Asymptotic properties of parametric and nonparametric probability density estimators of sample maximum

Taku MORIYAMA
Department of Management of Social Systems and Civil Engineering, Tottori University

Abstract

Asymptotic properties of three estimators of probability density function of sample maximum \( f(m) := mF^{m-1} \) are derived, where \( m \) is a function of sample size \( n \). One of the estimators is the parametrically fitted by the approximating generalized extreme value density function. However, the parametric fitting is misspecified in finite \( m \) cases. The misspecification comes from mainly the following two: the difference \( m \) and the selected block size \( k \), and the poor approximation \( f(m) \) to the generalized extreme value density which depends on the magnitude of \( m \) and the extreme index \( \gamma \). The convergence rate of the approximation gets slower as \( \gamma \) tends to zero. As alternatives two nonparametric density estimators are proposed which are free from the misspecification. The first is a plug-in type of kernel density estimator and the second is a block-maxima-based kernel density estimator. Theoretical study clarifies the asymptotic convergence rate of the plug-in type estimator is faster than the block-maxima-based estimator when \( \gamma > -\frac{1}{2} \). A numerical comparative study on the bandwidth selection shows the performances of a plug-in approach and cross-validation approach depend on \( \gamma \) and are totally comparable. Numerical study demonstrates that the plug-in nonparametric estimator with the estimated bandwidth by either approach overtakes the parametrically fitting estimator especially for distributions with \( \gamma \) close to zero as \( m \) gets large.

Keywords: Extreme value; kernel type estimator; mean squared error; nonparametric estimation

1 Parametric density estimation of sample maximum

Let \( X_1, X_2, \ldots \) be independent and identically distributed random variables with a continuous distribution function \( F \), which belongs to either one of the Hall class, the Weibull class and the bounded class shown in the following assumption.

Assumption 1 It holds that either one of the followings (i) \( \exists \alpha > 0, \exists \beta > 0, \exists A > 0, \exists B \neq 0 \) s.t.

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

(ii) \( \exists \kappa > 0, \exists C > 0 \) s.t.

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

(iii) \( \exists \mu < -2, \sigma < 0, \exists D \neq 0, \exists E \neq 0, \exists x^* \in \mathbb{R} \) s.t.

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

The Pareto distribution, \( t \)-distribution, Burr distribution, and extreme value distribution with \( \gamma > 0 \) belong to the Hall class. The Hall class of distributions are heavily tailed. Methods for checking whether the data are heavily tailed are provided in Resnick [1]. The gamma, logistic, and log-normal distributions belong to the Weibull class of distributions. The uniform distributions, beta distributions and the reversed Burr distribution belong to the bounded class of distributions. The three class of distributions are widely used as
representatives of the distributions belonging to the domain of attractions of the extreme value distribution (see Beirlant et al. [2]; Segers [3]).

Under Assumption 1, it follows from the Fisher-Tippett-Gnedenko theorem that

\[ F^m(x) - G_\gamma \left( \frac{x - b_m}{a_m} \right) \to 0 \quad \text{as} \quad m \to \infty, \]

(1)

where \( G_\gamma \) is the generalized extreme value distribution, \( \gamma = \begin{cases} \alpha^{-1} & \text{for (i)} \\ 0 & \text{for (ii)} \\ \mu^{-1} & \text{for (iii)} \end{cases} \),

\[ a_m := \begin{cases} \gamma(Am)^\gamma & \text{for (i)} \\ \kappa^{-1}C^{-1/\kappa}(\ln m)^{-\theta} & \text{for (ii)} \\ -\gamma(Dm)^\gamma & \text{for (iii)} \end{cases} \]

\[ b_m := \begin{cases} (Am)^\gamma & \text{for (i)} \\ (C^{-1}\ln m)^{1/\kappa} & \text{for (ii)} \\ x^* - (Dm)^\gamma & \text{for (iii)} \end{cases} \]

and \( \theta := 1 - (1/\kappa) \). Moriyama [4] discussed the distribution estimation of the sample maximum \( \max \{X_{n+1}, \ldots, X_{n+m}\} \), i.e., \( F^m \). Let the continuous density \( f \) exist. Then, the sample maximum density \( f_{sm} \) is given by \( f_{sm} := mf^m \). We suppose \( m = m_n \to \infty \) as \( n \to \infty \) and investigate asymptotic properties of several \( smd \) estimators. Beirlant and Devroye [5] discussed non-zero lower bounds of the total variation and the supremum norm in \( smd \) estimation for general settings. Novak [6] proved that there are no consistent distribution estimators of sample maximum unless \( m/n = o(1) \) without any assumptions.

