annals of Statistics*, **12**, 3, 1079–1084.

[13] Hall, P. (1987). On Kullback-Leibler loss and density estimation. *Annals of Statistics*, **15**, 3, 1491–1519.

[14] MacDonald, A.; Scarrott, C. J.; Lee, D.; Darlow, B.; Reale, M. and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics and Data Analysis*, **55**, 6, 2137–2157.

[15] Maiboroda, R. E. and Markovich, N. M. (2004). Estimation of heavy-tailed probability density function with application to Web data. *Computational Statistics*, **19**, 4, 569–592.

[16] Markovich, N. (2007). *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice.* Wiley Series in Probability and Statistics. Chichester: Wiley.

[17] Markovich, L. (2018). Light- and heavy-tailed density estimation by gamma-Weibull kernel. In *Nonparametric Statistics. ISNPS 2016. Springer Proceedings in Mathematics & Statistics, vol. 250*, pp. 145–158.

[18] Marron, J. S. and Ruppert, D. (1994). Transformations to reduce bound- ary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B*, **56**, 653–671.

[19] Moriyama, T. (2021). Parametric and nonparametric probability distribution estimators of sample maximum. *arXiv:2111.03765.*

[20] Segers, J. (2005). Generalized Pickands estimators for the extreme value index. *Journal of Statistical Planning and Inference*, **128**, 2, 381–396.

[21] Sheather, S. J. and Jones, M. C. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society, Series B (Methodological)*, **53**, 3, 683–690.

[22] Silverman, B. W. (1986). *Density Estimation.* London: Chapman and Hall.

[23] Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis*, **144**, 1–24.

[24] Tokdar, S. T.; Jiang, S. and Cunningham, E. L. (2022). Heavy-tailed density estimation. *Journal of the American Statistical Association*, 1-13.

[25] Wadsworth, J. L.; Tawn, J. A. and Jonathan, P. (2010). Accounting for choice of measurement scale in extreme value modeling. *The Annals of Applied Statistics*, **4**, 3, 1558–1578.

[26] Wang, S.; Li, A.; Wen, K. and Wu, X. (2020). Robust kernels for kernel density estimation. *Economics Letters*, **191**, 109138.

[27] Wand, P.; Marron, J. and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association*, **86**, 343–361.