# Asymmetric kernel density estimation of heavy tailed data with application to clustering

Yasmina ZIANE [1] & Nabil ZOUGAB [2] & Smail ADJABI [3]

[1,3] *Research Unit LaMOS, University of Bejaia, Operational Research Department, Faculty of Exact Sciences. Bejaia, Algeria.*
[2] *Research Unit LaMOS, University of Bejaia, Department of Electrical Engineering, Faculty of Technology. Bejaia, Algeria.*

**Résumé.** Dans ce travail, nous proposons d'estimer la fonction de densité des données à queue lourde avec un support non négatif. Comme les données à queue lourde se caractérisent par des observations rares dans la queue, nous proposons de subdiviser l'ensemble de données en deux sous-ensembles de densité forte et faible, en utilisant k-means, une méthode de classification non supervisée d'apprentissage machine. Pour cela, nous construisons un nouvel estimateur qui combine les deux sous-ensembles avec deux noyaux asymétriques BSPE et gamma. Cependant, le paramètre de lissage sera estimé par l'approche bayésienne adaptative, développée à l'aide des deux noyaux proposés et de la méthode classique UCV. Pour évaluer les performances de l'estimateur proposé, une étude comparative avec l'estimateur classique sur des données simulées et réelles est réalisée.

**Mots-clés.** Approche bayesienne, données à queue lourde, k-means, noyau BSPE, noyau gamma.

**Abstract.** In this work, we propose to estimate the density function of heavy tailed data with non-negative support. As the heavy tailed data are characterized by rare observations in the tail, we propose to classify them into two subsets with high and low density, using k-means method, an unsupervised machine learning classification method. To this end, we construct a new estimator that combines two asymmetric BSPE and gamma kernels. However, the smoothing parameter will be estimated by the adaptive Bayesian approach, developed using the two proposed kernels and the classical UCV method. A comparative study between the proposed estimator and the classical estimator on simulated and real data is performed to evaluate their performance.

**Keywords.** BSPE kernel, gamma kernel, bayesian approach, UCV, heavy tailed data, k-means.

# 1 Introduction

The estimation of the probability density of heavy-tailed data is very complex due to their specific characteristics, such as rare observations in the tail. Non-parametric probability

density estimation by the kernel method is one of the most important technique for understanding the properties of the data distribution. However, a good estimation of the density depends on the correct choice of its parameters, kernel $K$ and smoothing parameter $h$. The use of symmetric kernels in the case of asymmetric data estimation, causes a serious problem at the edges, which is the edge bias problem. Edge bias is due to the allocation of weights by the symmetric kernel outside the density support when smoothing is carried out near the boundary. For this reason, a family of asymmetric kernels has been proposed in the literature.

In this work, we are interested in estimating the density of heavy-tailed data with support $[0, \infty)$, as these are characterised by sparse observations in the tail, then we propose to classify the data into two clusters with high and low density, using unsupervised machine learning classification methods. A good clustering approach is one that provides high homogeneity within the cluster and heterogeneity between clusters (Xu and Wunsch, 2005), (Sharma and ShikhaRai, 2012). The type of cluster we are interested in is the one that aims at optimising a given merit function, which will lead to a good clustering, i.e. observations of the same similarity are grouped in the same cluster. Different algorithms are based on this principle, such as k-means, k-center clustering (Gonzalez, 1985), (Mohiuddin et al., 2020), (P and Miin-Shen, 2020), etc. The k-means method was used to divide the heavy tail data into two subsets with high (HDR) and low (LDR) density where we will associate the BSPE and Gamma kernels with the (HDR) and (LDR) regions respectively.

The smoothing parameter plays a key role in non-parametric kernel estimation, it controls the degree of smoothing. Several methods have been discussed in the literature. Classical methods which generates a global smoothing parameter, see for example (Saulo et al., 2013). However, a problem in using a global bandwidth is that the kernel methods often produce unsatisfactory results for complex or irregular densities. For this reason, several authors have been motivated to propose an estimator of the variable smoothing parameter, see for exemple, (Somé, 2022), (Ziane et al., 2015), (Yasmina et al., 2018) and (Belaid et al., 2016; Belaid et al., 2018) who have invested in the estimation of the variable smoothing parameter by the Bayesian approach in the asymmetric case and (Brewer, 2000); (Zhang et al., 2006) and (Zougab et al., 2014) for symmetric kernels.

