Yousri Slaoui*

Optimal bandwidth selection for recursive Gumbel kernel density estimators

Abstract: In this paper, we propose a data driven bandwidth selection of the recursive Gumbel kernel estimators of a probability density function based on a stochastic approximation algorithm. The choice of the bandwidth selection approaches is investigated by a second generation plug-in method. Convergence properties of the proposed recursive Gumbel kernel estimators are established. The uniform strong consistency of the proposed recursive Gumbel kernel estimators is derived. The new recursive Gumbel kernel estimators are compared to the non-recursive Gumbel kernel estimator and the performance of the two estimators are illustrated via simulations as well as a real application.

1 Introduction

In probability theory and statistics, the Gumbel distribution (called also Generalized Extreme Value distribution Type-I) introduced by [11, 12] is one of the most used parametric model in extreme value theory. This distribution can be used, for example, to represent the distribution of the maximum level of a watercourse. It can also be used to predict the probability of an extreme earthquake occurring. See [16] for a variety of Gumbel model applications.

Many approaches have been considered in the parametric framework, we can cite the work of [17], in which the model parameters were estimated using an inferential procedure based on likelihood, we can list also the work of [9], in which the parameters of the Gumbel function were estimated using the mode and the moments of the distributions. It should be noted that all these methods involve the estimation of unknown parameters. However, the parametric model may be limited for the analysis of the complex distribution of extreme values, and may give worse results for small and medium-sized samples. As an alternative to the Gumbel parametric model, we propose to use a non-parametric kernel method.

In the literature there has been some attention on non-parametric procedures for the tail density estimation. [6] proposes an estimator using the Champernowne transformation distribution and the classical kernel density estimator. [18] investigates kernel density estimation and its application to the Gumbel probability distribution. [4] considers an estimator based on classical transformation methods, the book includes real-life examples. [35] proposes two kernel density estimators based on a bias reduction technique. More recently, [2] develops a transformation kernel density estimator which is able to handle heavy tailed and bounded data.

The first objective of this article is to propose a recursive non-parametric Gumbel kernel estimators. The second objective is to establish the statistical properties of the new estimators and compare them with the non-recursive version of Gumbel kernel estimator and with the conventional kernel density estimators developed respectively by [26] and [23] for the non-recursive estimator and [21] and [28, 29] for the recursive estimator by using a Gaussian kernel and through a plug-in bandwidth selection.

*Corresponding Author: Yousri Slaoui: Université de Poitiers, E-mail: Yousri.Slaoui@math.univ-poitiers.fr

The layout of the present paper is as follows. Section 2 is devoted to introduce our proposed recursive Gumbel estimators. We establish the statistical properties of the proposed estimators and we develop a plug-in bandwidth selection approach in Section 3. Section 4 is dedicated to our application results, first by simulation (subsection 4.1) and second using a real dataset (subsection 4.2). We conclude the article in Section 5. To avoid interrupting the flow of this paper, all mathematical developments are relegated to Appendix A.

2 Recursive Gumbel kernel estimators

Let $X_1, \ldots, X_n$ be independent, identically distributed random variables, and let f be the probability of X. In order to construct a stochastic algorithm to recursively estimate the unknown density function of probability f at a point x, we consider at the beginning an algorithm of search of the zero of the function $h : y \to f(x) - y$. Then, we consider Robbins-Monro's procedure, the proposed algorithm is defined by setting $f_0(x) \in \mathbb{R}$, and, for all $n \geq 1$,

$$f_n^{\text{Gum}}(x) = f_{n-1}^{\text{Gum}}(x) + \beta_n T_n^{\text{Gum}}(x),$$

where $T_n^{\text{Gum}}(x)$ is considered as an observation of the function h at the point $f_{n-1}^{\text{Gum}}(x)$ and the stepsize (β_n) is a sequence of positive real numbers that goes to zero. To construct $T_n^{\text{Gum}}(x)$, we follow the approach of [24, 25], [37] and of [28, 30, 34], which are based on the classical properties of stochastic algorithms which is ($\mathbb{E}\left[T_n^{\text{Gum}}(x) | \mathcal{F}_{n-1}\right] = 0$, where $\mathcal{F}_{n-1}$ stands for the σ-algebra of the events occurring up the time $n - 1$) and we use the Gumbel kernel $K^{\text{Gum}}(z) = \exp(-z - \exp(-z))$, and a bandwidth (h_n) (that is, a sequence of positive real numbers that goes to zero), and set $T_n^{\text{Gum}}(x) = h_n^{-1} K^{\text{Gum}}\left(h_n^{-1}(X_n - x)\right) - f_{n-1}^{\text{Gum}}(x)$. Then, the estimator f_n^{Gum} to recursively estimate the probability density function using the Gumbel kernel at x can be written as

$$f_n^{\text{Gum}}(x) = (1 - \beta_n) f_{n-1}^{\text{Gum}}(x) + \beta_n h_n^{-1} K^{\text{Gum}}\left(h_n^{-1}[X_n - x]\right). \tag{1}$$

This recursive property (1) is very useful with a large sample size since f_n^{Gum} can be updated with each additional observation. In fact, if X_n is a new observation, the estimators $f_n^{\text{Gum}}(x)$ can be updated recursively by relation (1). From a computational cost point of view, this arrangement provides important savings in computational time and storage memory, which is a consequence of the fact that the estimate updating is independent of the history of the data. The main drawback of the non-recursive kernel estimators is the use of all data at each step of estimation. One can check that, if we consider $f_0^{\text{Gum}}(x) = 0$ and we let $Q_n = \prod_{j=1}^{n}\left(1 - \beta_j\right)$, then the last equation (1) can be written as follows:

$$f_n^{\text{Gum}}(x) \quad = \quad Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k h_k^{-1} K^{\text{Gum}}\left(\frac{X_k - x}{h_k}\right). \tag{2}$$

This equation will be very useful to construct a data-driven bandwidth selection. Moreover, we show that the optimal bandwidth which minimizes $\mathbb{E}\int_{\mathbb{R}}\left[f_n^{\text{Gum}}(x) - f(x)\right]^2 dx$ depends on the choice of the stepsize (β_n); more precisely, we show that under some conditions of regularity of the unknown density of probability function f and using the special stepsizes $(\beta_n) = \left(n^{-1}\right)$, the bandwidth (h_n) will be equal to

$$\left(\left(\frac{1}{24\gamma_0^2}\right)^{1/3}\left\{\frac{1}{\int_{\mathbb{R}}\left(f^{(1)}(x)\right)^2 dx}\right\}^{1/3} n^{-1/3}\right),$$

where γ_0 is the Euler-Mascheroni constant ($\gamma_0 \simeq 0.57721$). Then, in this paper we propose an automatic selection of such bandwidth through a plug-in method, and then give the conditions under which the recursive estimator f_n^{Gum} will be approximately similar to the non-recursive Gumbel density estimator defined as

$$\widetilde{f}_n^{\text{Gum}}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K^{\text{Gum}}\left(\frac{X_i - x}{h_n}\right). \tag{3}$$

The obtained results presented in Section 4 corroborate our theoretical results given in the next section.

3 Assumptions and main results

We define the following class of regularly varying sequences.

Definition 1. Let $\gamma \in \mathbb{R}$ and $(v_n)_{n \geq 1}$ be a nonrandom positive sequence. We say that $(v_n) \in \mathcal{GS}(\gamma)$ if

$$\lim_{n \to +\infty} n \left[1 - \frac{v_{n-1}}{v_n} \right] = \gamma. \tag{4}$$

Condition (4) was introduced by [10] to define regularly varying sequences (see also [3]) and by [20] in the context of stochastic approximation algorithms. Noting that the acronym $\mathcal{GS}$ stands for (Galambos and Seneta). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^\gamma (\log n)^b$, $n^\gamma (\log \log n)^b$, and so on.

In this section, we investigate asymptotic properties of the proposed estimators (1). The assumptions to which we shall refer to throughout this paper are the following:

(A1) *i)* $(\beta_n) \in \mathcal{GS}(-\beta)$ with $\beta \in (1/2, 1]$.
  ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in (0, 1)$.
  iii) $\lim_{n \to \infty} (n\beta_n) \in (\min \{a, (\beta - a)/2\}, \infty]$.
(A2) f is bounded, differentiable and $f^{(1)}$ is bounded.

Assumption $(A1)(iii)$ on the limit of $(n\beta_n)$ as n goes to infinity is standard in the framework of stochastic approximation algorithms. It implies in particular that the limit of $\left([n\beta_n]^{-1} \right)$ is finite. For simplicity, we introduce the following notations:

$$\xi = \lim_{n \to \infty} (n\beta_n)^{-1}, \tag{5}$$

$$I_1 = \int_{\mathbb{R}} f^2(x)\, dx, \qquad I_2 = \int_{\mathbb{R}} \left(f^{(1)}(x) \right)^2 f(x)\, dx,$$

where $L^{(1)}(x)$ is the first derivative of the function L at a point x. In this section, we explicit the choice of the optimal bandwidth (h_n) through a plug-in method, based on the minimization the Mean Weighted Integrated Squared Error *MWISE* of our proposed recursive Gumbel kernel density estimators f_n^{Gum} defined in (1). Moreover, we provide a comparison between our proposed estimators with the non-recursive Gumbel kernel density estimator (3) and with the conventional Gaussian kernel density estimators (see, [29]). In addition, it has been demonstrated in [21] and considered in [28] that to minimize the Mean Integrated Squared Error *MISE* of f_n^{Gum} ($MISE \left[f_n^{\mathrm{Gum}} \right] = \mathbb{E} \int_{\mathbb{R}} \left[f_n^{\mathrm{Gum}}(x) - f(x) \right]^2 dx$), the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$ and must satisfy $\lim_{n \to \infty} n\beta_n = 1$. We consider here the case $(\beta_n) = (n^{-1})$.