Since the convergence (1) is uniform (see Beirlant et al. [2]), the following convergence on \( smd \) holds

\[ f_{sm}(x) - g_{\gamma_m}(x) \to 0 \quad \text{as} \quad m \to \infty, \]

where

\[ g_{\gamma_m}(x) := \frac{1}{a_m} g_{\gamma} \left( \frac{x - b_m}{a_m} \right), \]

\[ g_{\gamma}(x) = \begin{cases} \{1 + \gamma x\}^{-(1/\gamma)-1} \exp(-\{1 + \gamma x\}^{-1/\gamma}), & 1 + \gamma x > 0, \quad \text{for (i),(iii)} \\ \exp(-x) \exp(-\exp(-x)), & x \in \mathbb{R}, \quad \text{for (ii)} \end{cases} \]

We estimate the parameters \( \gamma_m := (\gamma, a_m, b_m) \) by the maximum likelihood estimation based on the block maxima method (BM), where the block size is \( k \) and \( \exists N \in \mathbb{N} \) s.t. \( n = N \times k \), \( k \) can be different from \( m \); however, we do not consider block size selection.

**Assumption 2** Under Assumption 1, either \( (M_n \vee K_n) \to 0 \), \( (M_n \wedge K_n) \to \infty \), or \( \exists \delta > 0 \) s.t. \( M_n \to \delta \) and \( K_n \to \delta \) holds, where

\[ M_n := \begin{cases} Am x_n^{\alpha} & \text{for (i)} \\ m \exp(-Cx_n^\kappa) & \text{for (ii)} \\ Dm(x^* - x_n)^{-\mu} & \text{for (iii)} \end{cases} \]

\[ K_n := \begin{cases} Ak x_n^{\alpha} & \text{for (i)} \\ k\exp(-\kappa C^{1/\kappa}(\ln k)^{\theta} x_n) & \text{for (ii)} \\ Dk(x^* - x_n)^{-\mu} & \text{for (iii)} \end{cases} \]

**Proposition 1** Under Assumptions 1 and 2,

\[ \bar{\tau}_n := f_{sm}(x_n) - g_{\gamma_m}(x_n) \to 0. \]

**Assumption 3** Under Assumption 1, \( \exists \lambda \in \mathbb{R} \) s.t. \( \lambda_n \to \lambda \), where

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

2
Assumption 4 $N^{-1/2}k^{-\gamma}x_n \to 0$.

Set the MLE $\hat{\gamma}_k$. Regarding the asymptotic normality of the parametric estimator (PE) $g_{\hat{\gamma}_k}(x_n)$, the following theorem holds.

**Theorem 1** Under Assumptions 1, 2, 4 and 5,
\[
\sqrt{N}k^n(\tilde{\eta}_n^{-1}I_0^{-1}\eta_n)^{-1/2}\left\{ (f_m(x_n) - g_{\gamma_k}(x_n)) - (\tilde{\tau}_n + N^{-1/2}k^{-\gamma}\lambda_n\tilde{\eta}_n^{-1}I_0^{-1}b) \right\}
\]
converges in distribution to the standard normal, where $\tilde{\eta}_n := (\tilde{s}_n, \tilde{t}_n, \tilde{u}_n)^T$,
\[
\tilde{s}_n := K_n^{1+\gamma}\exp(-K_n) \{(1 - K_n)(1 - K_n^\gamma + \gamma \ln K_n) + \gamma (1 - K_n^\gamma)\},
\]
\[
\tilde{t}_n := K_n^{1+\gamma}\exp(-K_n)(K_n - 1) \left( K_n^\gamma - \frac{1}{\gamma} \right),
\]
\[
\tilde{u}_n := K_n^{1+\gamma}\exp(-K_n)K_n^\gamma(1 + \gamma - K_n).
\]

$b$ and the Fisher information matrix $I_0$ are given in Dombrzy and Ferrera [7].

This paper employs the symbols used in Moriyama [4] for clarity. We add a tilde on the symbols that are used in the same manner as in Moriyama [4] but are in fact different from those in Moriyama [4]. The following corollary on the convergence rate follows from Theorem 6. Proposition 3 and Theorem 10 are proved in the Appendices, and Corollary 7 is a direct consequence of Theorem 6.