The main objective of this work, is to improve the quality of the estimator, by proposing a new estimator that combines two classical estimators associated to each cluster (HDR and LDR) with the asymmetric BSPE and gamma kernels respectively. The smoothing parameter will be estimated adaptively by the Bayesian approach using the explicit forms determined by (Somé, 2022) with gamma kernel (LDR) and (Ziane et al., 2015) with BSPE kernel (HDR).

The rest of this paper is organized as follows. In Section 2, we give a brief review of S-PE and gamma kernels density estimator. In Section 3, we present the k-means classification method. In Section 4, we first introduce the proposed kernel density estimators based on BSPE and gamma kernels. Second, we show some properties of $\hat{f}_{BSPE-Gamma}$ (bias and variance). We adapt the adaptive bayesian for the choice of bandwidth in Section 5. The performance of the proposed estimator will be tested via real and simulated data sets in Section 6, while Section 7 concludes.

## 2 A brief review: BSPE and gamma kernels density estimator

In this section, we present a brief recall of the classical $\hat{f}$ estimator with BS-PE kernels and gamma kernel.
Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) continuous random variables with an unknown probability density function (pdf) $f$ on the support $[0, \infty)$. The kernel estimator based on asymmetric densities is expressed as

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i) \tag{1}$$

with kernels defined in table 1

Table 1: Kernels

| Distribution | Kernel |
|---|---|
| Gamma | $\frac{y^{\frac{x}{h}}}{\Gamma(1+\frac{x}{h})h^{1+\frac{x}{h}}} \exp\left(-\frac{y}{h}\right)$ |
| BS−PE | $\frac{\nu}{2^{\frac{1}{2\nu}}\Gamma(\frac{1}{2\nu})\sqrt{4h}} \left(\frac{1}{\sqrt{xy}} + \sqrt{\frac{x}{y^3}}\right) \exp\left(\frac{-1}{2h^{\nu}}\left(\frac{y}{x} + \frac{x}{y} - 2\right)^{\nu}\right)$ |

## 3 Data clustering with machine learning

Clustering is a common technique in statistical data analysis that is used in many fields. Data clustering is based on partitioning a dataset into clusters, where the elements of each cluster are similar to each other and different from other clusters. The case of unlabeled data, leads to an unsupervised machine learning classification. Several unsupervised classification algorithms have been proposed: hierarchical clustering algorithm, K-means algorithm, K-medoids algorithm, etc.
The objective of our work is to propose a subdivision of the data by unsupervised machine learning classification methods. As heavy-tailed data are characterised by low density at the tail, a partitioning of these data into two subsets with high and low density by the k-means method is interesting. The k-means clustering algorithm has the following steps:

