Journal of Innovative Applied Mathematics and Computational Sciences

J. Innov. Appl. Math. Comput. Sci. 2(2) (2022), 38–47. n2t.net/ark:/49935/jiamcs.v2i2.32

http://jiamcs.centre-univ-mila.dz/

Adaptive gamma-BSPE kernel density estimation for nonnegative heavy-tailed data

Ziane Yasmina [1], Zougab Nabil [2] and Adjabi Smail [1]

[1]Research unit LaMOS, Operational Research departement, Faculty of Exact Sciences, Bejaia University, Algeria

[2]Research unit LaMOS, Department of Technology, University of Bejaia, Algeria

Received 04 February 2022, Accepted 02 June 2022, Published xx September 2022

Abstract. In this work, we consider the nonparametric estimation of the probability density function for nonnegative heavy-tailed (HT) data. The objective is first to propose a new estimator that will combine two regions of observations (high and low density). While associating a gamma kernel to the high-density region and a BS-PE kernel to the low-density region. Then, to compare the proposed estimator with the classical estimator in order to evaluate its performance. The choice of bandwidth is investigated by adopting the popular cross-validation technique and two variants of the Bayesian approach. Finally, the performances of the proposed and the classical estimators are illustrated by a simulation study and real data.

Keywords: Bayesian bandwidth selector, BS-PE kernel, Cross validation, Gamma kernel, heavy-tailed data, MCMC method.

2020 Mathematics Subject Classification: 62G07, 62G99.

1 Introduction

In this work, we are interested in estimating the heavy-tailed data density with nonnegative support [7] and [8]. This data type requires special methods because of its specific characteristics: slow decay to zero and rare observations in the tail. As the parametric methods do not meet the characteristics of this data type, the nonparametric kernel method is proposed. The efficiency of the latter depends on the choice of its two parameters, the kernel K and the smoothing parameter h. The most used kernels in the literature are the symmetric kernels, such as the Gaussian kernel and the Epanechnicov kernel for unbounded support densities. However, when we want to estimate densities with unbounded support, the classical kernel estimator becomes non-consistent, because of edge effects. This problem is due to symmetric kernels, which assign a weight outside the support when the smoothing is considered near the edge. To address this problem, several authors have proposed a new family of asymmetric kernels. See [2] (gamma and modified gamma kernels), [6] (inverse and reciprocal inverse

✉ Corresponding author. Email: yasmina-ziane@hotmail.com

ISSN (electronic): 2773-4196
© 2022 Published under a Creative Commons Attribution-Non Commercial-NoDerivatives 4.0 International License by the Institute of Science and Technology, Mila University Center Publishing

Gaussian kernels), [3] (lognormal and Birnbaum-Saunders (BS) kernels) and recently [4] proposed a Generalized Birnbaum-Saunders (GBS) kernel for estimating densities with nonnegative support, which includes BS-power-exponential (BS-PE) and BS-Student (BS-t) kernels. It is proposed for analyzing nonnegative heavy-tailed (HT) data.

The performance of the associated kernel density estimator depends crucially on the smoothing parameter, which controls the smoothing quality of the estimator. Classical methods have been proposed for the smoothing parameter choice. The cross-validations methods are interesting in practice because they are guided only by observations. However, the drawback of these methods is that they tend to provide under or over-smoothed estimators when the data are small or medium size or when we want to estimate complex functions. So, to deal with this problem, the Bayesian approach has been proposed.

As a base for this work, we followed the idea in [9], where the authors proposed a subdivision of the HT dataset into two subsets (two regions) with low and high density (Low Density Region (LDR) and High Density Region (HDR)), and associated to each region a smoothing parameter (h_{LDR} and h_{HDR}). We propose an estimator composed of two different kernels, gamma, and BSPE. The gamma kernel is associated with the high density region, and the BSPE kernel is associated with the high density region (see also [5]). The new Gamma-BSPE kernel density has two smoothing parameters (bandwidths) that will be selected using the adaptive Bayesian approach. A comparative study is conducted with the work of [9], where they considered a single BSPE kernel for both regions.