Our first result is the following proposition, which gives the bias and the variance of f_n^{Gum}.

Proposition 1 (Bias and variance of f_n^{Gum}). Let Assumptions $(A1) - (A2)$ hold.

$$\mathbb{E} \left[f_n^{\mathrm{Gum}}(x) \right] - f(x) = \frac{\gamma_0}{(1 - a\xi)} f^{(1)}(x) [1 + o(1)] h_n \mathbb{1}_{\{a \leq \beta/3\}} + \mathbb{1}_{\{a > \beta/3\}} o \left(\sqrt{\beta_n h_n^{-1}} \right). \tag{6}$$

$$Var \left[f_n^{\mathrm{Gum}}(x) \right] = \frac{1}{4 (2 - (\beta - a)\xi)} f(x) [1 + o(1)] \frac{\beta_n}{h_n} \mathbb{1}_{\{a > \beta/3\}} + \mathbb{1}_{\{a \leq \beta/3\}} o \left(\frac{\beta_n}{h_n} \right). \tag{7}$$

Remark 1.

1. When $\lim_{n \to \infty} (n\beta_n) > \max \{a, (a - \beta)/2\}$, (6) and (7) hold simultaneously.
2. The bias and the variance of f_n^{Gum} defined by the stochastic approximation algorithm (1) depends on the choice of the stepsize (β_n).

Let us now state the following theorem, which gives the weak convergence rate of the estimators f_n^{Gum} defined in (1).

Theorem 1 (Weak pointwise convergence rate). Let Assumptions $(A1) - (A2)$ hold.

1. If there exists $c \geq 0$ such that $\beta_n^{-1} h_n^3 \to c$, then

$$\sqrt{\beta_n^{-1} h_n} \left(f_n^{\text{Gum}}(x) - f(x) \right) \overset{\mathcal{D}}{\to} \ \mathcal{N}\left(\frac{c^{1/2} \gamma_0}{(1 - 2a\xi)} f^{(1)}(x), \frac{1}{4(2 - (\beta - a)\xi)} f(x) \right).$$

2. If $\beta_n^{-1} h_n^3 \to \infty$, then

$$\frac{1}{h_n} \left(f_n^{\text{Gum}}(x) - f(x) \right) \overset{\mathbb{P}}{\to} \frac{c^{1/2} \gamma_0}{2(1 - 2a\xi)} f^{(1)}(x),$$

where $\overset{\mathcal{D}}{\to}$ denotes the convergence in distribution, $\mathcal{N}$ the Gaussian-distribution and $\overset{\mathbb{P}}{\to}$ the convergence in probability.

Now, in order to measure the quality of our recursive estimators f_n^{Gum} defined in (1), we consider the following amount,

$$\begin{aligned}
MWISE\left[f_n^{\text{Gum}}\right] \ &= \ \mathbb{E} \int_{\mathbb{R}} \left[f_n^{\text{Gum}}(x) - f(x) \right]^2 f(x)\, dx \\
&= \ \int_{\mathbb{R}} \left(\mathbb{E}\left(f_n^{\text{Gum}}(x) \right) - f(x) \right)^2 f(x)\, dx + \int_{\mathbb{R}} Var\left(f_n^{\text{Gum}}(x) \right) f(x)\, dx.
\end{aligned}$$

The next proposition gives the *MWISE* of our recursive Gumbel kernel density estimators f_n^{Gum} defined in (1).

Proposition 2 (*MWISE* of f_n^{Gum}). Let Assumptions $(A1) - (A2)$ hold, then

$$MWISE\left[f_n^{\text{Gum}}\right] \ = \ \frac{\gamma_0^2}{(1 - a\xi)^2} I_2 h_n^2 \mathbb{1}_{\{a \leq \beta/3\}} + \frac{1}{4(2 - (\beta - a)\xi)} \frac{\beta_n}{h_n} I_1 \mathbb{1}_{\{a \geq \beta/3\}} + o\left(h_n^2 \right).$$

Theorem 2 (Uniform convergence). Let Assumptions $(A1) - (A2)$ hold, f is uniformly continuous and there exists $\eta > 0$ such that $z \to \|z\|^\eta |f(x)|$ is a bounded function. We let $\mathcal{C}$ be a compact set of $\mathbb{R}$. Then, we have

$$\sup_{x \in \mathcal{C}} \left| f_n^{\text{Gum}}(x) - f(x) \right| = o(1) \quad \text{a.s as } n \to \infty.$$

3.1 Bandwidth selection

The bandwidth selection methods studied in the literature as part of the non-parametric kernel estimators can be divided into three main classes: cross-validation techniques, plug-in methods and the bootstrap idea. [7] proposed a detailed comparison of the three practical bandwidth selection. The last authors reaches the following conclusions; chosen appropriately, the plug-in and bootstrap selectors both methods exceed the cross-validation bandwidth, and neither of the two is best in all cases. More recently, [33] compared a plug-in bandwidth selector of a semi recursive kernel hazard function to a cross-validation bandwidth selector. The last author conclude that the plug-in method outperforms the cross-validation method. In this section, we developed a plug-in bandwidth selector based on the minimization of the *MWISE* of the proposed recursive Gumbel kernel density estimator by using the function $f(x)$ as a weight function.

The following corollary indicates that the bandwidth which minimizes the *MWISE* of f_n^{Gum} depends on the stepsize (β_n) and consequently the corresponding *MWISE* depends also on the stepsize (β_n).

Corollary 1. Let Assumptions $(A1) - (A2)$ hold. To minimize the *MWISE* of f_n^{Gum}, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$, $(\beta = 1)$, the bandwidth (h_n) must equal

$$\left(\left\{ \frac{(1 - a\xi)^2}{8\gamma_0^2 (2 - (1 - a)\xi)} \frac{I_1}{I_2} \right\}^{1/3} \beta_n^{1/3} \right).$$

Then, the corresponding *MWISE* is equal to

$$MWISE\left[f_n^{\text{Gum}}\right] \ = \ \frac{3}{4} \gamma_0^2 (2 - (1 - a)\xi)^{-2/3} (1 - a\xi)^{-2/3} I_1^{2/3} I_2^{1/3} \beta_n^{2/3} + o\left(\beta_n^{2/3} \right).$$

The following corollary shows that, for a special choice of the stepsize $(\beta_n) = (\beta_0 n^{-1})$, which fulfilled that $\lim_{n\to\infty} n\beta_n = \beta_0$ and that $(\beta_n) \in \mathcal{GS}(-1)$, the optimal bandwidth (h_n) depends on β_0 and then the corresponding *MWISE* depend on β_0.

Corollary 2. Let Assumptions $(A1) - (A2)$ hold. To minimize the *MWISE* of f_n^{Gum}, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$, $\lim_{n\to\infty} n\beta_n = \beta_0$, the bandwidth (h_n) must equal

$$\left(\left\{ \frac{(\beta_0 - 1/3)}{16\gamma_0^2} \frac{I_1}{I_2} \right\}^{1/3} n^{-1/3} \right)$$

and we then have

$$MWISE\left[f_n^{\text{Gum}} \right] = \frac{3}{4} 2^{-2/3} \gamma_0^2 \beta_0^2 (\beta_0 - 1/3)^{-4/3} I_1^{2/3} I_2^{1/3} n^{-2/3} + o\left(n^{-2/3} \right).$$

Remark 2.

1. Clearly the optimal bandwidth h_n belong to $\mathcal{GS}(-1/3)$ and then $a = 1/3$, which explain that the bandwidth h_n and the corresponding *MWISE* do not depend on the quantity a anymore.
2. Since the minimum of $\beta_0^2 (\beta_0 - 1/3)^{-4/3}$ is reached at $\beta_0 = 1$, the optimal bandwidth (h_n) must be equal to

$$\left(\left\{ \left(\frac{1}{24\gamma_0^2} \right) \frac{I_1}{I_2} \right\}^{1/3} n^{-1/3} \right) \tag{8}$$

   and therefore we have

$$MWISE\left[f_n^{\text{Gum}} \right] = \frac{3^{7/3}}{16} \gamma_0^2 I_1^{2/3} I_2^{1/3} n^{-2/3} + o\left(n^{-2/3} \right).$$

3. We consider in the rest of this paper only the stepsize $(\beta_n) = (n^{-1})$.

At this stage in order to estimate the optimal bandwidth (h_n) given in (8), we must estimate I_1 and I_2. We followed the approach of [1] and [29, 30, 33, 34], which is called plug-in estimation, and we consider the following kernel estimators for I_1 and I_2:

$$\widehat{I_1} = \frac{Q_n}{n} \sum_{i,k=1}^{n} Q_k^{-1} \beta_k b_k^{-1} \exp\left\{ \left(\frac{X_i - X_k}{b_k} \right) - \exp\left(\frac{X_i - X_k}{b_k} \right) \right\} \tag{9}$$

$$\widehat{I_2} = \frac{Q_n'^2}{n} \sum_{\substack{i,j,k=1 \\ j\neq k}}^{n} Q_k'^{-1} Q_j'^{-1} \beta_k' \beta_j' b_k'^{-2} b_j'^{-2}$$

$$\times \left\{ 1 - \exp\left(\frac{X_i - X_k}{b_k'} \right) - \exp\left(\frac{X_i - X_j}{b_j'} \right) + \exp\left(\frac{X_i - X_k}{b_k'} + \frac{X_i - X_j}{b_j'} \right) \right\}, \tag{10}$$

where b_n and b_n' are two pilot bandwidth, β_n and β_n' are two pilot stepsizes for the estimation respectively of I_1 and I_2. $Q_n = \prod_{i=1}^{n} (1 - \beta_i)$ and $Q_n' = \prod_{i=1}^{n} (1 - \beta_i')$.
In practice, in the first step of our method we let

$$b_n = n^{-\beta} \min\left\{ \widehat{s}, \frac{Q_3 - Q_1}{\mathcal{Q}(0.75) - \mathcal{Q}(0.25)} \right\}, \quad \beta \in \,]0, 1[\tag{11}$$