**Corollary 1** Under the assumptions of Theorem 6, $(f_m(x_n) - g_{\gamma_k}(x_n))$ converges with the rate $(N^{-1/2}k^{-\gamma}\lambda_n\tilde{\eta}_n^{-1}1 + \tilde{\tau}_n + \zeta_n)$, where
\[
\zeta_n := N^{-1/2}k^{-\gamma} \times \begin{cases} K_n^{1+\gamma} \ln K_n & \text{for } (M_n \lor K_n) \to 0 \\ 1 & \text{for } M_n \to \delta, K_n \to \delta \\ K_n^{2(1+\gamma)} \exp(-K_n) & \text{for } (M_n \land K_n) \to \infty. \end{cases}
\]

The misspecification of the limitation of $smd$ to the generalized extreme value distribution appears in the three bias term $\lambda_n$, $\tilde{\tau}_n$ and $f_m(x_n) - g_{\gamma_m}(x_n)$. $\lambda_n$ corresponds to the asymptotic bias of $\hat{\gamma}_k$ which depends on the second order parameter. The bias $g_{\gamma_m}(x_n) - g_{\gamma_k}(x_n)$ comes from the asymptotic difference between $m$ and $k$. Though large $k$ yields large variance of $\hat{\gamma}_k$, the precise approximation of $g_{\gamma_m}(x_n)$ to $f_m(x_n)$ (i.e. $\tilde{\tau}_n$) needs large $m$. As the extreme index tends to zero i.e. $\gamma \to 0$, the convergence rate of the approximation of $g_{\gamma_m}(x_n)$ to $f_m(x_n)$ becomes slow. In the Weibull cases, the rate is slower than any polynomial. Therefore, the convergence rate is sensitive to the first order parameter $\gamma$. Numerical examples in some cases are given later in Table 1 in Section 4.

We next consider $smd$ estimation without limitation to the generalized extreme value distribution. By using the nonparametric approach, the plug-in type of the kernel density estimator of $smd$ defined as $f_m := m\tilde{f}F^{m-1}$ is constructed, where $\tilde{f}$ and $F$ are the kernel density and kernel distribution estimators of $f$ and $F$. We also consider the Block-Maxima-based (BM-based) kernel density estimator $f_m$ which is another candidate for the $smd$ estimator. The BM-based estimator is the kernel density estimator of $smd$ itself, which is a function of Block-Maxima. From the definition, we see that the nonparametric estimators are consistent for quite wide class of distributions in fixed $m$ cases, where PE is inconsistent. The nonparametric estimators can be good candidate for the $smd$ at least for relatively small $m$.

Two nonparametric estimators of $smd$ aim at the $smd$ itself directly, and they are free from the misspecification of the approximation of $g_{\gamma_m}(x_n)$ to $f_m(x_n)$ (i.e. $\tilde{\tau}_n$). We theoretically investigate asymptotic properties of the two kernel type of the distribution estimator. Section 2 and 3 gives asymptotic properties of the plug-in kernel density estimator $f_m$ and the Block-Maxima-based (BM-based) kernel density estimator $f_m$. The plug-in kernel density estimator is proved to be theoretically superior to the Block-Maxima-based (BM-based) kernel density estimator. We compare the numerical accuracies of PE $g_{\hat{\gamma}_k}$ and the plug-in estimator $f_m$ in Section 4. The proofs are given in Appendices.
2 Nonparametric plug-in type of estimation of smd

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

\[ \hat{f}(x; h_1, h_2) := m \hat{f}(x; h_1) \hat{F}^{m-1}(x; h_2) \]

where \( \hat{f} \) and \( \hat{F} \) are the kernel estimator given by

\[ \hat{f}(x; h_1) = \frac{1}{nh_1} \sum_{i=1}^{n} k \left( \frac{x - X_i}{h_1} \right), \quad \hat{F}(x; h_2) = \frac{1}{n} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_2} \right), \]

where \( k \) is a symmetric and bounded density function whose support is bounded for (iii). \( K \) is the cumulative distribution function, \( h_j \) are bandwidths that satisfy \( nh_j \to \infty \), \( h_j \to 0 \) for \( j = 1, 2 \) and \( h_j(x^* - x_n) \to 0 \) under Assumption 3 for \( j = 1, 2 \).

**Assumption 5** \( f \) is twice continuously differentiable at \( x_n \).

Assumption 8 for the bounded class of distributions requires \( \mu \) is integer or \( \mu < -3 \).