The paper is structured as follows. Section 2 presents the classical BSPE kernel estimator. In section 3, we introduce the new gamma-BSPE kernel estimator. Section 4 presents the procedure proposed for deriving the adaptive bandwidths. Simulation studies and application of real data are presented in Sections 5 and 6. Section 7 concludes the paper.

2 The classical BS-PE Kernel estimator

Given a random sample $X_1, \ldots, X_n$, the BS-PE kernel estimator of an unknown pdf f with nonnegative support is given by:

$$\begin{aligned} \hat{f}_{BS-PE,h}(x) &= \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i) \\ &= \frac{\nu}{n 2^{\frac{1}{2\nu}}\Gamma(\frac{1}{2\nu})\sqrt{4h}}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{xX_i}}+\sqrt{\frac{x}{X_i^3}}\right)\exp\left(\frac{-1}{2h^{\nu}}\left(\frac{X_i}{x}+\frac{x}{X_i}-2\right)^{\nu}\right), \end{aligned} \tag{2.1}$$

where $x > 0$ is the point where the density is estimated, $h > 0$ is a smoothing parameter and $\nu > 0$ is a fixed parameter.

The expressions of the bias and variance for $\hat{f}_{BS-PE}(x)$ are derived by Marchant et al. [4]. The asymptotic bias when $h \to 0$ is given by:

$$Bias(\hat{f}_{BS-PE}(x)) = \frac{hu_1(g)}{2}\left(xf'(x) + x^2 f''(x)\right) + o(h),$$

$$Var(\hat{f}_{BS-PE}(x)) = \frac{c^2}{c_{g^2} n h^{1/2} x} f(x) + o\left(\frac{1}{nh^{1/2}}\right),$$

where, $u_1(g) = \frac{2^{\frac{1}{\nu}}\Gamma(\frac{3}{2\nu})}{\Gamma(\frac{1}{2\nu})}$, $c^2 = \frac{\nu}{2^{\frac{1}{2\nu}}\Gamma(\frac{1}{2\nu})}$ and $c_{g^2} = \frac{\nu}{\Gamma(\frac{1}{2\nu})}$.

3 The gamma-BSPE Kernel estimator

In this section, we present a new density estimator for heavy-tailed data, which is flexible on the domain near the zero boundary and estimates the heavy tail of the distribution. The latter is based on: dividing the observations into two regions, namely the low-density region (LDR) and high-density region (HDR), and assigning two different bandwidths to these two regions [9]. We also propose to combine two asymmetric gamma and BS-PE kernels ([2] and [4]) as follows: associate a gamma kernel for the high-density region (HDR) (near bord) and BS-PE kernel for the low-density region (LDR).

3.1 Gamma kernel

The gamma kernel is nonnegative and possesses good boundary properties for a wide class of densities. It is given by:

$$K_{Gam(x,h)}(y) = \frac{y^{\frac{x}{h}}}{\Gamma(1+\frac{x}{h})h^{1+\frac{x}{h}}} \exp\left(-\frac{y}{h}\right) \mathbf{1}_{\{0\leq x<\infty\}}(y), \tag{3.1}$$

where $\Gamma(y) = \int_0^\infty t^{y-1}\exp(-t)dt$ is the classical gamma function with $y > 0$, and $\mathbf{1}_{\{0\leq x<\infty\}}$ denotes the indicator function.
The classical gamma kernel estimator of an unknown pdf f with nonnegative support is given by:

$$\hat{f}_{Gam,h}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i^{\frac{x}{h}}}{\Gamma(1+\frac{x}{h})h^{1+\frac{x}{h}}} \exp\left(-\frac{X_i}{h}\right) \mathbf{1}_{\{0\leq x<\infty\}}(X_i). \tag{3.2}$$