(see [27]) where $\widehat{s}$ is the sample standard deviation, Q_1, Q_3 denotes respectively the first and third quartiles of the length-biased distribution and $\mathcal{Q}$ represent the quantile function of the standard normal distribution). We followed the same steps as in [29, 34] and we showed that in order to minimize the *MISE* of $\widehat{I_1}$ the pilot bandwidth (b_n) must belong to $\mathcal{GS}(-2/3)$ and the pilot stepsize $(\beta_n) = (1.36 \, n^{-1})$ and in order to minimize

the *MISE* of $\widehat{I}_2$ the pilot bandwidth (b'_n) must belong to $\mathcal{GS}(-3/8)$ and the pilot stepsize $(\beta'_n) = (1.48\, n^{-1})$. Consequently, the proposed plug-in estimator of the bandwidth (h_n) using the recursive estimators defined in (1) with the choice $(\beta_n) = (n^{-1})$ is equal to

$$
\left(\left\{ \left(\frac{1}{24\gamma_0^2} \right) \frac{\widehat{I}_1}{\widehat{\widetilde{I}}_2} \right\}^{1/3} n^{-1/3} \right), \tag{12}
$$

and the corresponding *MWISE* is equal to

$$
\widehat{MWISE}\left[f_n^{\text{Gum}} \right] \quad = \quad \frac{3^{7/3}}{16} \gamma_0^2 I_1^{2/3} I_2^{1/3} n^{-2/3} + o\left(n^{-2/3} \right).
$$

Further, using the choice which minimize the variance (7) of the proposed recursive Gumbel density estimators, which is given by $(\beta_n) = ((1-a)\, n^{-1})$, the plug-in estimator of the bandwidth (h_n) must be equal to

$$
\left(\left\{ \left(\frac{1}{48\gamma_0^2} \right) \frac{\widehat{I}_1}{\widehat{\widetilde{I}}_2} \right\}^{1/3} n^{-1/3} \right) \tag{13}
$$

and the corresponding *MWISE* must be equal to

$$
\widehat{MWISE}\left[f_n^{\text{Gum}} \right] \quad = \quad \left(\frac{3}{4} \right)^{1/3} \gamma_0^2 \widehat{I}_1^{2/3} \widehat{I}_2^{1/3} n^{-2/3} + o\left(n^{-2/3} \right). \tag{14}
$$

One can check easily that by considering the well known non-recursive kernel density estimator introduced by [26] and [23] together with the Gumbel kernel, the bias and the variance of the non-recursive Gumbel density estimator $\widetilde{f}_n^{\text{Gum}}$ can be equal to the following expression (see the next proposition).

Proposition 3 (Bias and variance of $\widetilde{f}_n^{\text{Gum}}$). Let Assumptions $(A1)\, ii) - (A2)$ hold.

$$
\mathbb{E}\left[\widetilde{f}_n^{\text{Gum}}(x) \right] - f(x) = \gamma_0 h_n f^{(1)}(x) + o(h_n),
$$

and

$$
Var\left[\widetilde{f}_n^{\text{Gum}}(x) \right] \quad = \quad \frac{1}{4nh_n} f(x) + o\left(\frac{1}{nh_n} \right).
$$

It follows that,

$$
MWISE\left[\widetilde{f}_n^{\text{Gum}} \right] \quad = \quad \gamma_0^2 h_n^2 I_2 + \frac{1}{4nh_n} I_1 + o\left(h_n^2 + \frac{1}{nh_n} \right).
$$

Then, to minimize the *MWISE* of $\widetilde{f}_n^{\text{Gum}}$, the bandwidth (h_n) must equal to

$$
\left(\left\{ \frac{1}{8\gamma_0^2} \frac{I_1}{I_2} \right\}^{1/3} n^{-1/3} \right) \tag{15}
$$

and the corresponding *MWISE* is equal to

$$
MWISE\left[\widetilde{f}_n^{\text{Gum}} \right] \quad = \quad \frac{3}{4} \gamma_0^{2/3} I_1^{2/3} I_2^{2/3} n^{-2/3} + o\left(n^{-2/3} \right).
$$

Now, in order to estimate the optimal bandwidth (15), we must estimate I_1 and I_2. For this purpose, we consider the following kernel estimators, to estimate I_1 and I_2 respectively:

$$
\widetilde{I}_1 \quad = \quad \frac{1}{n\,(n-1)\,b_n} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \exp\left\{ \left(\frac{X_i - X_k}{b_n} \right) - \exp\left(\frac{X_i - X_k}{b_n} \right) \right\}, \tag{16}
$$

$$
\widetilde{I}_2 \quad = \quad \frac{1}{n^3 b_n'^4} \sum_{\substack{i,j,k=1 \\ j\neq k}}^{n} \left\{ 1 - \exp\left(\frac{X_i - X_k}{b_n'} \right) - \exp\left(\frac{X_i - X_j}{b_n'} \right) \right.
$$

$$+ \exp\left(\frac{X_i - X_k}{b'_n} + \frac{X_i - X_j}{b'_n}\right)\bigg\}, \tag{17}$$

where b_n and b'_n are two pilot bandwidth defined by (11).

We show that in order to minimize the *MISE* of $\widetilde{I}_1$ and $\widetilde{I}_2$ the pilot bandwidth (b_n) must belong to $\mathcal{GS}\left(-2/3\right)$, and in order to minimize the *MISE* of $\widetilde{I}_2$ the pilot bandwidth (b'_n) must belong to $\mathcal{GS}\left(-3/8\right)$.

Then the plug-in estimator of the bandwidth (h_n) using the non-recursive estimator (3), is given by

$$\left(\left\{\frac{1}{8\gamma_0^2}\frac{\widetilde{I}_1}{\widetilde{I}_2}\right\}^{1/3} n^{-1/3}\right) \tag{18}$$

and the plug-in of the *MWISE* of the non-recursive Gumbel density estimator (3), is given by

$$\widetilde{MWISE}\left[\widetilde{f}_n^{\text{Gum}}\right] = \frac{3}{4}\gamma_0^{2/3}\widetilde{I}_1^{2/3}\widetilde{I}_2^{2/3}n^{-2/3} + o\left(n^{-2/3}\right). \tag{19}$$

An additional motivation of using Gumbel kernel density estimators rather than Gaussian kernel density estimators is given in the following paragraph.

3.1.0.1 Large and moderate deviation

The theory of large and moderate deviations focuses on rare events and the asymptotic calculation of their probability on an exponential scale.

We recall that the rate function obtained for the Moderate Deviation Principles (MDP) of the non-recursive kernel estimator ([26] and [23]) is equal to $J_x\left(t\right) = \frac{t^2}{2f(x)\int_{\mathbb{R}} K^2(z)dz}$, when $f\left(x\right) \neq 0$ (see [19]). By considering a normal kernel $K\left(z\right) = \exp\left(-z^2/2\right)/\sqrt{2\pi}$ we infer that $J_x\left(t\right) = \frac{t^2\sqrt{\pi}}{f(x)}$.

We underline that the rate function obtained for the MDP of the recursive estimators proposed in [21] and [29], by using the stepsize which minimize the asymptotic variance and by considering a normal kernel, is equal to $J_x\left(t\right)$ (see [28]).

Now by using the Gumbel kernel, the rate function for the MDP of the Gumbel kernel density estimators are equal to $\frac{2t^2}{f(x)}$. Then, the rate function obtained for the Gaussian kernel density estimators are smaller than the one obtained for the Gumbel kernel density estimators. We conclude then that the Gumbel kernel density estimators are more concentrated around the density function $f\left(x\right)$ than the Gaussian kernel density estimators.

4 Applications

The aim of our applications is to compare the performance of the proposed (R-Gumbel) recursive Gumbel kernel density estimator defined in (1) using the stepsize $(\beta_n) = \left(n^{-1}\right)$ together with the proposed plug-in bandwidth selection procedure given in (12) to three others methods: The first one (NR-Gaussien) consists of using the non-recursive kernel density estimator proposed by [26] and [23] by using the Gaussian kernel together with the plug-in bandwidth selection method proposed in [29]; the second one (R-Gaussien) consists in using the recursive kernel density estimator proposed by [21] by using the Gaussian kernel together with the plug-in bandwidth selection method proposed in [29]; the third one (NR-Gumbel) consists in using the non-recursive Gumbel kernel density estimator defined in (3) and the considered plug-in bandwidth selection procedure given in (18).

4.1 Simulations

In order to investigate the comparison between the four approaches, we consider three sample sizes: $n = 50$, 200, and 400. In each case, the number of simulations is fixed to $N = 500$. Moreover, we

consider eight models : Model 1- the Gumbel distribution $\mathcal{G}um\,(0, 0.5)$, Model 2- the Gumbel distribution $\mathcal{G}um\,(-3, 1)$, Model 3- the Gumbel distribution $\mathcal{G}um\,(7, 3)$, Model 4- the mixture of two Gumbel distributions $0.5\mathcal{G}um\,(1, 0.5) + 0.5\mathcal{G}um\,(4, 2)$, Model 5- the mixture of two Gumbel distributions $0.25\mathcal{G}um\,(5, 1) + 0.75\mathcal{G}um\,(10, 2)$, Model 6- the Fréchet distribution $\mathcal{F}re\,(0, 1, 1)$, Model 7- the Fréchet distribution $\mathcal{F}re\,(0, 1, 3)$, and Model 8- the Fréchet distribution $\mathcal{F}re\,(0, 2, 3)$. We denote by $f_i^{\star}$ the reference density, and by f_i the density test, and then we compute the following measures: Mean Absolute Error ($MAE = n^{-1} \sum_i |f_i - f_i^{\star}|$), Mean Squared Error ($MSE = n^{-1} \sum_i (f_i - f_i^{\star})^2$) and Mean Relative Absolute Error ($MRAE = n^{-1} \sum_i \left| \frac{f_i^{\star}}{f_i} - 1 \right|$).