**Assumption 6**

\[ \int z^2k(z)dz < \infty, \quad \int k^2(z)dz < \infty, \quad \int zK(z)k(z)dz < \infty. \]

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

**Theorem 2** Given Assumption 1, 8 and 9 If \( b_{n,1} \to 0 \) and \( v_{n,1} \to 0 \), then \( v_{n,1}^{1/2} \{(f_{(m)}(x_n) - \hat{f}_{(m)}(x_n; h_1, h_2)) - b_{n,1} \} \) converges in distribution to the standard normal, where

\[ b_{n,1} := \frac{h_1^2}{2} \exp(-M_n)M_n^{1+3\gamma}m^{-3\gamma}\psi_n \int z^2k(z)dz + \frac{h_2^2}{2}M_n m \xi_n f(x_n) \int z^2k(z)dz, \]

\[ v_{n,1} := \frac{m^2}{nh_1} \exp(-2M_n)f(x_n) \int k^2(z)dz + \frac{m^2}{n}M_n^2f^2(x_n)(M_n^{-1} - 2h_2\omega_n \int zK(z)k(z)dz), \]

\[ \psi_n := \begin{cases} \alpha(\alpha + 1)(\alpha + 2)A^{-3\gamma} & \text{for (i)} \\ \kappa^3C^3x_n^{3\kappa-3} & \text{for (ii)} \\ -\mu(\mu + 1)(\mu + 2)D^{-3\gamma} & \text{for (iii)}, \end{cases} \]

\[ \xi_n := \begin{cases} \alpha(\alpha + 1)x_n^{-2} & \text{for (i)} \\ \kappa^2C^2x_n^{2\kappa-2} & \text{for (ii)} \\ -\mu(\mu + 1)(x^* - x_n)^{-2} & \text{for (iii)} \\ A^{-1}\alpha x_n^{\alpha-1} & \text{for (i)} \end{cases} \]

\[ \omega_n := \begin{cases} \kappa C x_n^{\kappa-1} \exp(Cx_n^\kappa) & \text{for (ii)} \\ -D^{-1}\mu(x^* - x_n)^{\mu-1} & \text{for (iii)}. \end{cases} \]

**Corollary 2** The asymptotically optimal values of \( h_1 \) and \( h_2 \) are given by

\[ \left( M_n^{-2-6\gamma}m^{2+6\gamma}\psi_n^{-2}f(x_n)n^{-1} \left( \int z^2k(z)dz \right)^{-2} \int k^2(z)dz \right)^{1/5} \]

and

\[ \left( 2\xi_n^{-2}\omega_n^{-1} \left( \int z^2k(z)dz \right)^{-2} \int zK(z)k(z)dz \right)^{1/3}, \]
which are asymptotically identical to those of \( \hat{f}(x_n; h_1) \) and \( \hat{F}(x_n; h_2) \). Under the assumption of Theorem

\[ f_{(m)}(x_n) - \hat{f}_{(m)}(x_n; h_1, h_2) \]

with the optimal bandwidths is asymptotically non-degenerate normal with the following asymptotic mean:

\[
\nu_1 \exp(-M_n)M_n^{(1+3\gamma)/5}m^{(4-3\gamma)/5}\psi_n^{1/5}f_n^{2/5}(x_n)n^{-2/5} + \nu_2M_n\xi_n^{-1/3}f_n\omega_n^{2/3}n^{-2/3}
\]

where

\[
\nu_1 := \frac{1}{2} \left( \int z^2 k(z) dz \right)^{1/5} \left( \int k^2(z) dz \right)^{2/5}, \nu_2 := \frac{2}{\int z^2 k(z) dz}^{1/3} \left( \int zK(z)k(z)dz \right)^{2/3}.
\]

**Corollary 3** The asymptotically optimal bandwidths for \( M_n = O(1) \) are of the order

\[
\left( \frac{m}{n} \right)^{1/5} \times \left\{ \frac{m^\gamma}{\{C^{-1}\ln m\}^{-\theta}} \right\} \quad \text{for (i), (ii)}
\]

and

\[
\left( \frac{m}{n} \right)^{1/3} \times \left\{ \frac{m^\gamma}{\{C^{-1}\ln m\}^{-\theta}} \right\} \quad \text{for (i), (iii)}
\]

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

\[
\left( \frac{m}{n} \right)^{2/5} \left\{ \frac{m^{-\gamma}}{(\ln m)^\theta} + \left( \frac{m}{n} \right)^{2/3} \times \left\{ \frac{m^{-\gamma}}{(\ln m)^\theta} \right\} \right\} \quad \text{for (i), (iii)}
\]

The proof of Theorem 10 is found in Appendices. Corollary 11 and 12 are directly obtained from Theorem 10.