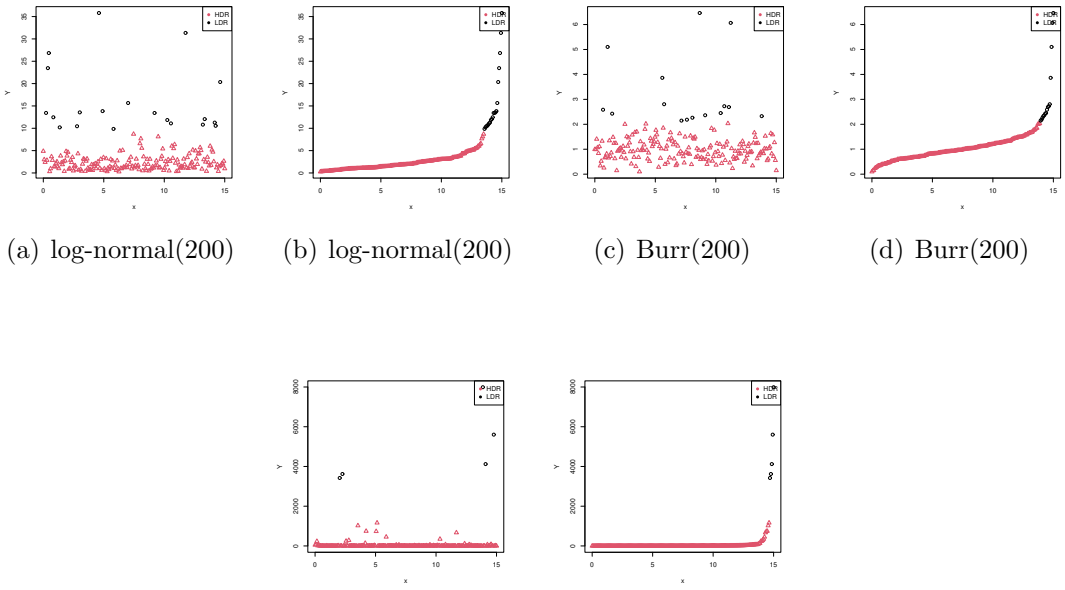
**a)** Define the number of clusters

**b)** Establish the centroid coordinates

**c)** Determine the distance between each observation and the centroid.

**d)** Grouping observations according to the minimum distance

In the case of heavy-tailed data, the number of clusters $k = 2$. Figure 1, illustrates the distribution of the data into two subsets of the different laws, which are characterised by a heavy tail (log-normal, burr and levy) and for sample sizes (n=200).




(a) log-normal(200) (b) log-normal(200) (c) Burr(200) (d) Burr(200)

(e) Levy(200) (f) Levy(200)

Figure 1: K-means repartition in high density region (HDR) and low density region (LDR).

From the graphs in figure 1, it can be seen that there is a low density at the extreme right (rare observations in the tail), in contrast to the extreme left where there is a high density.

# 4 The BSPE-Gamma kernel density estimation

We propose a new estimator of the density function, with a kernel function that would be flexible near zero and another kernel function that could estimate the tail of the density. Let $d$ be a threshold value that determines the proportion of the high density region in the sample space that is determined by the unsupervised machine learning classification method. The kernel density estimator is given by:

$$\hat{f}_{BSPE-Gamma}(x) = \begin{cases} \hat{f}_{(h,BSPE)}(x), & \text{if } x \in [0,d]; \\ \hat{f}_{(h,Gamma)}(x), & \text{if } x > d. \end{cases} \tag{2}$$

The bias and variance of the estimator $\hat{f}_{BSPE-Gamma}$ given by (2) are

$$biais\left(\hat{f}_{BSPE-Gamma}(x)\right) = \begin{cases} biais\left(\hat{f}_{(h,BSPE)}(x)\right), & \text{if } x \in [0,d]; \\ biais\left(\hat{f}_{(h,Gamma)}(x)\right), & \text{if } x > d. \end{cases} \tag{3}$$

$$biais\left(\hat{f}_{BSPE-Gamma}(x)\right) = \begin{cases} \frac{h\mu_1(g)}{2}\left(xf'(x) + x^2 f''(x)\right) + o(h), & \text{if } x \in [0,d]; \\ h\left(f'(x) + \frac{1}{2}xf''(x)\right) + o(h), & \text{if } x > d. \end{cases} \tag{4}$$

$$Var\left(\hat{f}_{BSPE-Gamma}(x)\right) = \begin{cases} Var\left(\hat{f}_{(h,BSPE)}(x)\right), & \text{if } x \in [0,d]; \\ Var\left(\hat{f}_{(h,Gamma)}(x)\right), & \text{if } x > d. \end{cases} \tag{5}$$

$$Var\left(\hat{f}_{BSPE-Gamma}(x)\right) = \begin{cases} \frac{c^2}{C_{g^2} nh^{1/2}x} f(x) + o\left(\frac{1}{nh^{1/2}}\right), & \text{if } x \in [0,d]; \\ \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1}h^{-1/2}x^{-1/2} f(x) & \text{if } \frac{x}{h} \to \infty \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)} n^{-1}h^{-1} f(x) & \text{if } \frac{x}{h} \to \kappa \end{cases}, & \text{if } x > d. \end{cases} \tag{6}$$

where $\kappa$ is a nonnegative constant.

| $\mu_1(g)$ | $c$ | $C_{g^2}$ |
|---|---|---|
| $\frac{2^{1/nu}\Gamma(\frac{3}{2\nu})}{\Gamma(\frac{1}{2\nu})}$ | $\frac{\nu}{2^{1/2\nu}\Gamma(\frac{1}{2\nu})}$ | $2^{1/\nu}\Gamma(\frac{2\nu+1}{2\nu})\cos(\frac{\pi}{2\nu})$ |

# 5 Adaptive bayesian bandwidth selection

In this section, we present the adaptive smoothing parameters $(h_i)$ estimated by the adaptive Bayesian approach. The adaptive smoothing parameter represents the window $h$ estimated in each observation $X_i$ which gives a vector of the smoothing parameter $(h_1, h_2, \ldots, h_n)$. The explicit form of $h_i$ with the BSPE and gamma kernels developed respectively by (Ziane et al., 2015) and (Somé, 2022), are given by:

$$\hat{h}_{i,BSPE} = \frac{1}{\beta^{1/\nu}} \frac{\Gamma(\alpha - \frac{1}{2}\nu) \sum\limits_{j=1, i\neq j}^{n} \left(\frac{1}{\sqrt{x_i x_j}} + \sqrt{\frac{X_i}{X_j}}\right) \left[\frac{1}{2}\left(\frac{X_j}{X_i} + \frac{X_i}{X_j} - 2\right)^{\nu} + 1\right]^{-\alpha+\frac{1}{2\nu}}}{\Gamma(\alpha + \frac{1}{2}\nu) \sum\limits_{j=1, i\neq j}^{n} \left(\frac{1}{\sqrt{X_i X_j}} + \sqrt{\frac{X_i}{X_j}}\right) \left[\frac{1}{2}\left(\frac{X_j}{X_i} + \frac{X_i}{X_j} - 2\right)^{\nu} + 1\right]^{-\alpha-\frac{1}{2\nu}}} \tag{7}$$

$$\hat{h}_{i,Gamma} = \frac{1}{D_{ij}(\alpha,\beta)} \sum_{j=1, i\neq j}^{n} \left\{ \frac{(X_j+\beta)C_j(\alpha,\beta)}{\alpha} \mathbf{1}_{\{0\}}(X_i) + \frac{A_{ij}(\alpha,\beta)B_{ij}(\beta)}{\alpha - 1/2} \mathbf{1}_{(0,\infty)}(X_i) \right\} \tag{8}$$

where

$$\begin{aligned} A_{ij}(\alpha,\beta) &= \frac{\Gamma(\alpha+1/2)}{\beta^{\alpha} X_i^{1/2}\sqrt{2\pi}(B_{ij}(\beta))^{\alpha+\frac{1}{2}}}, \\ B_{ij}(\beta) &= X_i \log X_i - X_i \log X_j + X_j - X_i + \beta, \\ C_j(\alpha,\beta) &= \frac{\Gamma(\alpha+1)}{\beta^{-\alpha}(X_j+\beta)^{\alpha+1}}, \\ D_{ij}(\alpha,\beta) &= \sum_{j=1, i\neq j}^{n} \left(A_{ij}\mathbf{1}_{(0,\infty)}(X_i) + C_j \mathbf{1}_{(0)}(X_i)\right). \end{aligned}$$

$h_{i,BSPE}$ is the smoothing parameter vector associated to the high density region and $h_{i,Gamma}$ is the smoothing parameter vector associated to the low density region.