We recall that the extreme value type I distribution is referred to as the Gumbel distribution. The general formula for the probability density function of the Gumbel distribution given parameters (μ and σ) is equal to:

$$g_G\,(x; \mu, \sigma) = \frac{1}{\sigma} \exp\,(-z - \exp\,(-z)), \quad \text{where} \quad z = \frac{x - \mu}{\sigma}; \; -\infty < x < \infty,$$

where μ is the mode value (location parameter), σ is the scale parameter (see [11, 12]). As μ increases, the distribution shifts to the left, whereas when μ decreases, it shifts to the right. Probability distributions that have thinner tails than an exponential distribution are called light tail distributions. They go to zero much faster than the exponential, and therefore have less mass in the tail. The Gumbel distribution and the normal distribution are two examples of light tail distributions. Formally, a distribution F is called light-tailed if and only if

$$\int_{\mathbb{R}} \exp\,(\lambda x)\,F\,(dx) < \infty, \quad \text{for some } \lambda > 0.$$

However, a heavy tailed distribution has a tail that is heavier than an exponential distribution (see [5]). Formally, a distribution F is called heavy-tailed if and only if

$$\int_{\mathbb{R}} \exp\,(\lambda x)\,F\,(dx) = \infty, \quad \text{for all } \lambda > 0.$$

Many distributions are heavy tailed, including; Fréchet distribution, Lognormal distribution and Pareto distribution.

We recall that the probability density of the Fréchet distribution given parameters (μ, σ and τ) is equal to:

$$g_F\,(x; \mu, \sigma, \tau) = \frac{\tau}{\sigma} \left(\frac{x - \mu}{\sigma} \right)^{-1-\tau} \exp\left(- \left(\frac{x - \mu}{\sigma} \right)^{-\tau} \right),$$

for $x \in (\mu, \infty)$, μ, σ and τ are the location, scale and shape parameters (see [8]).

Performing all the considered methods, we report the MAE, MSE and $MRAE$ values for each considered distribution function and for each sample size in Tables 1 and 2.

From Tables 1 and 2, we observe that

- using Model 1, the `NR-Gaussian` kernel estimator gives better results for small sample size $n = 50$ and for large sample size $n = 400$. However, the `R-Gumbel` kernel estimator gives better results for moderate sample size $n = 200$.
- using Model 2 and Model 3, the `NR-Gumbel` kernel estimator gives better results for small sample size. However the `R-Gumbel` kernel estimator gives better results for moderate and large sample size regarding to the MAE, MSE and $MRAE$, with an exception for the second model for large sample size regarding to the $MRAE$.
- using Model 4 and Model 5, the Gumbel kernel estimators outperformed the Gaussian kernel estimators for all the sample sizes regarding to the MAE and MSE. However, the `NR-Gaussian` kernel estimator give better results regarding to $MRAE$ for both moderate and large sample size using for the fourth model and for large sample size for the last model.

Table 1: Quantitative comparison between the four considered methods. The bold values indicate the smallest values regarding the *MAE*, *MSE* and *MRAE*.

		Model 1: $X \sim \mathcal{G}um(0, 0.5)$							
	$n = 50$			$n = 200$			$n = 400$		
	MAE	MSE	MRAE	MAE	MSE	MRAE	MAE	MSE	MRAE
NR-Gaussian	**0.0220**	**0.0009**	**0.0931**	0.0191	0.0008	0.1100	**0.0117**	**3e^{-04}**	**0.1499**
R-Gaussian	0.0450	0.0038	0.1693	0.0286	0.0017	0.1577	0.0207	1e^{-03}	0.2224
NR-Gumbel	0.0331	0.0021	0.1554	0.0223	0.0010	0.2056	0.0440	4e^{-03}	0.4879
R-Gumbel	0.0347	0.0020	0.1257	**0.0138**	**0.0004**	**0.0923**	0.0130	4e^{-04}	0.1542

		Model 2: $X \sim \mathcal{G}um(-3, 1)$							
	$n = 50$			$n = 200$			$n = 400$		
	MAE	MSE	MRAE	MAE	MSE	MRAE	MAE	MSE	MRAE
NR-Gaussian	0.0163	4e^{-04}	0.1403	0.0093	2e^{-04}	0.1204	0.0060	7e^{-05}	**0.0981**
R-Gaussian	0.0241	1e^{-03}	0.2080	0.0142	4e^{-04}	0.1811	0.0104	2e^{-04}	0.1705
NR-Gumbel	**0.0063**	**1e^{-04}**	**0.0573**	0.0119	3e^{-04}	0.2189	0.0222	1e^{-03}	0.4761
R-Gumbel	0.0151	3e^{-04}	0.1316	**0.0076**	**1e^{-04}**	**0.1021**	**0.0057**	**6e^{-05}**	0.1455

		Model 3: $X \sim \mathcal{G}um(7, 3)$							
	$n = 50$			$n = 200$			$n = 400$		
	MAE	MSE	MRAE	MAE	MSE	MRAE	MAE	MSE	MRAE
NR-Gaussian	0.0054	5e^{-05}	0.1405	0.0031	2e^{-05}	0.1369	0.0029	2e^{-05}	0.1210
R-Gaussian	0.0078	1e^{-04}	0.2021	0.0051	5e^{-05}	0.2042	0.0043	4e^{-05}	0.1670
NR-Gumbel	**0.0018**	**6e^{-06}**	**0.0496**	0.0044	4e^{-05}	0.2549	0.0072	1e^{-04}	0.4469
R-Gumbel	0.0045	3e^{-05}	0.1143	**0.0024**	**9e^{-06}**	**0.1294**	**0.0024**	**1e^{-05}**	**0.1109**

		Model 4: $X \sim 0.5\mathcal{G}um(1, 0.5) + 0.5\mathcal{G}um(4, 2)$							
	$n = 50$			$n = 200$			$n = 400$		
	MAE	MSE	MRAE	MAE	MSE	MRAE	MAE	MSE	MRAE
NR-Gaussian	0.0151	0.0013	0.0876	0.0084	5e^{-04}	**0.0948**	0.0057	2e^{-04}	**0.1557**
R-Gaussian	0.0195	0.0021	0.1290	0.0122	1e^{-03}	0.1159	0.0089	6e^{-04}	0.1559
NR-Gumbel	**0.0043**	**0.0001**	**0.0501**	**0.0065**	**1e^{-04}**	0.2478	0.0111	3e^{-04}	0.5663
R-Gumbel	0.0147	0.0009	0.1321	0.0071	3e^{-04}	0.1023	**0.0056**	**1e^{-04}**	0.1841

		Model 5: $X \sim 0.25\mathcal{G}um(5, 1) + 0.75\mathcal{G}um(10, 2)$							
	$n = 50$			$n = 200$			$n = 400$		
	MAE	MSE	MRAE	MAE	MSE	MRAE	MAE	MSE	MRAE
NR-Gaussian	0.0120	4e^{-04}	0.1694	0.0053	1e^{-04}	0.0952	0.0041	4e^{-05}	**0.1137**
R-Gaussian	0.0156	6e^{-04}	0.2429	0.0089	3e^{-04}	0.1613	0.0067	1e^{-04}	0.1486
NR-Gumbel	**0.0024**	**1e^{-05}**	**0.0492**	0.0057	6e^{-05}	0.1968	0.0081	1e^{-04}	0.3857
R-Gumbel	0.0099	2e^{-04}	0.1873	**0.0044**	**5e^{-05}**	**0.0992**	**0.0033**	**1e^{-05}**	0.1361

- using Model 6, the `NR-Gaussian` kernel estimator gives better results for small sample size, while the `R-Gumbel` kernel estimator gives better results for moderate and large sample size.
- using Models 7 and 8, the `NR-Gumbel` kernel estimator gives better results for small sample size, while the `R-Gumbel` kernel estimator gives better results for moderate and large sample size.

Table 2: Quantitative comparison between the four considered methods. The bold values indicate the smallest values regarding the *MAE*, *MSE* and *MRAE*.