### 3 Nonparametric block-maxima-based estimation of smd

Suppose that the block size is \( m \) and \( \frac{3}{\bar{n}} \in \mathbb{N} \) s.t. \( n = \bar{n} \times m \). We consider the following BM-based kernel-type of nonparametric density estimator (NE2)

\[
\hat{f}_{(m)}(x) = \frac{1}{\bar{n}h} \sum_{i=1}^{\bar{n}} k \left( \frac{x - Y_i}{h} \right).
\]

where \( Y_j := \max\{X_{m(j-1)+1}, X_{m(j-1)+2}, \cdots, X_{mj}\} \space (j = 1, \cdots, \bar{n}) \). The kernel function \( k \) is a symmetric and bounded density function whose support is bounded for (ii). \( h \) is a bandwidth that satisfies all \( h \to 0 \), \( \bar{n}h \to \infty \) and \( h(x^* - x_n) \to 0 \) under Assumption 3. The following theorem on the asymptotic normality of the naive nonparametric density estimator holds.

**Theorem 3** Given Assumption 1, 8 and 9. If \( x_n^{\kappa-1}h \to 0 \),

\[
b_{n,2} := \frac{mh^2}{2} f(x_n) \phi_n \int z^2 k(z) dz \to 0
\]

and \( v_{n,2} := (\bar{n}h)^{-1} f_{(m)}(x_n) \int k^2(z) dz \to 0 \), then \( v_{n,2}^{1/2} \{(f_{(m)}(x_n) - \hat{f}_{(m)}(x_n; h)) - b_{n,2}\} \) converge in distribution to the standard normal, where

\[
\phi_n := \left\{ \frac{m^2}{(\kappa C)^2 x_n^{2(\kappa-1)} - 2mC x_n^{\kappa-1} + m^2} \right\} \quad \text{for (i), (iii)}
\]

\[
\phi_n := \left\{ \frac{m^2}{(\kappa C)^2 x_n^{2(\kappa-1)} - 2mC x_n^{\kappa-1} + m^2} \right\} \quad \text{for (ii)}
\]
Corollary 4 The asymptotically optimal value of $h$ is given by

$$
\left(n^{-1}\{f(x_n)\}^{-1} \phi_n^{-2} \left( \int z^2 k(z) dz \right)^{-2} \int k^2(z) dz \right)^{1/5}.
$$

Under the assumption of Theorem 13, $f(m(x_n)) - \hat{f}(m(x_n); h)$ with the optimal bandwidths is asymptotically non-degenerate normal with the following asymptotic mean:

$$
\nu_n^{-2/5} m\{f(x_n)\}^{3/5} \phi_n^{1/5}.
$$

Corollary 5 The asymptotically optimal bandwidth for $m = O(1)$ is of the order

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

Under the assumption of Theorem 13, $\hat{f}(m(x_n))$ has the asymptotic bias of the order:

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

Theorem 13 is proved in the Appendices. Corollaries 14 and 15 are the direct consequences.

4 Comparative study

Performances of the three estimators PE, NE1 and NE2 are compared in this section. The first is theoretically and the second is numerically. $M_n \equiv K_n \equiv \sigma$ is supposed throughout this section. NE1 and NE2 with their respective optimal bandwidths converge on the order of $m^{-2\gamma}(m/n)^{4/5}$ and $m^{-(6/5)\gamma}(m^2/n)^{4/5}$. The following Theorem 16 is immediately seen from Theorems 10 and 13.

Theorem 4 The convergence rate of NE1 $\hat{f}(m(x_n); h_1, h_2)$ with the optimal bandwidths is faster than NE2 $\hat{f}(m(x_n))$ with the optimal bandwidth for $\gamma > -1$.

PE converges with the rate $m^{-2\gamma} \times (N^{-1}m^{2-4\beta} + m^{-2\gamma} + m^{-2} + N^{-1})$ for the Hall class. For the Weibull class, PE does not satisfy both Assumption 2 and 4 with any block size $k$ and is inconsistent. For the bounded class, PE has the convergence rate $m^{-2\gamma} \times (N^{-1}m^{2+4\sigma} + m^{-2\gamma} + m^{-2} + N^{-1})$. Assumption 5 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 the estimators without terms slower than any polynomial are summarized in Table 1. The hyphens mean the distribution breaks the assumption of the theorems. The settings in the density estimation and distribution estimation surveyed in Moriyama [4] are same. The difference is all the target values $f(m(x_n))$ tend to zero with the exception of the bounded class with $\mu = -1$ while all $F^m(x_n) \rightarrow \exp(-\delta) > 0$ in distribution estimation. The convergence rates divided by $f(m(x_n))$ are given in Table 2, where the hyphens are also used to mean that the estimators are inconsistent. All the convergence rates of NE2 are slower than or same as those of NE1 in the Table 2 as Theorem 16 stated. While for $m = n^{1/4}$ all the rates of NE1 are faster than PE, the differences become small as $m$ gets large, which coincides with the results in distribution estimation given in Moriyama [4]. Both the rates of PE and NE1 become slow as $\gamma$ gets small.