3.2 Gamma-BSPE kernel estimator

After dividing the data set into two subsets, we present the estimator associated with this subdivision by associating different kernels to the two regions, gamma kernel for the HDR region and BSPE for the LDR region. The gamma-BSPE Kernel estimator is given by:

$$\begin{aligned} \hat{f}_{h^{(0)},h^{(1)}}(x) &= \frac{1}{n}\sum_{j=1}^{n}\left\{ I_j K_{x,h^{(1)}}(x_j) + (1-I_j)K_{x,h^{(0)}}(x_j)\right\}, \\ &= \frac{1}{n}\sum_{j=1}^{n}\left\{ I_j K_{Gam(x,h^{(1)})}(x_j) + (1-I_j)K_{BS-PE(x,h^{(0)})}(x_j)\right\}. \end{aligned} \tag{3.3}$$

where

$$I_j = \begin{cases} 1, & if \quad x_j \in S_{(HDR)}, \quad j = 1,\dots,n; \\ 0, & else. \end{cases}$$

$S_{(HDR)}$: the observations of the high-density region (HDR), $S_{(LDR)}$: the observations of the low-density region (LDR)
and $h^{(1)}$ denotes the bandwidth assigned to the observations of $S_{(HDR)}$, and $h^{(0)}$ is the bandwidth assigned to the observations of $S_{(LDR)}$.

4 Adaptive Bayesian bandwidth selection

In this section, we derive the variable Bayesian bandwidths at each subset ($S_{(HDR)}$ and $S_{(LDR)}$) (Bayesian adaptive approach) for Equation (3.3) in the kernel density estimation context, with positive support using the gamma-BSPE kernels. We treat $h^{(1)}$ and $h^{(0)}$ as random quantities with prior distributions $\pi_1(\cdot)$ and $\pi_0(\cdot)$. As proposed by [7], we assume that the variable bandwidths $h^{(1)}$ and $h^{(0)}$ have prior distributions with parameters α, β, and $\nu = 2$; this prior is defined by

$$\pi(h^{(0)}) = \frac{\nu}{\Gamma(\alpha)\beta^{\alpha}} \frac{1}{(h^{(0)})^{\alpha\nu+1}} \exp\left(\frac{-1}{\beta(h^{(0)})^{\nu}}\right), \quad h^{(0)} > 0 \tag{4.1}$$

and

$$\pi(h^{(1)}) = \frac{\nu}{\Gamma(\alpha)\beta^{\alpha}} \frac{1}{(h^{(1)})^{\alpha\nu+1}} \exp\left(\frac{-1}{\beta(h^{(1)})^{\nu}}\right), \quad h^{(1)} > 0 \tag{4.2}$$

The posterior of $h^{(1)}$ and $h^{(0)}$ for given $\{x_1, x_2, \ldots, x_n\}$ is

$$\hat{\pi}(h^{(1)}, h^{(0)}|x_1, x_2, \ldots x_n) \propto \left\{\prod_{i=1}^{n} \hat{f}_{h^{(0)},h^{(1)}}(x_i)\right\} \pi(h^{(0)})\pi(h^{(1)}). \tag{4.3}$$

Under the squared error loss, the Bayes estimator of the smoothing parameters $h^{(1)}$ and $h^{(0)}$ is the mean of the posterior density given by:

$$\left(\hat{h}^{(1)}, \hat{h}^{(0)}\right) = \int\int (h^{(1)}, h^{(0)})\hat{\pi}(h^{(1)}, h^{(0)}|x_1, x_2, \ldots x_n) dh^{(1)} dh^{(0)}. \tag{4.4}$$

We cannot derive an analytical expression as the estimate of $\left(\hat{h}^{(1)}, \hat{h}^{(0)}\right)$ from the formula (4.3) and (4.4). However, we propose using the Markov Chain Monte Carlo method (MCMC) for the approximation. We use a random walk metropolis algorithm to sample $\left\{h^{(1)}, h^{(0)}\right\}$ and the sampling algorithm is briefly described below :

Step 01 Initialize $\mathbf{h}_{(0)}$, where $\mathbf{h} = \left(h^{(1)}, h^{(0)}\right)$.