Model 6: $X \sim \mathcal{F}re\,(0, 1, 1)$									
	$n = 50$			$n = 200$			$n = 400$		
	MAE	MSE	MRAE	MAE	MSE	MRAE	MAE	MSE	MRAE
NR-Gaussian	**$6.8e^{-04}$**	**$2.0e^{-05}$**	**0.3430**	$5.4e^{-0.4}$	$5.6e^{-05}$	0.2126	$3.6e^{-04}$	$4.4e^{-05}$	0.1024
R-Gaussian	$9.6e^{-04}$	$4.3e^{-05}$	0.3445	$5.2e^{-04}$	$5.2e^{-05}$	0.2185	$3.6e^{-04}$	$5.2e^{-05}$	0.1018
NR-Gumbel	$7.6e^{-04}$	$3.0e^{-05}$	0.6271	$2.1e^{-04}$	$6.0e^{-06}$	0.1109	$1.4e^{-04}$	$4.0e^{-06}$	0.1014
R-Gumbel	$1.2e^{-03}$	$6.8e^{-05}$	0.4826	**$5.2e^{-05}$**	**$1.2e^{-06}$**	**0.1095**	**$1.3e^{-04}$**	**$2.0e^{-06}$**	**0.1012**

Model 7: $X \sim \mathcal{F}re\,(0, 1, 3)$									
	$n = 50$			$n = 200$			$n = 400$		
	MAE	MSE	MRAE	MAE	MSE	MRAE	MAE	MSE	MRAE
NR-Gaussian	0.0221	0.0031	0.0860	0.0125	0.0019	0.2731	0.0072	0.0005	0.0882
R-Gaussian	0.0318	0.0059	0.0837	0.0174	0.0024	0.2625	0.0086	0.0007	0.0848
NR-Gumbel	**0.0102**	**0.0003**	**0.0827**	0.0121	0.0011	0.1060	0.0089	0.0022	0.0821
R-Gumbel	0.0232	0.0023	0.0979	**0.0105**	**0.0005**	**0.0528**	**0.0063**	**0.0004**	**0.0431**

Model 8: $X \sim \mathcal{F}re\,(0, 2, 3)$									
	$n = 50$			$n = 200$			$n = 400$		
	MAE	MSE	MRAE	MAE	MSE	MRAE	MAE	MSE	MRAE
NR-Gaussian	0.0111	0.0008	0.1153	0.0048	$1.4e^{-04}$	0.0853	**0.0035**	$1.2e^{-04}$	0.0854
R-Gaussian	0.0156	0.0015	0.1125	0.0057	$3.2e^{-04}$	0.0695	0.0054	$2.2e^{-03}$	0.0724
NR-Gumbel	**0.0062**	**0.0001**	**0.1106**	0.0130	$7.4e^{-04}$	0.0892	0.0165	$1.4e^{-03}$	0.0744
R-Gumbel	0.0111	0.0005	0.1260	**0.0046**	**$1.2e^{-04}$**	**0.0545**	0.0045	**$1.1e^{-04}$**	**0.0633**

- using all the considered models, the R-Gumbel kernel estimator outperformed the NR-Gumbel kernel estimator for moderate and large sample size. However, the NR-Gumbel kernel estimator gives better results for small sample size.
- using all the considered models and by performing the four approaches, the *MAE*, *MSE* and *MRAE* decrease as the sample size increase.
- by considering models with light tail distribution (see Model 1, Model 2, Model 3), we observe that the more the location parameter moves away from zero and the more the scale parameter moves away from one, the more the *MAE*, *MSE* and *MRAE* decrease.
- by considering models with heavy tailed distribution (see Model 6, Model 7, Model 8), we observe that the more the scale and shape parameter move away from one, the more the *MAE*, *MSE* and *MRAE* decrease.

Remark 3. Form Tables 1 and 2, we can observe that the Gumbel kernel estimators outperfomed the Gaussian kernel estimators for all the considered models including those with light tail distribution (see Model 1, Model 2, Model 3, Model 4 and Model 5) and those with heavy tailed distribution (see Model 6, Model 7 and Model 8), an exception is made when the location is equal to zero (see Model 1 and Model 6).

Figure 1 shows that the Gumbel kernel estimators are closer to the true density compared to the Gaussian kernel estimators. Moreover, Figure 2 shows clearly that the recursive Gumbel kernel estimator is closer to the true density compared to the three other methods.
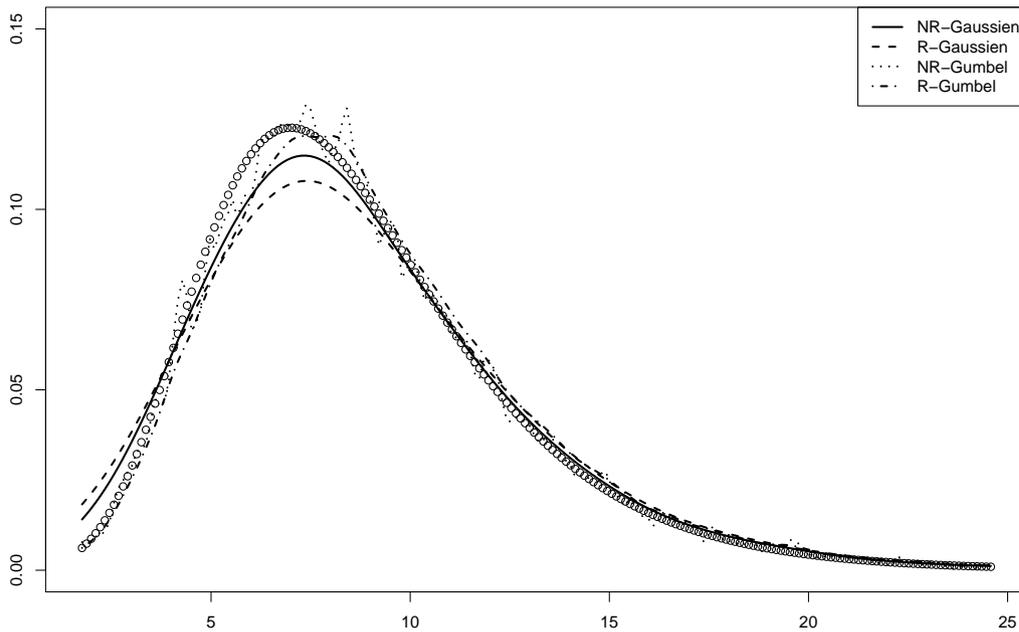
Figure 1: Qualitative comparison between the four considered estimators. The circles represent the true distribution. Here we consider the Gumbel distribution $\mathcal{G}um\,(7,3)$ and the sample size $n = 200$.

Table 3: The comparison between the plug-In *MWISE* of the four considered estimators using the proposed real data. The bold values indicate the smallest values regarding the plug-In *MWISE*

|	I_1	I_2	h_n	*MWISE*
NR-Gaussian	0.1680	0.0145	0.5478	0.001638
R-Gaussian.	0.1633	0.0172	0.4140	0.001756
NR-Gumbel	0.1768	0.0429	0.2861	0.001215
R-Gumbel	0.1637	0.0246	0.2327	**0.000976**

4.2 Caniapiscau River Daily Flows

We considered `caniapiscau` dataset which appears in the `R` package `FlowScreen`.

This data contain the mean daily streamflow for the Caniapiscau River for the period 1954-1999. The Caniapiscau River is located in Nunavik, Quebec, Canada, and flows northward. The headwaters were dammed to create the Caniapiscau Reservoir, which started filling in 1981. In 1985, the reservoir was diverted to the west into the La Grande hydroelectric complex. This flow time series is used as an example of a river with a known change point to demonstrate the package's screening capabilities.

We are interested by the estimation of the probability density of the Mean daily streamflow, measured in m3/s.

Remark 4.

1. The results given in Table 3 are based on the Plug-In estimator of I_1 and I_2 (see, (9), (10), (16) and (17) for the estimation of I_1 and I_2 and (14) and (19) for the plug-In *MWISE*). This estimators are very useful when we use real data since we do not need to know the density of our sample.
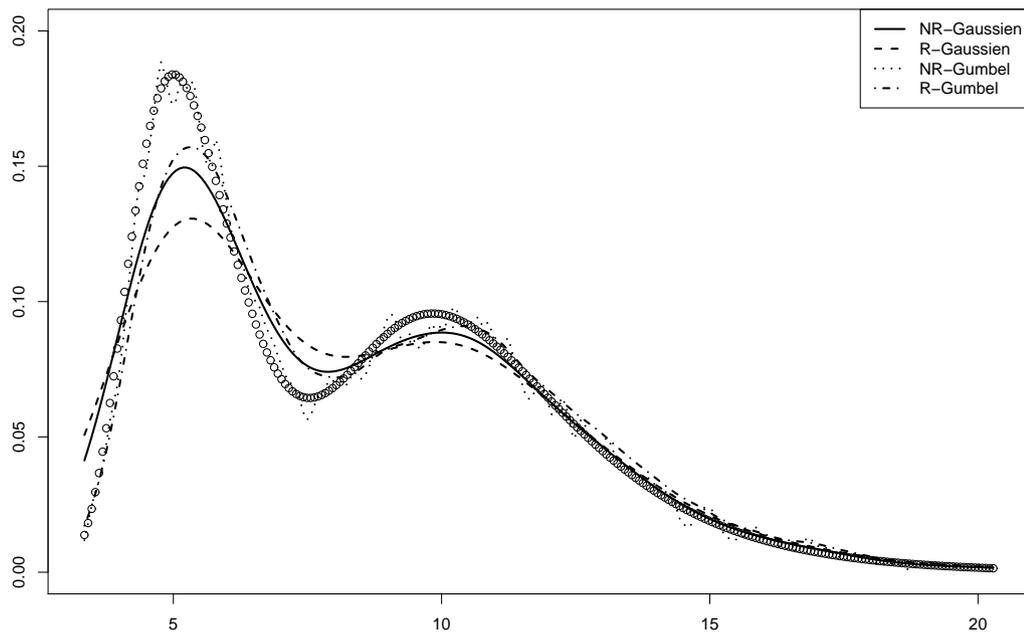
Figure 2: Qualitative comparison between the four considered estimators. The circles represent the true distribution. Here we consider the mixture of two Gumbel distributions $0.25\mathcal{G}um\,(5, 1) + 0.75\mathcal{G}um\,(10, 2)$ and the sample size $n = 200$.