By simulating the mean integrated squared error ($MISE$) of PE $g_{\eta_n}$:

$$
L_m \int_{Q_m(0.1)}^{Q_m(0.9)} \left( g_{\eta_n}(x) - f(m(x)) \right)^2 dx.
$$
and that of \( \hat{f}(m) \), we studied the numerical accuracy in small-sample cases, where \( L_m := Q_m(0.9) - Q_m(0.1) \) and \( Q_m(q) \) denotes the \( q \)th quantile of the \( smd \). Tables 3–4 show the mean values and their standard deviation (sd) of the obtained MISE values by simulating 1000 times. The underlying distributions \( F \) were Pareto distributions, \( T \)-distributions, Burr distributions, Fréchet distributions, Weibull distributions and reversed Burr distributions. The forecast periods were \( m = n^{1/4} \), \( m = n^{1/2} \), and \( m = n^{3/4} \). All kernel functions were the Epanechnikov for the bounded class and the 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 in the Tables ‘CV’ and ‘PI’ respectively. CV means the unbiased cross-validation estimator (see e.g. Silverman [8]) and Bowman et al [9]’s cross-validation estimator. As the plug-in approaches, the Sheather and Jones [10]’s method and the Altman and Leger [11]’s method were employed. The sample sizes were \( (n = )2^8 \) in Tables 3 and \( (n = )2^{12} \) in Tables 4.

Comparing CV with PI, we see CV greatly outperforms for \( 1 \leq \gamma \leq 2 \) while PI is much better for \( \gamma \geq 4 \). In the other cases, they are numerically comparable. Comparing NE1 with PE, we find that the MISE values of PE were smaller in most of the cases with \( m = n^{1/4} \). For \( \gamma \approx 0 \) those of NE1 were smaller as \( m \) get large. For the light-tailed distributions with \( \gamma < 0 \), the MISE values of NE1 were comparable with those of PE, though so-called boundary bias in NE1 for bounded distributions with \( \gamma \geq -1 \) was observed. Due to the boundary bias, the results especially for distributions with large extreme indices (e.g. \( \gamma \geq -1/3 \)) are different from those of distribution estimation shown in Moriyama [4]. Comparing the cases \( (n = )2^{12} \) with \( (n = )2^8 \), it is seen that PE substantially reduces the MISE values, which are smaller than those of NE1.