Step 02 For $i \in \{1, \ldots, M\}$,

- **a)** Generate $\tilde{\mathbf{h}} \sim$ truncate Normal $(\mathbf{h}_{(i-1)}, \sigma^2)$.
- **b)** Calculate the acceptance probability $\alpha = min\{1, \frac{\pi(\tilde{\mathbf{h}}/x)}{\pi(\mathbf{h}_{(i-1)}/x)} \frac{truncate\ Normal(\mathbf{h}_{(i-1)}, \sigma^2)}{truncate\ Normal(\tilde{\mathbf{h}}, \sigma^2)}\}$.

$$\mathbf{h}_{(i)} = \begin{cases} \tilde{\mathbf{h}}, & \mu < \alpha,\ \mu \sim \mathrm{U}_{[0,1]}; \\ \mathbf{h}_{(i-1)}, & \text{else.} \end{cases}$$

Step 03 $i = i + 1$ and go to step 2.

Reject $(\mathbf{h}_{(0)}, \mathbf{h}_{(1)}, \ldots, \mathbf{h}_{(M_0)})$ which represents burn-in period, and estimate $\mathbf{h}$ by

$$\hat{\mathbf{h}} = \frac{1}{M - M_0} \sum_{i=M_0+1}^{M} \mathbf{h}_{(i)}.$$

Table 5.1: Distributions in the simulation study.

	Distribution	Density	Parameters
D1	*lognormal*(μ, σ)	$f_1(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(\ln(x)-\mu)^2\right).$	$(\mu, \sigma) = (1, 1)$
D2	*Burr*(k, r)	$\frac{kx^{k-1}}{(1+rx^k)^{r+1}}$	$(k, r) = (3, 1)$
D3	Mixture of *pGamma*(α_1) and *pGamma*(α_2)	$p \times \frac{x^{\alpha_1 - 1}\exp(-x)}{\Gamma(\alpha_1)} + p \times \frac{x^{\alpha_2 - 1}\exp(-x)}{\Gamma(\alpha_2)}$	$(\alpha_1, \alpha_2, p) = (2.5, 10, 0.5)$

5 Simulation study

In this section, we examine and compare the performances of the adaptive bandwidth approach for (gamma-BSPE kernel estimator and BSPE kernel estimator proposed in [9]), with a global bandwidth approach (bayesian global and classical UCV method), by using several nonnegative heavy-tailed distributions.
The optimal bandwidth selected by classical method UCV was obtained by :

$$h_{UCV} = \arg\min_h UCV(h),$$

where :

$$\mathrm{UCV}(h) \quad = \quad \int \hat{f}^2_{\{h^{(0)},h^{(1)}\}}(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{\{h^{(0)},h^{(1)},i\}}(X_i),$$

and $\hat{f}_{\{h^{(0)},h^{(1)},i\}}$ being computed as $\hat{f}_{\{h^{(0)},h^{(1)}\}}$ by excluding X_i.
We consider the target densities labeled **D1**, **D2** and **D3**. Functional forms of these densities are given by: table 5.1.
This comparison is based on the data simulated from **D1**, **D2** and **D3** and five samples sizes $n = 10$, 25, 50, 100 and $n = 500$, using $Nsim = 100$ replications. We examine the performance of these methods via the integrated square error (ISE) criterion, defined by:

$$ISE = \int \{\hat{f}_h(x) - f(x)\}^2 dx. \tag{5.1}$$

Table 2 presents the average ISE ($\overline{ISE}$) and the average bandwidth ($\bar{h}$) based on 100 replications for the estimators of **D1**, **D2** and **D3**. The burn-in period contains $M_0 = 1500$ iterations and the following $M = 3000$ iterations were recorded. From Table 2, we observe that:

- For all estimators, the means of ISE and h based on 100 replications decrease as sample size n increases.
- For all sample sizes and models considered, the adaptive Bayesian approach (BSPE and Gamma-BSPE) outperforms the global Bayesian approach and the classical UCV method.
- We notice that, mean h associated with the high density region (HDR) is smaller than the mean h associated with the low density region (HDR) for both adaptive Bayesian approaches (BSPE and Gamma-BSPE).
- A comparison between the UCV and global bayesian approaches. We notice that for almost all the considered models, the global bayesian approach is better than the UCV for small sample sizes, but for medium and large sample sizes, the UCV works better.