2. Table 3 provides the estimation of the two unknown quantities I_1 and I_2 before given the optimal bandwidth of the four considered methods respectively, we then use the *MWISE* to compare the four considered approaches, we infer that the Gumbel kernel estimators outperformed the Gaussian kernel estimators with a particular preference for using the recursive version.
3. Qualitative comparison between the Gumbel and Gaussian kernel estimators are given in Figures 3 and 4.
4. Figure 4 illustrates that the Gumbel kernel estimator provides globally the best fit, whilst the Gaussian kernel estimator tends to underestimate the density function.
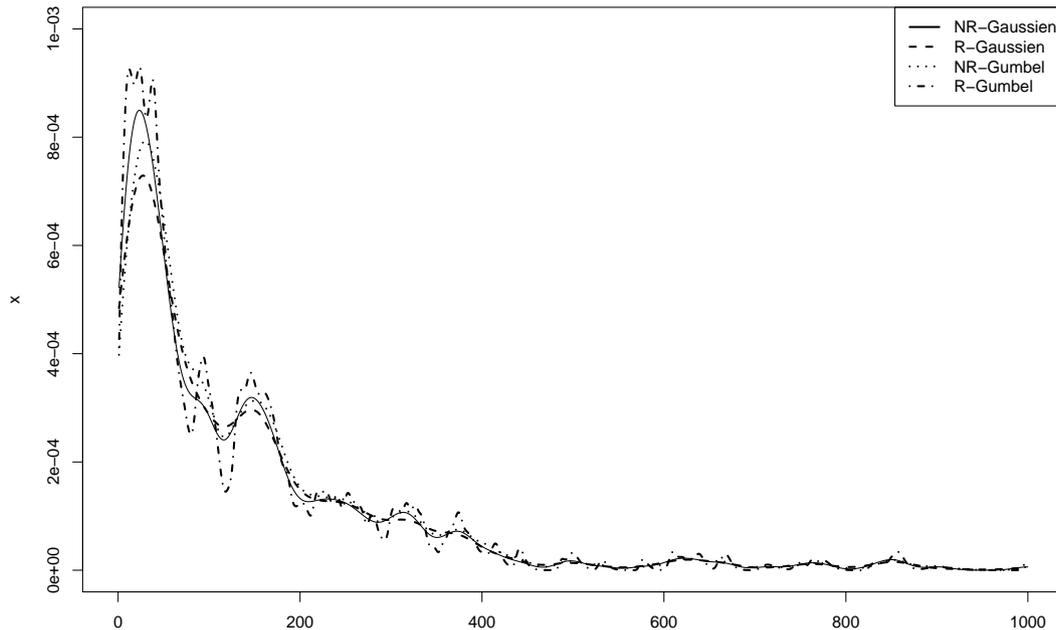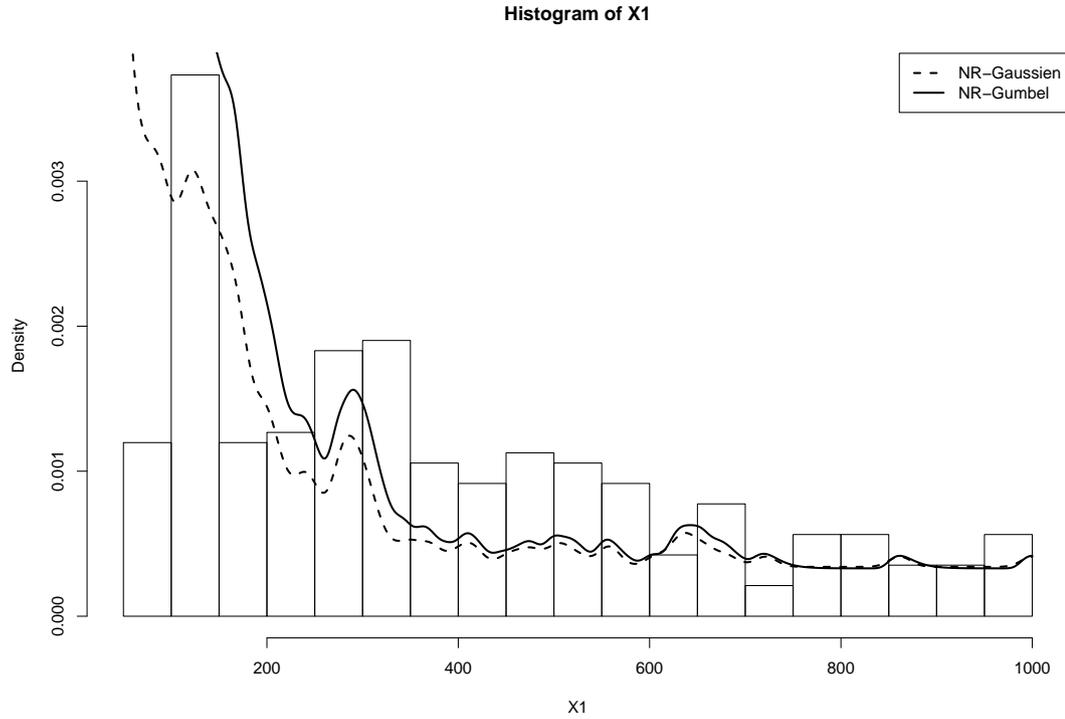
Figure 3: Qualitative comparison between the four considered estimators. Here we consider the `caniapiscau` dataset of the package `FlowScreen`.

5 Conclusion

This paper propose an automatic selection of the bandwidth of two proposed Gumbel kernel density estimators, the first one is recursive, however the second one is non-recursive. The two proposed estimators asymptotically follow normal distributions. The estimators are compared to the two classical Gaussian density estimators, the first one is non-recursive and introduced by [26] and [23], the second one is recursive and proposed by [21] and investigated by [29]. We showed that, using some selected bandwidth and some particularly stepsizes, the two proposed Gumbel estimators are very competitive to the two classical Gaussian density estimators. The simulation study confirms the benefits of our two proposed Gumbel estimators.

In conclusion, the proposed estimators are very competitive to the two classical Gaussian density estimators. We plan to extend this work to the multi-dimensional context with a comparison to the work of [2]. We plan also to consider the bias correction techniques with application to heavy tailed data.

We plan also to consider Bernstein polynomials rather than kernels and to propose an adaptation of the estimators developed in [15] and [36] with application to heavy tailed data. Moreover, we plan to make an extensions of our proposed plug-in method in future with application on extreme value and to consider the case of the averaged Révész's regression estimators (see [22] and [31, 32]) and the semi-recursive kernel regression estimators proposed by [34] and the case of time series as in [13] in recursive way (see [14]).

Histogram of X1

Figure 4: Qualitative comparison between Gumbel and Gaussian kernel estimators. Here we consider the `caniapiscau` dataset of the package `FlowScreen`.

A Proofs

Throughout this section we use the following notation:

$$Q_n = \prod_{j=1}^{n} \left(1 - \gamma_j\right),$$

$$K_{\text{Gum}}(z) = \exp\left(-x - \exp(x)\right).$$

$$Z_n^{\text{Gum}}(x) = h_n^{-1} K_{\text{Gum}}\left(\frac{X_n - x}{h_n}\right). \tag{20}$$

$$Y_n^{\text{Gum}}(x) = Q_n^{-1} \beta_n \left(Z_n^{\text{Gum}}(x) - \mathbb{E}\left(Z_n^{\text{Gum}}(x)\right)\right). \tag{21}$$

Let us first state the following technical lemma.

Lemma 1. Let $(v_n) \in \mathcal{GS}\left(v^\star\right)$, $(\beta_n) \in \mathcal{GS}\left(-\beta\right)$, and $m > 0$ such that $m - v^\star \xi > 0$ where ξ is defined in (5). We have

$$\lim_{n \to +\infty} v_n Q_n^m \sum_{k=1}^{n} Q_k^{-m} \frac{\beta_k}{v_k} = \frac{1}{m - v^\star \xi}.$$

Moreover, for all positive sequence (β_n) such that $\lim_{n \to +\infty} \beta_n = 0$, and all $\delta \in \mathbb{R}$,

$$\lim_{n \to +\infty} v_n Q_n^m \left[\sum_{k=1}^{n} Q_k^{-m} \frac{\beta_k}{v_k} \beta_k + \delta\right] = 0.$$

Remark 5.

1. The application of Lemma 1 requires Assumption $(A1)(iii)$ on the limit of $(n\beta_n)$ as n goes to infinity.
2. The application of Lemma 1 ensures that the bias and the variance depend only on h_n and not on $h_1, \ldots, h_n$.

A.1 Proof of Proposition 1

First, it follows from (1) and (20), that

$$
\begin{aligned}
f_n^{\text{Gum}}(x) - f(x) &= (1 - \beta_n)\left(f_{n-1}^{\text{Gum}}(x) - f(x)\right) + \beta_n\left(Z_n^{\text{Gum}}(x) - f(x)\right) \\
&= \sum_{k=1}^{n-1}\left[\prod_{j=k+1}^{n}(1 - \gamma_j)\right]\beta_k\left(Z_k^{\text{Gum}}(x) - f(x)\right) + \beta_n\left(Z_n^{\text{Gum}}(x) - f(x)\right) \\
&\quad + \left[\prod_{j=1}^{n}(1 - \gamma_j)\right]\left(f_0^{\text{Gum}}(x) - f(x)\right) \\
&= Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k\left(Z_k^{\text{Gum}}(x) - f(x)\right) + Q_n\left(f_0^{\text{Gum}}(x) - f(x)\right). \quad (22)
\end{aligned}
$$

Moreover, we have

$$
\mathbb{E}\left(f_n^{\text{Gum}}(x)\right) - f(x) = Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k\left(\mathbb{E}\left(Z_k^{\text{Gum}}(x)\right) - f(x)\right) + Q_n\left(f_0^{\text{Gum}}(x) - f(x)\right).
$$

Further, since we have $\int_{\mathbb{R}} K_{\text{Gum}}(z)\,dz = 1$, Taylor's expansion with integral remainder ensures that

$$
\begin{aligned}
\mathbb{E}\left[Z_k^{\text{Gum}}(x)\right] - f(x) &= \int_{\mathbb{R}} h_k^{-1}K_{\text{Gum}}\left(h_k^{-1}(y - x)\right)f(y)\,dy - f(x) \\
&= \int_{\mathbb{R}} K_{\text{Gum}}(z)\left[f(x + zh_k) - f(x)\right]dz \\
&= h_k f^{(1)}(x)\int_{\mathbb{R}} zK_{\text{Gum}}(z)\,dz + h_k\delta_k(x)
\end{aligned}
$$

with

$$
\delta_k(x) = h_k^{-1}\int_{\mathbb{R}} K_{\text{Gum}}(z)\left[f(x + zh_k) - f(x) - zh_k f^{(1)}(x)\right]dz,
$$

and, since $f^{(1)}$ is bounded and continuous at x, we have $\lim_{k\to\infty}\delta_k(x) = 0$. Moreover, since

$$
\int_{\mathbb{R}} zK_{\text{Gum}}(z)\,dz = -\int_{0}^{+\infty} \exp(-z)\ln(z)\,dz = \int_{0}^{1} H_z\,dz = \gamma_0,
$$

where H_z is the fractional harmonic number and γ_0 is the Euler-Mascheroni constant ($\gamma_0 \simeq 0.57721$). When $a \leq \beta/3$, we have $\lim_{n\to\infty}(n\beta_n) > 2a$, we apply Lemma 1 to infer that

$$
\begin{aligned}
\mathbb{E}\left[f_n^{\text{Gum}}(x)\right] - f(x) &= \int_{\mathbb{R}} zK_{\text{Gum}}(z)\,dz f^{(1)}(x)Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k h_k[1 + o(1)] + Q_n\left(f_0(x) - f(x)\right) \\
&= \frac{\gamma_0}{(1 - a\xi)}f^{(1)}(x)h_n\left[1 + o(1)\right].
\end{aligned}
$$

Moreover, in the case $a > \beta/3$, we have $h_n = o\left(\sqrt{\beta_n h_n^{-1}}\right)$. Whence, since $\lim_{n\to\infty}(n\beta_n) > (\beta - a)/2$, the application of Lemma 1 gives

$$
\mathbb{E}\left[f_n^{\text{Gum}}(x)\right] - f(x) = Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k o\left(\sqrt{\beta_k h_k^{-1}}\right) + O(Q_n)
$$

$$= o\left(\sqrt{\beta_n h_n^{-1}}\right),$$

which gives (6). Now, we have

$$
\begin{aligned}
Var\left[f_n^{\mathrm{Gum}}(x)\right] &= Q_n^2 \sum_{k=1}^n Q_k^{-2}\beta_k^2 Var\left[Z_k(x)\right] \\
&= Q_n^2 \sum_{k=1}^n \frac{Q_k^{-2}\beta_k^2}{h_k}\left[\int_{\mathbb{R}} K_{\mathrm{Gum}}^2(z) f(x+zh_k)\,dz - h_k\left(\int_{\mathbb{R}} K_{\mathrm{Gum}}(z) f(x+zh_k)\,dz\right)^2\right] \\
&= Q_n^2 \sum_{k=1}^n \frac{Q_k^{-2}\beta_k^2}{h_k}\left[f(x)\int_{\mathbb{R}} K_{\mathrm{Gum}}^2(z)\,dz + v_k(x) - h_k\tilde{v}_k(x)\right]
\end{aligned}
$$

with

$$
v_k(x) = \int_{\mathbb{R}} K_{\mathrm{Gum}}^2(z)\left[f(x+zh_k) - f(x)\right]dz \quad \text{and} \quad \tilde{v}_k(x) = \left(\int_{\mathbb{R}} K_{\mathrm{Gum}}^2(z) f(x+zh_k)\,dz\right)^2.
$$

In view of $(A2)$, we have $\lim_{k\to\infty} v_k(x) = 0$ and $\lim_{k\to\infty} h_k\tilde{v}_k(x) = 0$. In the case $a \geq \beta/3$, we have $\lim_{n\to\infty}(n\beta_n) > (\beta - a)/2$, we make use of Lemma 1 to infer that

$$
\begin{aligned}
Var\left[f_n^{\mathrm{Gum}}(x)\right] &= Q_n^2 \sum_{k=1}^n \frac{Q_k^{-2}\beta_k^2}{h_k}\left[f(x)\int_{\mathbb{R}} K_{\mathrm{Gum}}^2(z)\,dz + o(1)\right] \\
&= \frac{1}{2-(\beta-a)\xi}\frac{\beta_n}{h_n}\left[f(x)\int_{\mathbb{R}} K_{\mathrm{Gum}}^2(z)\,dz + o(1)\right].
\end{aligned}
$$

Moreover, by using the following variable change $t = \exp(-z)$, we have

$$
\int_{\mathbb{R}} K_{\mathrm{Gum}}^2(z)\,dz = \int_{\mathbb{R}} \exp(-2z)\exp(-2\exp(-z))\,dz = \frac{1}{4}.
$$

Further, in the case when $a < \beta/3$, we have $\beta_n h_n^{-1} = o\left(h_n^2\right)$, and since $\lim_{n\to\infty}(n\beta_n) > a$, Lemma 1 ensures that

$$
\begin{aligned}
Var\left[f_n^{\mathrm{Gum}}(x)\right] &= Q_n^2 \sum_{k=1}^n Q_k^{-2}\beta_k o\left(h_k^2\right) \\
&= o\left(h_n^2\right),
\end{aligned}
$$

which gives (7).

A.2 Proof of Theorem 1

Let us at first assume that, if $a \geq \beta/3$, then

$$
\sqrt{\beta_n^{-1} h_n}\left(f_n^{\mathrm{Gum}}(x) - \mathbb{E}\left[f_n^{\mathrm{Gum}}(x)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4(2-(\beta-a)\xi)}f(x)\right). \tag{23}
$$

In the case when $a > \beta/3$, Part 1 of Theorem 1 follows from the combination of (6) and (23). In the case when $a = \beta/3$, Parts 1 and 2 of Theorem 1 follow from the combination of (6) and (23). In the case $a < \beta/3$, (7) implies that

$$
h_n^{-2}\left(f_n^{\mathrm{Gum}}(x) - \mathbb{E}\left(f_n^{\mathrm{Gum}}(x)\right)\right) \xrightarrow{\mathbb{P}} 0,
$$

and the application of (6) gives Part 2 of Theorem 1.

Let us now prove (23). In view of (22), we have

$$
\begin{aligned}
f_n^{\mathrm{Gum}}(x) - \mathbb{E}\left[f_n^{\mathrm{Gum}}(x)\right] &= (1-\beta_n)\left(f_{n-1}^{\mathrm{Gum}}(x) - \mathbb{E}\left[f_{n-1}^{\mathrm{Gum}}(x)\right]\right) + \beta_n\left(Z_n^{\mathrm{Gum}}(x) - \mathbb{E}\left[Z_n^{\mathrm{Gum}}(x)\right]\right) \\
&= Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k\left(Z_k^{\mathrm{Gum}}(x) - \mathbb{E}\left[Z_k^{\mathrm{Gum}}(x)\right]\right).
\end{aligned}
$$

Moreover, the application of Lemma 1 ensures that

$$
\begin{aligned}
v_n^2 &= \sum_{k=1}^{n} Var\left(Y_k^{\mathrm{Gum}}(x)\right) = \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 Var\left(Z_k^{\mathrm{Gum}}(x)\right) \\
&= \sum_{k=1}^{n} \frac{Q_k^{-2}\beta_k^2}{h_k}\left[f(x)\int_{\mathbb{R}} K_{\mathrm{Gum}}^2(z)\,dz + o(1)\right] = O\left(\frac{1}{Q_n^2}\frac{\beta_n}{h_n}\right).
\end{aligned}
$$

Further, we have, for all $p > 0$,

$$
\mathbb{E}\left[\left|Z_k^{\mathrm{Gum}}(x)\right|^{2+p}\right] = O\left(\frac{1}{h_k^{1+p}}\right). \tag{24}
$$

We let $\mathcal{V}_n = \frac{h_n^{1+p}}{\beta_n^{1+p}}$, then $\mathcal{V}_n \in \mathcal{GS}\left((\beta - a)(1 + p)\right)$. Moreover, since $\lim_{n\to\infty}(n\beta_n) > (\beta - a)/2$, then, there exists $p > 0$ such that $\lim_{n\to\infty}(n\beta_n) > \frac{1+p}{2+p}(\beta - a) > (\beta - a)/2$. Then, the application of Lemma 1 ensures that

$$
\mathcal{V}_n Q_n^{2+p} \sum_{k=1}^{n} Q_k^{-2-p}\frac{\beta_k}{\mathcal{V}_k} = O(1). \tag{25}
$$

Then, the combination of (21), (24) and (25), ensure that

$$
\begin{aligned}
\sum_{k=1}^{n} \mathbb{E}\left[\left|Y_k^{\mathrm{Gum}}(x)\right|^{2+p}\right] &= O\left(\sum_{k=1}^{n} Q_k^{-2-p}\beta_k^{2+p}\mathbb{E}\left[\left|Z_k^{\mathrm{Gum}}(x)\right|^{2+p}\right]\right) \\
&= O\left(\sum_{k=1}^{n} \frac{Q_k^{-2-p}\beta_k^{2+p}}{h_k^{1+p}}\right) = O\left(\frac{\beta_n^{1+p}}{Q_n^{2+p}h_n^{1+p}}\right),
\end{aligned}
$$

we thus get

$$
\frac{1}{v_n^{2+p}} \sum_{k=1}^{n} \mathbb{E}\left[\left|Y_k^{\mathrm{Gum}}(x)\right|^{2+p}\right] = O\left(\left[\beta_n h_n^{-1}\right]^{p/2}\right) = o(1).
$$

The convergence in (23) then follows from the application of Lyapounov's Theorem.

A.3 Proof of Theorem 2

First, using the compactness property of the set $\mathcal{C}$, we infer that, for some $(x_k)_{1\le k\le\gamma_n}$, $\mathcal{C} \subset \bigcup_{k=1}^{\gamma_n} B(x_k, a_n)$, where $\gamma_n \sim a_n^{-1}$ with $a_n = h_n^{\frac{2}{a}+1}$.

Now, for any $x \in \mathcal{C}$, we set $\widetilde{k}(x) = \arg\min_k \|x_k - x\|$. Then, for any $x \in \mathcal{C}$, we have

$$
\begin{aligned}
\sup_{x\in\mathcal{C}}\left|f_n^{\mathrm{Gum}}(x) - \mathbb{E}\left[f_n^{\mathrm{Gum}}(x)\right]\right| \le\;&\sup_{x\in\mathcal{C}}\left|f_n^{\mathrm{Gum}}(x) - f_n^{\mathrm{Gum}}(x_{\widetilde{k}})\right| \\
&+ \sup_{x\in\mathcal{C}}\left|f_n^{\mathrm{Gum}}(x_{\widetilde{k}}) - \mathbb{E}\left[f_n^{\mathrm{Gum}}(x_{\widetilde{k}})\right]\right| \\
&+ \sup_{x\in\mathcal{C}}\left|\mathbb{E}\left[f_n^{\mathrm{Gum}}(x_{\widetilde{k}})\right] - \mathbb{E}\left[f_n^{\mathrm{Gum}}(x)\right]\right|
\end{aligned}
$$

$$=: \mathcal{T}_{1,n} + \mathcal{T}_{2,n} + \mathcal{T}_{3,n}. \tag{26}$$

First, let α denote the Hölder order of K^{Gum} and $\left\|K^{\mathrm{Gum}}\right\|_H$ its corresponding Hölder norm. Then, it follows from (2) that for any $x \in \mathcal{C}$

$$
\begin{aligned}
\left|f_n^{\mathrm{Gum}}(x) - f_n^{\mathrm{Gum}}\left(x_{\widetilde{k}}\right)\right| &\leq Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k h_k^{-1}\left|K^{\mathrm{Gum}}\left(\frac{X_k - x}{h_k}\right) - K^{\mathrm{Gum}}\left(\frac{X_k - x_{\widetilde{k}}}{h_k}\right)\right| \\
&\leq 2\left\|K^{\mathrm{Gum}}\right\|_H Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k h_k^{-1}\left(\frac{\left\|x - x_{\widetilde{k}}\right\|}{h_k}\right)^{\alpha} \\
&\leq 2\left\|K^{\mathrm{Gum}}\right\|_H Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k h_k,
\end{aligned}
$$

we then get $\mathcal{T}_{1,n} = o(1)$ and $\mathcal{T}_{3,n} = o(1)$. Now, we set $\rho > 0$ and M such that $\|f_\infty\| \int_{\|z\|>M}\left|K^{\mathrm{Gum}}(z)\right| dz \leq \rho/2$. Moreover, Lemma 1 ensures that $Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k = 1 + o(1)$, then, it comes that

$$
\begin{aligned}
\left|f_n^{\mathrm{Gum}}\left(x_{\widetilde{k}}\right) - \mathbb{E}\left[f_n^{\mathrm{Gum}}\left(x_{\widetilde{k}}\right)\right]\right| &\leq Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k h_k^{-1}\left|\mathbb{E}\left[K^{\mathrm{Gum}}\left(\frac{X_k - x_{\widetilde{k}}}{h_k}\right)\right]\right| \\
&\leq \frac{\rho}{2} + \int_{\|z\|\leq M}\left|K^{\mathrm{Gum}}(z)\right|\left|f\left(x_{\widetilde{k}}\right)\right| dz \\
&\quad + Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k \int_{\|z\|>M}\left|K^{\mathrm{Gum}}(z)\right|\left|f\left(x_{\widetilde{k}} - zh_k\right) - f\left(x_{\widetilde{k}}\right)\right| dz.
\end{aligned}
$$

Then, the uniform continuity of f combined with the dominate convergence and the existence of $\eta > 0$ such that $z \to \|z\|^{\eta}|f(x)|$ is a bounded function ensure that $\mathcal{T}_{2,n} = o(1)$. Then the combination of (6) and (26) concludes the proof of Theorem 2.

References

[1] Altman, N. and C. Léger (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inference 46*(2), 195–214.

[2] Béranger, B., T. Duong, S. E. Perkins-Kirkpatrick, and S. A. Sisson (2019). Tail density estimation for exploratory data analysis using kernel methods. *J. Nonparametr. Stat. 31*(1), 144–174.

[3] Bojanic, R. and E. Seneta (1973). A unified theory of regularly varying sequences. *Math. Z. 134*(2), 91–106.

[4] Bolancé, C., M. Guillén, J. Gustafsson, and J. P. Nielson (2012). *Quantitative Operational Risk Models*. CRC Press, Boca Raton FL.

[5] Bryson, M. C. (1974). Heavy-tailed distributions: properties and tests. *Technometrics 16*(1), 61–68.

[6] Buch-Larsen, T., J. P. Nielsen, M. Guillén, and C. Bolancé (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics 39*(6), 503–518.

[7] Delaigle, A. and I. Gijbels (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal. 45*(2), 249–267.

[8] de Gusmão, F. R. S., E. M. M. Ortega, and G. M. Cordeiro (2011). The generalized inverse Weibull distribution. *Statist. Papers 52*(3), 591–619.

[9] Gorgoso-Varela, J. J and A. Rojo-Alboreca (2014). Use of Gumbel and Weibull functions to model extreme values of diameter distributions in forest stands. *Ann. For. Sci. 71*(7), 741–750.

[10] Galambos, J. and E. Seneta (1973). Regularly varying sequences. *Proc. Amer. Math. Soc. 41*(1), 110–116.

[11] Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques. *Ann. Inst. H. Poincaré 5*(2), 115–158.

[12] Gumbel, E. J. (1941). The return period of flood flows. *Ann. Math. Statist. 12*(2), 163–190.

[13] Hart, J. D. and P. Vieu (1990). Data-driven bandwidth choice for density estimation based on dependent data. *Ann. Statist.* *18*(2), 873–890.

[14] Huang, Y., X. Chen, and W. B. Wu (2014). Recursive nonparametric estimation for times series. *IEEE Trans. Inform. Theory* *60*(2), 1301–1312.

[15] Jmaei, A., Y. Slaoui, and W. Dellagi (2017). Recursive distribution estimators defined by stochastic approximation method using Bernstein polynomials. *J. Nonparametr. Stat. 29*(4), 792–805.

[16] Kotz, S. and S. Nadarajah (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London.

[17] Kumar, R., A. K. Srivastava, and V. Kumar (2012). Analysis of Gumbel model for software reliability using Bayesian paradigm. *Int. J. Adv. Res. Artif. Intell. 1*(9), 39–45.

[18] Miladinovic, B (2008). Kernel Density Estimation of Reliability with Applications to Extreme Value Distribution. PhD thesis, University of South Florida. Available at https://scholarcommons.usf.edu/etd/408/.

[19] Mokkadem, A., M. Pelletier, M., and J. Worms (2005). Large and moderate deviations principles for kernel estimation of a multivariate density and its partial derivatives. *Aust. N. Z. J. Stat. 47*(4), 489–502.

[20] Mokkadem, A. and M. Pelletier (2007). A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *Ann. Statist. 35*(4), 1749–1772.

[21] Mokkadem, A., M. Pelletier, and Y. Slaoui (2009a). The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference 139*(7), 2459–2478.

[22] Mokkadem, A., M. Pelletier, and Y. Slaoui (2009b). Revisiting Révész's stochastic approximation method for the estimation of a regression function. *ALEA. Lat. Am. J. Probab. Math. Stat. 6*, 63–114.

[23] Parzen, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist. 33*(3), 1065–1076.

[24] Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Stud. Sci. Math. Hung. 8*(3-4), 391–398.

[25] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Statistics 8*(1), 119–126.

[26] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist. 27*(3), 832–837.

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

[28] Slaoui, Y. (2013). Large and moderate deviation principles for recursive kernel density estimators defined by stochastic approximation method. *Serdica Math. J. 39*(1), 53–82.

[29] Slaoui, Y. (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *J. Probab. Stat.* 2014, Article ID 739640, 11 pages.

[30] Slaoui, Y. (2014b). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist. 23*(4), 306–325.

[31] Slaoui, Y. (2015a). Moderate deviation principles for recursive regression estimators defined by stochastic approximation method. *Int. J. Math. Stat. 16*(2), 51–60.

[32] Slaoui, Y. (2015b). Large and moderate deviation principles for averaged stochastic approximation method for the estimation of a regression function. *Serdica Math. J. 41*(2-3), 307–328.

[33] Slaoui, Y. (2016a). On the choice of smoothing parameters for semirecursive nonparametric hazard estimators. *J. Stat. Theory Pract. 10*(4), 656–672.

[34] Slaoui, Y. (2016b). Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat. Interface 9*(3), 375–388.

[35] Slaoui, Y. (2018). Bias reduction in kernel density estimation. *J. Nonparametr. Stat. 30*(2), 505–522.

[36] Slaoui, Y. and A. Jmaei (2019). Recursive density estimators based on Robbins-Monro's scheme and using Bernstein polynomials. *Stat. Interface 12*(3), 439–455.

[37] Tsybakov, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution. *Probl. Inf. Transm. 26*(1), 31–37.