References

[1] Resnick, S. I. (1997), Discussion of the danish data on large fire insurance losses. ASTIN Bulletin, 27(1), 139–151.
[2] Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004), Statistics of Extremes. John Wiley & Sons, Ltd.
[3] Segers, J. (2005), Generalized Pickands estimators for the extreme value index. Journal of Statistical Planning and Inference, 128(2), 381–396.
[4] Moriyama, T. (2021), Parametric and nonparametric probability distribution estimators of sample maximum. arXiv:2111.03765
[5] Beirlant, J. Devroye, L. (1999), On the impossibility of estimating densities in the extreme tail, Statistics & Probability Letters, 43, 57–64.
[6] Novak, S. Y. (2009), Impossibility of consistent estimation of the distribution function of a sample maximum. Statistics, 44(1), 25–30.
[7] Dombry, C. and Ferreira, A. (2019), Maximum likelihood estimators based on the block maxima method. Bernoulli, 25(3), 1690–1723.
[8] Silverman, B. W. (1986), Density Estimation. London: Chapman and Hall.
[9] Bowman, A. W., Hall, P. and Prvan, T. (1998), Cross-validation for the smoothing of distribution functions. Biometrika, 85, 799–808.
[10] 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.
[11] Altman, N., Leger, C. (1995), Bandwidth selection for kernel distribution function estimation. Journal of Statistical Planning and Inference 46, 195–214.
| Table 1: The polynomial convergence rates of MSE of the estimators |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | \( f \)         | \( \alpha \)    | \( \beta \)    | \( m = n^{1/4} \) | \( m = n^{1/2} \) | \( m = n^{2/4} \) |
| **Pareto**      |                 |                 |                 |                 |                 |                 |
| \( 1/2 \)       | 1/2            | 1               | -1/2            | \(-8/5\)        | \(-1\)          | \(-12/5\)       |
| \( 1 \)         | 1              | 1               | -1/2            | \(-11/10\)       | -1/5            | \(-12/5\)       |
| \( 3 \)         | 3              | 1               | -1/2            | \(-23/30\)       | -1/2            | \(-2/3\)        |
| 10              | 10             | 1               | -1/10           | \(-13/20\)       | \(-1/5\)        | \(-3/50\)       |
| **Tdist**       |                 |                 |                 |                 |                 |                 |
| \( 1/2 \)       | 1/2            | 2               | -1/2            | \(-8/5\)        | \(-1\)          | \(-12/5\)       |
| \( 1 \)         | 1              | 2               | -1/2            | \(-11/10\)       | -1/2            | \(-2/3\)        |
| 10              | 10             | 2               | -1/10           | \(-13/20\)       | \(-1/5\)        | \(-3/50\)       |
| **Burr**        |                 |                 |                 |                 |                 |                 |
| \( c, \ell \)   | 1/2            | 1/2             | 1/2             | \(-8/5\)        | \(-9/2\)        | \(-24/15\)      |
| \( 1 \)         | 1              | 1               | -1/2            | \(-8/5\)        | \(-3\)          | \(-12/5\)       |
| 1/2             | 1/2            | 1               | -1/2            | \(-11/10\)       | -1/5            | \(-2/3\)        |
| 3               | 3              | 3               | -1/3            | \(-23/30\)       | \(-1\)          | \(-3/50\)       |
| **Frechet**     |                 |                 |                 |                 |                 |                 |
| \( \gamma \)    | 1/2            | 1/2             | 1/2             | \(-8/5\)        | \(-3/2\)        | \(-12/5\)       |
| \( 1 \)         | 1              | 1               | -1/2            | \(-11/10\)       | -1/5            | \(-3/50\)       |
| 1/2             | 1/2            | 1               | -1/2            | \(-17/20\)       | \(-1\)          | \(-3/50\)       |
| 1/4             | 1/4            | 1               | -1/4            | \(-29/40\)       | \(-1/2\)        | \(-3/50\)       |
| **Weibull**     |                 |                 |                 |                 |                 |                 |
| \( k, \rho \)   | 1/2            | 0               | 0               | \(-3/5\)        | \(-1\)          | \(-2\)          |
| 1               | 1              | 0               | 0               | \(-3/5\)        | \(-1\)          | \(-2\)          |
| 1/2             | 1/2            | 1               | 1               | \(-3/5\)        | \(-1\)          | \(-2\)          |
| 1/4             | 1/4            | 1               | 1               | \(-3/5\)        | \(-1\)          | \(-2\)          |
| **rev.Burr**    |                 |                 |                 |                 |                 |                 |
| \( c, \ell \)   | 1/2            | -1/3            | -6              | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        |
| \( 1 \)         | 1/3            | -1              | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        |
| \( 1/2 \)       | -1/3           | -1              | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        |
| \( 3, -1 \)     | -1/3           | -1/3            | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        |
| \( 1/2, -2 \)   | -1/3           | -1              | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        |
| \( -1 \)        | -1/2           | -1/2            | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        |
| \( -1 \)        | -1/2           | -1/2            | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        |
| \( -3, -2 \)    | -1/3           | -1/3            | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        | \(-1/2\)        |
Table 2: The normalized polynomial convergence rates of MSE of the estimators

| Parameter | $m = n^{1/4}$ | $m = n^{1/2}$ | $m = n$ |
|-----------|----------------|----------------|---------|
| $\ell$    | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ |
| 1/2       | 1/2     | 1/2 | -1/4 | -1/2 | -3/2 | -1/2 | -3/2 | -1/2 |
| 3         | 3/2     | 3/2 | -1/4 | -5/8 | -5/8 | -5/8 | -5/8 | -5/8 |
| $dist.$   | $o_1$ | $\beta$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ |
| 1/2       | 1/2     | 2/2 | 1/4  | -1/3 | -3/2 | -1/2 | -3/2 | -1/2 |
| 3         | 3/2     | 3/2 | -1/4 | -41/60 | -5/12 | -1/2 | -17/30 | -3/10 |
| $c_1, \ell$ | $c_1, \ell$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ |
| 1/4, 1/2 | 1/4     | 1/2 | 1/4  | -3/2 | -3/2 | -5/8 | -5/8 | -5/8 |
| 1/2, 1/2 | 1/2     | 1/2 | -1/2 | -5/8 | -5/8 | -5/8 | -5/8 | -5/8 |
| $\gamma$  | $\alpha$ | $\beta$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ |
| 1/2       | 1/2     | 1/2 | -1/2 | -5/8 | -5/8 | -5/8 | -5/8 | -5/8 |
| 1/4       | 1/4     | 1/2 | -1/2 | -3/2 | -3/2 | -3/2 | -3/2 | -3/2 |
| $K_{weibull}$ | $\omega_1$ | $\beta_1$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ |
| 1/2       | 1/2     | 1/2 | -1/2 | -3/2 | -3/2 | -3/2 | -3/2 | -3/2 |
| 1/4       | 1/4     | 1/2 | -1/2 | -3/2 | -3/2 | -3/2 | -3/2 | -3/2 |
| rev. $Burr$ | $c_1, \ell$ | $\mu$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ |
| 1/2, -1/3 | -6     | -2 | -1/8 | -20/4 | -17/30 | -1/2 | -17/30 | -1/2 |
| 3, -1/3  | -3     | -1 | -1  | -17/30 | -3/8 | -3/8 | -3/8 | -3/8 |
| $c_1, \ell$ | $\sigma$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ | $n_{E1}$ | $n_{E2}$ |
| 1/2, -1/2 | -2     | -2 | -1/2 | -1/2 | -1/2 | -1/2 | -1/2 | -1/2 |
| k | m | m = \frac{1}{n} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} |
|---|---|---|---|---|---|---|---|---|
| 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 |
| 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
| 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 |