# 6 Simulation study

In this section, we present the simulation results obtained by evaluating the performance of the proposed estimator, compared to the classical estimator. This simulation study is based on samples of random variables simulated from heavy-tailed distributions, lognormal, burr and Levy presented in the table 2, for different sample sizes $n = 10, 50, 100$ and 500, and for a number of repetitions $N = 100$. The performances of the estimators are examined via integrated squared error (ISE) given by:

$$ISE := \int \left\{ \hat{f}(x) - f(x) \right\}^2 \tag{9}$$

where $\hat{f}$ is the BSPE-gamma, BSPE or gamma kernel estimators. the smoothing parameter $h$ is estimated by the adaptive Bayesian approach with the combination of BSPE-Gamma kernels which will be compared with the classical UCV approach.

Table 2: Distributions in the simulation study.

| | Distribution | Density | Parameters |
|---|---|---|---|
| **D1** | lognormal | $\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}\left(\ln(x)-\mu\right)^2\right),\ x>0$ | $(\mu,\sigma) = (1,1)$ |
| **D2** | Burr | $\frac{kx^{k-1}}{(1+rx^k)^{r+1}},\ x>0$ | $(k,r) = (3,1)$ |
| **D3** | Levy | $\sqrt{\frac{c}{2\pi}} \frac{1}{(x-\mu)^{3/2}} \exp(-\frac{c}{2(x-\mu)}),\ x>\mu$ | $(\mu,c) = (0,1/2)$ |

**remark:**
The performance of the adaptive Bayesian approach, depends on the choice of the a priori law parameters, used to develop the explicit form of the smoothing parameter (7) and (8). In this work, we followed the same principle as (Ziane et al., 2015) and (Somé, 2022).

The results of the comparative study between the proposed estimator and the classical BSPE and Gamma kernel estimator, with UCV for the bandwidth $h$ selection are presented in table 3, which reports averages of ISE and standard deviations values. From these we can observe:

- For the **D3** model, the proposed estimator is better than the classical estimator for all sizes of the sample.
- The proposed estimator outperforms the classical estimator for the **D1** and **D4** models for all sample sizes except for a large size $n = 500$.
- For the **D2** model, the classical estimator works best for medium and large sample sizes (100 and 500).

Table 3: Some expected values of ISE and their standard errors between parentheses based on 100 replications for the **D1**, **D2**, **D3** and **D4** distributions.

| $f$ | $n$ | $ISE_{BSPE-Gamma(UCV)}$ | $ISE_{BSPE(UCV)}$ | $ISE_{Gamma(UCV)}$ |
|---|---|---|---|---|
| | 10 | 0.0224 (0.0184) | 0.0331 (0.0229) | 0.0306 (0.0327) |
| **D1** | 50 | 0.0090 (0.0056) | 0.0094 (0.0062) | 0.0130 (0.0103) |
| | 100 | 0.0058 (0.0039) | 0.0061 (0.0053) | 0.0074 (0.0060) |
| | 500 | 0.0031 (0.0011) | 0.0017 (0.0012) | 0.0023 (0.0014) |
| | 10 | 0.0931 (0.0439) | 0.0984 (0.0807) | 0.1122 (0.0872) |
| **D2** | 50 | 0.0409 (0.0487) | 0.0260 (0.0169) | 0.0371 (0.0805) |
| | 100 | 0.0247 (0.0191) | 0.0160 (0.0116) | 0.0197 (0.0266) |
| | 500 | 0.0119 (0.0075) | 0.0068 (0.0046) | 0.0103 (0.0103) |
| | 10 | 0.0808 (0.0418) | 0.0918 (0.0441) | 0.1014 (0.0598) |
| **D3** | 50 | 0.0331 (0.0279) | 0.0341 (0.0358) | 0.0483 (0.0274) |
| | 100 | 0.0124 (0.0145) | 0.0159 (0.0141) | 0.0289 (0.0183) |
| | 500 | 0.0043 (0.0080) | 0.0053 (0.0042) | 0.0078 (0.0040) |
| | 10 | 0.0299 (0.0454) | 0.0485 (0.0456) | 0.0306 (0.0329) |
| **D4** | 50 | 0.0086 (0.0054) | 0.0147 (0.0163) | 0.0096 (0.0095) |
| | 100 | 0.0081 (0.0081) | 0.0090 (0.0069) | 0.0086 (0.0093) |
| | 500 | 0.0041 (0.0014) | 0.0029 (0.0021) | 0.0028 (0.0029) |

- If we compare the classical estimators with the BSPE and gamma kernels, we notice that the BSPE kernel, estimates the density better, in the case of models **D2**, **D3** and for model **D1** just for $n = 50$ and 100.

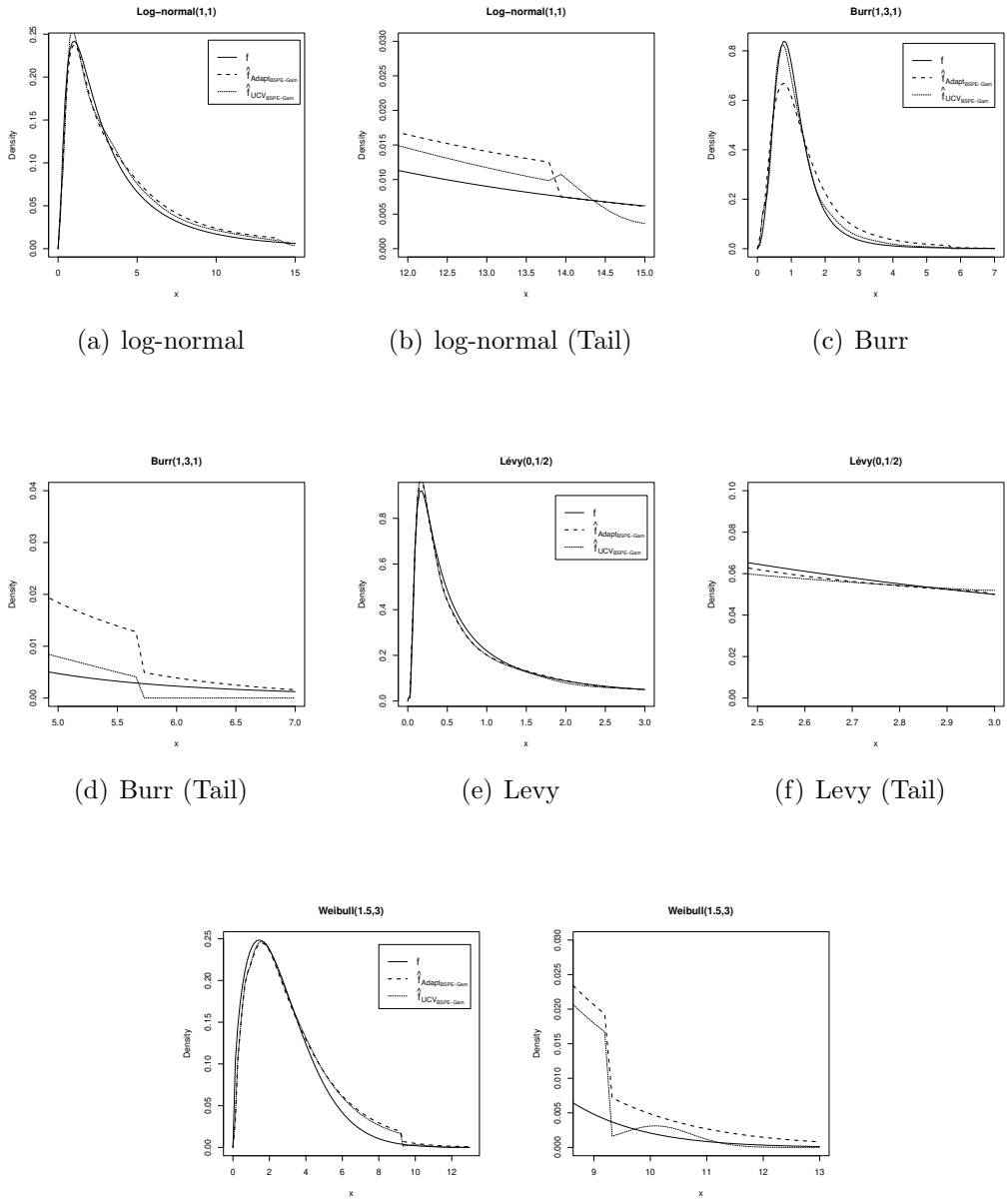
Table 4 presents averages ISE and standard deviations values resulting from comparison of the proposed estimator $\hat{f}_{BSPE-Gamma}$ with two smoothing parameter selection methods, the adaptive Bayesian approach and the classical UCV approach. From the results obtained, we can see that, the adaptive Bayesian approach is better than the classical UCV approach for all the models considered except **D2** for the sizes $n = 100$ and 500.

Table 4: Some expected values of ISE and their standard errors between parentheses based on 100 replications for the **D1**, **D2**, **D3** and **D4** distributions.

| $f$ | $n$ | $ISE_{BSPE-Gamma(bayes-adaptif)}$ (Std) | $ISE_{BSPE-Gamma(UCV)}$ (Std) |
|---|---|---|---|
| | 10 | 0.0215 (0.0176) | 0.0292 (0.0256) |
| **D1** | 50 | 0.0078 (0.0043) | 0.0122 (0.0057) |
| | 100 | 0.0047 (0.0028) | 0.0064 (0.0037) |
| | 500 | 0.0042 (0.0012) | 0.0056 (0.0023) |
| | 10 | 0.0873 (0.0355) | 0.1056 (0.0993) |
| **D2** | 50 | 0.0783 (0.0224) | 0.0934 (0.0360) |
| | 100 | 0.0580 (0.0184) | 0.0394 (0.0109) |
| | 500 | 0.0477 (0.0116) | 0.0348 (0.0057) |
| | 10 | 0.0774 (0.0458) | 0.1056 (0.0688) |
| **D3** | 50 | 0.0434 (0.0225) | 0.0452 (0.0463) |
| | 100 | 0.0100 (0.0134) | 0.0163 (0.0129) |
| | 500 | 0.0057 (0.0099) | 0.0057 (0.0038) |
| | 10 | 0.0246 (0.0206) | 0.0380 (0.0436) |
| **D4** | 50 | 0.0104 (0.0090) | 0.0193 (0.0331) |
| | 100 | 0.0064 (0.0031) | 0.0080 (0.0059) |
| | 500 | 0.0041 (0.0009) | 0.0057 (0.0024) |

In addition, We also compared the estimators graphically as shown in Figure 2, which plots the estimators of the lognormal, Burr, Levy and Weibull distributions for $n = 200$ and one replication, with Bayesian adaptive and UCV approaches for the choice of $h$. The left is the estimators of the distributions, and on the right is the zoom of the estimators tail. We note that the estimators are able to reproduce the uni-modality of the considered models. We can observe that the quality of smoothing is satisfactory, in particular at the tail of the

distributions, where the estimators estimate the tail well, whether it is with UCV or adaptive bayes approach.




(a) log-normal (b) log-normal (Tail) (c) Burr

(d) Burr (Tail) (e) Levy (f) Levy (Tail)

(g) Weibull (h) Weibull (Tail)

Figure 2: True pdf and kernel estimators for **D1**, **D2**, **D3** and **D4** with BSPE-gamma kernel, adaptive bayesian and UCV for $h$, for a sample size $n = 100$.

# 7 Real data

This section illustrates the performances of new estimators for real data set. These data present the vinyl chloride data obtained from clean upgrading, monitoring wells in mg/l; this

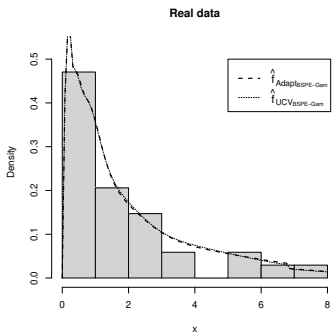



Figure 3: $\hat{f}_{BSPE-Gamma}$ estimator of vinyl chloride data with bayesian and UCV approaches

data set was used by (Ziane et al., 2021). Table 5 provides the descriptive summaries of vinyl chloride data.

Table 5: Descriptive summary of the vinyl chloride data set.

| Data set | $n$ | Max | Min | Median | Mean | SD | CS | CK |
|---|---|---|---|---|---|---|---|---|
| vinyl chloride | 34 | 8.000 | 0.100 | 1.150 | 1.879 | 1.952 | 1.603 | 5.005 |

Figure 3, shows the histogram and the $\hat{f}_{BSPE-Gamma}$ estimator with the adaptive Bayesian and UCV approaches, for the selection of the smoothing parameter on the real Vinyl chloride data. From this result, we can see that the estimators were able to reproduce the uni-modality and the tail of the data, the quality of the estimation is satisfactory with both considered approaches.

# 8 CONCLUSION

In this work, we proposed to estimate the density of data that are characterized by a heavy tail. Given this characteristic, we opted to divide the data set into two clusters (high density region (HDR) and low density region (LDR)) using the unsupervised k-means classification method. A new density estimator is proposed, associating to HRD the BSPE kernel and LDR the gamma kernel, the smoothing parameter is selected by the UCV and adaptive Bayesian approaches. A comparative study is performed to show the advantages of the proposed estimator over the classical estimator on simulated and real data.

# References

Belaid, N., Adjabi, S., Kokonendji, C. C., and Zougab, N. (2018). Bayesian adaptive bandwidth selector for multivariate discrete kernel estimator. *Communications in Statistics-Theory and Methods*, 47:2988–3001.

Belaid, N., Adjabi, S., Zougab, N., and Kokonendji, C. C. (2016). Bayesian bandwidth selection in discrete multivariate associated kernel estimators for probability mass functions. *Journal of the Korean Statistical Society*, 45:557–567.

Brewer, M. J. (2000). A bayesian model for local smoothing in kernel density estimation. *Statistics and Computing*, 10:299–309.

Gonzalez, T. F. (1985). Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306.

Mohiuddin, A., Raihan, S., and Shamsul, I. S. M. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9:1295.

P, S. K. and Miin-Shen, Y. (2020). Unsupervised k-means clustering algorithm. *IEEE access*, 8:80716–80727.

Saulo, H., Leiva, V., Ziegelmann, F. A., and Marchant, C. (2013). A nonparametric method for estimating asymmetric densities based on skewed birnbaum–saunders distributions applied to environmental data. *Stochastic Environmental Research and Risk Assessment*, 27:1479–1491.

Sharma, S. and ShikhaRai (2012). Genetic k-menas algorithm implementation and analysis. *International Journal of Recent Technology and Engineering (IJRTE)*, 1(2):2277–3878.

Somé, S. M. (2022). Bayesian selector of adaptive bandwidth for gamma kernel density estimator on [0,): simulations and applications. *Communications in Statistics-Simulation and Computation*, 51:7287–7297.

Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms. *IEEE transactions on neural networks*, 16(3):645–678.

Yasmina, Z., Nabil, Z., and Smail, A. (2018). Birnbaum–saunders power-exponential kernel density estimation and bayes local bandwidth selection for nonnegative heavy tailed data. *Computational Statistics*, 33:299–318.

Zhang, X., King, M. L., and Hyndman, R. J. (2006). A bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics and Data Analysis*, 50:3009–3031.

Ziane, Y., Adjabi, S., and Zougab, N. (2015). Adaptive bayesian bandwidth selection in asymmetric kernel density estimation for nonnegative heavy-tailed data. *Journal of Applied Statistics*, 42:1645–1658.

Ziane, Y., Zougab, N., and Adjabi, S. (2021). Body tail adaptive kernel density estimation for nonnegative heavy-tailed data. *Monte Carlo Methods and Applications*, 27(1):57–69.

Zougab, N., Adjabi, S., and Kokonendji, C. C. (2014). Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. *Computational Statistics and Data Analysis*, 75:28–38.