Table 5.2: Average ISE ($\overline{ISE}$) (with average h ($\bar{h}$) brackets) based on 100 replications for D1, D2 and D3 distributions

Density	n	$\overline{ISE}_{UCV}$ ($\bar{h}_{UCV}$)	$\overline{ISE}_{Bayes-global}$ ($\bar{h}_{Bayes-global}$)	$\overline{ISE}_{Bayes-Adap_{BSPE}}$ ($\bar{h}_{HDR}, \bar{h}_{LDR}$)	$\overline{ISE}_{Bayes-Adap_{Gam-BSPE}}$ ($\bar{h}_{HDR}, \bar{h}_{LDR}$)
	10	0.07298109	0.03339965	0.03116670	**0.02446340**
		(1.29691400)	(0.41601620)	(0.41572240,0.43166210)	(0.42456180, 0.43005550)
	50	0.01992840	**0.00867429**	0.00944265	0.00944339
D1		(0.66709700)	(0.56194740)	(0.21892850, 0.35730160)	(0.15519320, 0.35793270)
	100	0.00812029	0.013134670	0.00545582	**0.00481987**
		(0.14878390)	(1.23848300)	(0.18921140, 0.26361370)	(0.12187420, 0.26551030)
	250	0.00464116	0.01115482	0.00421646	**0.00226334**
		(0.09514660)	(1.09769077)	(0.90787800, 0.25401930)	(0.07194098, 0.24740547)
	10	0.18070430	0.09793666	**0.08833576**	0.09152228
		(0.63006541)	(0.47316392)	(0.42567424, 0.42613821)	(0.35051259, 0.41623607)
	50	0.03882595	0.02937818	**0.01933364**	0.02062382
D2		(0.06591049)	(0.23530874)	(0.33260233, 0.33640982)	(0.28967948, 0.34092399)
	100	0.04370892	0.01372667	**0.00810524**	0.01026966
		(0.02410700)	(0.20874785)	(0.24873629, 0.29847000)	(0.21676625, 0.29696059)
	250	0.02080307	0.00948743	0.00499513	0.00684447
		(0.018870869)	(0.18647112)	(0.18162649, 0.1989421610)	(0.16955425, 0.18021157)
	10	0.0232216915	0.02009696	0.01920234	**0.01701824**
		(1.167033916)	(0.80816964)	(0.36240572, 0.92421295)	(0.37011558, 0.40727976)
	50	0.02737774	0.00871603	0.00807053	**0.00781293**
D3		(1.09355602)	(0.15824070)	(0.15292585, 0.29172937)	(0.15211303,0.27504394)
	100	0.015952670	0.00477119	0.00370813	**0.00364616**
		(0.08969938)	(0.16248214)	(0.10336488, 0.29077918)	(0.10264756, 0.26801841)
	250	0.00968110	0.00428933	0.00238951	**0.00237538**
		(0.03910971)	(0.12370519)	(0.07658197, 0.19733326)	(0.09667585, 0.16960317)

- The adaptive Bayesian approach with two different kernels (gamma for HDR and BSPE for LDR), outperforms the adaptive Bayesian approach with the same kernel (BSPE for both HDR and LDR regions), for models D1 and D3 for almost all sizes considered. Contrary to the D2 model, where the adaptive Bayesian approach with the same kernel (BSPE) is better.

The comparison is also given in Figures 5.1 and 5.2, presenting the plots of the pdf estimates for D1, D2 and D3, with UCV and Bayesian method for the choice of bandwidth parameter. The results are given for sample size $n = 200$ and one replication. We can observe that the smoothing quality is satisfactory for the adaptive Bayesian approach, practically for the three considered models. The adaptive Bayesian Gamma-BSPE approach reproduces well the bimodality of the D3 model. We also notice that the smoothing quality by the classical UCV approach is poor for the D2 model.

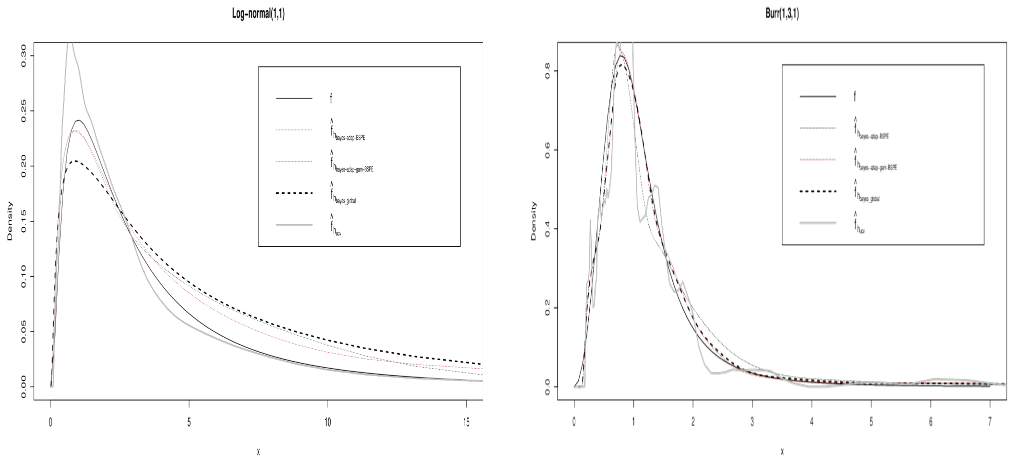


Figure 5.1: The estimators of heavy-tailed densities **D1** and **D2** with $n = 200$, with BS-PE kernel and (global, adaptive and UCV) methods.

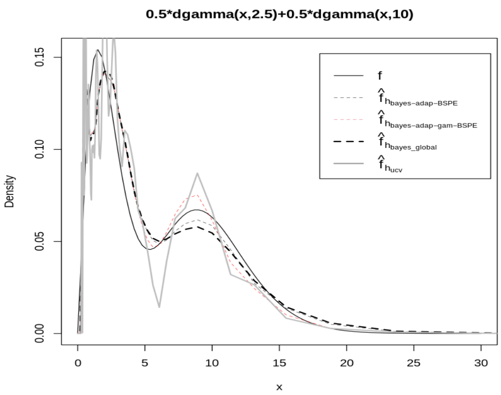


Figure 5.2: The estimators of heavy-tailed densities **D3** with $n = 200$, with BS-PE kernel and (global, adaptive and UCV) methods.

Table 6.1: Descriptive summary of the Web-traffic data set [8].

Data set	n	Max	Min	Median	Mean	SD	CS	CK
Web-traffic	312	65.613	0.042	1.362	4.081	8.520	5.044	27.897

Table 6.2: 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

6 Application to real HT data

In this section, we illustrate the performance of the proposed estimator on two real HT data sets defined below:

- **Web-traffic HT data:** These data represent the size of different web files (pdf, html, images, video, etc.) measured in Kilo Octet from the world cup (French, June 1998) server. These data are collected for $n = 312$ queries [8].

- **Vinyl chloride data:** These data present the vinyl chloride data obtained from clean upgrading and monitoring wells in mg/L; this data set was used by [1].

Tables 6.1 and 6.2 provide the description summaries for Web-traffic and vinyl chloride data, respectively.

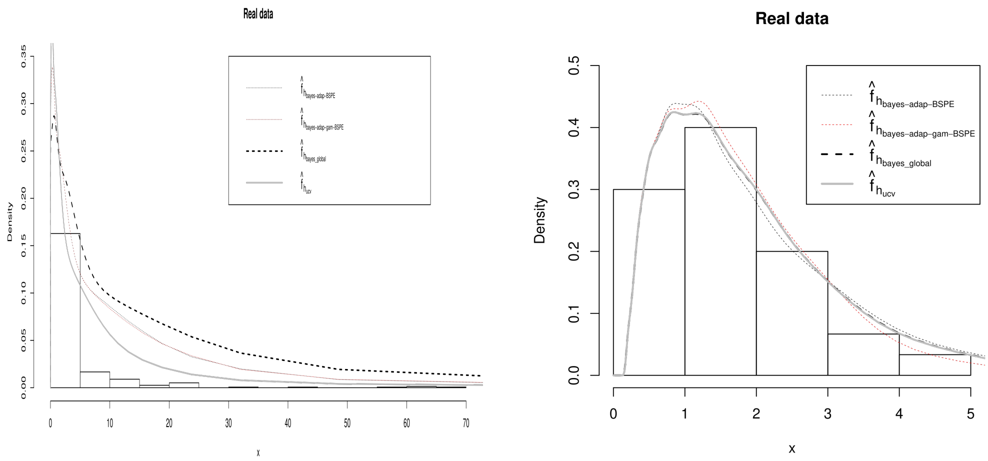
Now, we apply kernel estimators to estimate the density for traffic web and vinyl chloride data based on different selection methods of the smoothing parameter (UCV, Bayesian(global), and Bayesian adaptive(BS-PE($\nu = 2$) kernel and Gamma-BSPE($\nu = 2$) kernel)). The Bayes variable bandwidths estimates were obtained with prior parameters $\alpha = 2.5$ and $\beta = n^{4/5}$. The figure 6.1 shows, that all the methods can reproduce the unimodality of these data. We observe that the smoothing quality is satisfactory for almost all the considered methods.

Conclusions

In this paper, we have proposed a new gamma-BSPE kernel estimator. It is based on the principle of subdividing the HT dataset into two regions (LDR and HDR) and associating to each region the gamma and BSPE kernels. The smoothing parameter is determined using the adaptive Bayesian approach. The simulation study showed that the adaptive Bayesian approach with the gamma-BSPE kernels and the same BSPE kernels performs better than the global Bayesian and the classical UCV approaches. This study also showed that in some cases, the estimator with the gamma-BSPE kernels performs better than the estimator with the same BSPE kernels for both regions, contrary to other cases.

Acknowledgments

This paper has been presented in the International Conference on Mathematics and Applications (ICMA'2021), December 7-8, 2021, Blida1 University, Algeria.



(a) Web-traffic data (b) vinyl chloride data

Figure 6.1: kernel estimator for Web-traffic and vinyl chloride data with sample size $n = 312$ and $n = 34$ respectively.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

[1] D. K. Bhaumik, K. Kapur and R. D. Gibbons, *Testing parameters of a gamma distribution for small samples*, Technometrics, **51**(3) (2009), 326–334. DOI

[2] S. X. Chen, *Probability Density Function Estimation Using Gamma Kernels*, Ann. Inst. Statist. Math., **52**(3) (2000), 471–480. DOI

[3] X. Jin and J. Kawczak, *Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data*, Annals of Economics and Finance, **4** (2003), 103–124. URL

[4] C. Marchant, K. Bertin, V. Leiva and H. Saulo, *Generalized BirnbaumâĂŞSaunders kernel density estimators and an analysis of financial data*, Comput. Statist. Data Anal., **63** (2013), 1–15. [DOI]

[5] L. Markovich, *Gamma-weibull kernel estimation of the heavy-tailed densities*, arXiv preprint arXiv: 1604.06522. (2016), 1–10. DOI

[6] O. Scaillet, *Density estimation using inverse and reciprocal inverse Gaussian kernels*, J. Nonparametr. Stat., **16**(1-2) (2004), 217–226. DOI

[7] Y. Ziane, S. Adjabi and N. Zougab, *Adaptive Bayesian bandwidth selection in asymmetric kernel density estimation for nonnegative heavy-tailed data*, J. Appl. Stat., **42**(8) (2015), 1645–1658. DOI

[8] Y. Ziane, N. Zougab and S. Adjabi, *BirnbaumâĂŞSaunders power-exponential kernel density estimation and Bayes local bandwidth selection for nonnegative heavy-tailed data*, Comput. Statist., **33**(1) (2018), 299–318. DOI

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