| k | m | m = \frac{1}{n} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} |
|---|---|---|---|---|---|---|---|---|
| 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 |
| 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
| 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 |

| k | m | m = \frac{1}{n} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} |
|---|---|---|---|---|---|---|---|---|
| 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 |
| 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
| 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 |

| k | m | m = \frac{1}{n} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} |
|---|---|---|---|---|---|---|---|---|
| 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 |
| 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
| 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 |

| k | m | m = \frac{1}{n} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} | \frac{\bar{m} - m}{\sqrt{\text{CV}(m)}} |
|---|---|---|---|---|---|---|---|---|
| 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 | 0.952 |
| 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
| 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 |
## Table 4: Scaled MISE values and sd values where $n = 212$

| Pareto | $m = n^{1/6}$ | $m = n^{1/7}$ | $m = n^{1/8}$ | $m = n^{1/9}$ | $m = n^{1/10}$ |
|--------|---------------|---------------|---------------|---------------|---------------|
| $t$    | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  |
| 1/4    | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 1      | 0.004 | 0.004 | 0.004  | 0.004 | 0.004 | 0.004 | 0.004  | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| 3      | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| 10     | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |

| Tfrist | $m = n^{1/4}$ | $m = n^{1/6}$ | $m = n^{1/7}$ | $m = n^{1/8}$ | $m = n^{1/9}$ |
|--------|---------------|---------------|---------------|---------------|---------------|
| $t$    | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  |
| 1/4    | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 1      | 0.005 | 0.005 | 0.005  | 0.005 | 0.005 | 0.005 | 0.005  | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| 3      | 0.003 | 0.003 | 0.003  | 0.003 | 0.003 | 0.003 | 0.003  | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| 10     | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |

| Burr | $m = n^{1/4}$ | $m = n^{1/6}$ | $m = n^{1/7}$ | $m = n^{1/8}$ | $m = n^{1/9}$ |
|------|---------------|---------------|---------------|---------------|---------------|
| $t$  | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  |
| 1/4  | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 1    | 0.005 | 0.005 | 0.005  | 0.005 | 0.005 | 0.005 | 0.005  | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| 3    | 0.003 | 0.003 | 0.003  | 0.003 | 0.003 | 0.003 | 0.003  | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| 10   | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |

| Fréchet | $m = n^{1/4}$ | $m = n^{1/6}$ | $m = n^{1/7}$ | $m = n^{1/8}$ | $m = n^{1/9}$ |
|--------|---------------|---------------|---------------|---------------|---------------|
| $t$    | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  |
| 5     | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 1     | 0.004 | 0.004 | 0.004  | 0.004 | 0.004 | 0.004 | 0.004  | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| 3     | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| 5     | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |

| Weibull | $m = n^{1/4}$ | $m = n^{1/6}$ | $m = n^{1/7}$ | $m = n^{1/8}$ | $m = n^{1/9}$ |
|--------|---------------|---------------|---------------|---------------|---------------|
| $t$    | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  |
| 1/2   | 0.007 | 0.007 | 0.007  | 0.007 | 0.007 | 0.007 | 0.007  | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 |
| 1     | 0.004 | 0.004 | 0.004  | 0.004 | 0.004 | 0.004 | 0.004  | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| 2     | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| 10    | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |

| rev.Burr | $m = n^{1/4}$ | $m = n^{1/6}$ | $m = n^{1/7}$ | $m = n^{1/8}$ | $m = n^{1/9}$ |
|---------|---------------|---------------|---------------|---------------|---------------|
| $t$     | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  | PE  | sd  | CV  | sd  |
| 1/2, 1/3 | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001  | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 1      | 0.004 | 0.004 | 0.004  | 0.004 | 0.004 | 0.004 | 0.004  | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| 3      | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| 10     | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002  | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |