Asia Pac. J. Math. 2025 **12**:106

GLOBAL BAYESIAN APPROACH IN THE KERNEL ESTIMATION OF CONDITIONAL DENSITY

NOUR EL-HAYET LADAOURI[1,*], MOULOUD CHERFAOUI[2]

[1]Laboratory of Applied Mathematics, University of Biskra, Biskra, Algeria
[2]Research unit LaMOS (System Modeling and Optimization), University of Béjaïa, Béjaïa, Algeria
[*]Corresponding author: nourelhayet.ladaouri@univ-biskra.dz

Received Oct. 8, 2025

Abstract. The aim of this paper is to apply the Bayesian alternative to calculate the global optimal smoothing parameter in order to estimate the conditional density, using the kernel technique, defined under the hypothesis H_1: "*the smoothing parameters in the* X *and in the* Y *direction are independent*", based upon the likelihood cross-validation technique. Secondly, is to compare its performance (the average of the ISE and the calculation time) with the cross-validation method and the estimator defined under the hypothesis H_2: "the smoothing parameter in the X direction is the same as the Y direction". To achieve our objective, we will realize a numerical study based on simulated samples from two different conditional models, using the Gaussian kernel.

2020 Mathematics Subject Classification. 62G05; 62F15.

Key words and phrases. kernel density estimation; conditional density; UCV method; Bayesian alternative; *MCMC* methods; simulation.

1. Introduction

Let (X, Y) be a couple of random variables in $\mathbb{R} \times \mathbb{R}$, with unknown joint density $g(x, y)$ and marginal density $m(x)$, we consider $(x_1, y_1), ..., (x_n, y_n)$ a sample of independent observations from the same distribution as (X, Y). The kernel estimator of the conditional density of Y knowing $X = x$ was first considered by Rosenblatt [21] in 1969, the author proposed the following estimator:

$$\hat{f}(y/x) = \frac{\frac{1}{nab}\sum\limits_{j=1}^{n} K\left(\frac{x-x_j}{a}\right)K\left(\frac{y-y_j}{b}\right)}{\frac{1}{na}\sum\limits_{j=1}^{n} K\left(\frac{x-x_j}{a}\right)} = \frac{1}{b}\sum_{j=1}^{n} W_{j,a}(x)K\left(\frac{y-y_j}{b}\right), \quad (1)$$

with,

DOI: 10.28924/APJM/12-106

$$W_{j,a}(x) = \frac{K\left(\frac{x-x_j}{a}\right)}{\sum\limits_{j=1}^{n} K\left(\frac{x-x_j}{a}\right)}, \tag{2}$$

where, the function K is a kernel on $\mathbb{R}$ and the couple (a, b) represent a positive parameter controls the degree of smoothing applied to the density estimate. More precisely, a controls the smoothness between conditional densities in the X direction, and b controls the smoothness of each conditional density in the Y direction.

To ensure the convergence of this estimator, Hyndman et al. [13] showed that the two smoothing parameters a and b must check the following conditions: $(a, b) \to (0, 0)$ and $nab \to \infty$ when $n \to \infty$.

By a simple research in the literature, we can find another simplified version of (1) proposed under the hypothesis that the two smoothing parameters in the X and in the Y direction are equal ($h = a = b$). So, the expression (1) can be rewritten as:

$$\hat{f}(y/x) = \frac{\frac{1}{nh^2}\sum\limits_{j=1}^{n} K\left(\frac{x-x_j}{h}\right) K\left(\frac{y-y_j}{h}\right)}{\frac{1}{nh}\sum\limits_{j=1}^{n} K\left(\frac{x-x_j}{h}\right)} = \frac{1}{h}\sum_{j=1}^{n} W_{j,h}(x) K\left(\frac{y-y_j}{h}\right), \tag{3}$$

where $W_{j,h}(x)$ is given by (2), the function K is a kernel on $\mathbb{R}$, and $h > 0$ is a smoothing parameter. For this new version, Youndje [27] in 1996, found the necessary conditions on the smoothing parameter h, so that the estimator $\hat{f}(y/x)$ defined in (3) is punctually consistent when n goes to ∞, $h \to 0$ and $nh^2 \to \infty$. The author also showed that if these conditions on the smoothing parameter h are satisfied, then the estimator $\hat{f}(y/x)$ is close to $f(y/x)$ uniformly, in L^1 and converges almost surely in L^2.

From the two expressions (1) and (3), it is clear that the implementation of the estimators is based on the prior fixing of their parameters (the kernel and the smoothing parameter).

The authors working on the kernel estimation of the conditional density emphasize that the quality of the estimates in question is not significantly affected by changing the kernel. So, the popular choice of the kernel function remains the same as the classical kernel density estimator (see Silverman [24]). However, the choice of the smoothing parameter is crucial, and any change in its value leads to a significant change in the numerical and/or graphical quality of the estimator, where low values of this parameter, compared to the optimal smoothing parameter, generate the phenomenon of under-smoothing and large values of this parameter generate the phenomenon of over-smoothing. In the literature there are two categories to select this parameter, the first is the plug-in method which based on the minimization of the mean integrate squared error ($MISE$), this method is interesting in theory, but practically it depends on unknown quantities and more than this, it is only valid asymptotically. The second category is of the cross-validation type, this technique is interesting in practice because it is guided only by observations. Unfortunately, this technique suffers from the phenomenon of several local minimum and the lack of robustness in relation to changes in sample size.

Both of these techniques tend to provide under- or over-smoothed estimators when the sample is small or medium sized, or when we want to estimate complex functions. So, recently, the Bayesian approach has been proposed as an alternative to these classical methods (see [2,19,28,30]).

The Bayesian approach in the kernel density estimation framework is relatively recent, first of all it only concerns probability densities with unbounded support in univariate and multivariate cases, using the classical Gaussian kernel: [2,28]. Immediately after these works, this alternative used in different field: incomplete data [17, 18], multivariate framework [5, 25, 29], regression function [31, 33, 34], discrete functions [32], semiparametric framework [14] and in conditional density [19].

Similar to the work of Ladouri and Cherfaoui [19], which proposed to use the Bayesian technique to estimate the smoothing parameter associated with the estimator (3), in the present work, we aim to apply the Bayesian method to compute the global optimal smoothing parameter to estimate the conditional density (1) defined under the hypothesis H_1, and to compare its performance with the cross-validation method and with the results obtained under the hypothesis H_2 which presented by Ladaouri and Cherfaoui in [19].

This work is organized as follow: In the next section, we will briefly discus the kernel choice and the classical method to select the smoothing parameter (the cross-validation method). In the section 3, we will cover the main steps to apply the Bayesian technique to select the smoothing parameter. In the last section, we will describe our numerical application and present the numerical and graphical results obtained.

2. Choice of kernel and smoothing parameter

As we underline in the introduction, the change in the kernel has little effect on the quality of the estimator, so to simplify the choice of this function, whether for the estimator (1) or (3), it suffices that it be defined as in the case of the univariate density estimation. Thus, the choice must only be adapted to the support of the target density to be estimated (see: [3,4,15,16,24]).

Unlike the choice of the kernel, the selection of the smoothing parameter plays a crucial role in determining the quality of the estimator in question and the appropriate choice of this parameter is one of the main challenges in non-parametric density estimation. In the literature there are several smoothing parameter selection methods, the most common technique used in application is the **Cross-validation technique**.

Habbema *et al.* [12] in 1974 were the first to apply the cross-validation method to estimate the smoothing parameter for univariate kernel density estimation, later several other works were proposed in this framework [1,22,23]. In conditional density estimation there are various work proposed to apply the cross-validation technique in the estimation of the smoothing parameter [6,9,10].

Fan *et al.* [6] in 2004 proposed to select the couple of smoothing parameter (a, b) in the estimator (1) by the cross-validation technique, the authors have shown that if the integrated squared error (ISE) is retained as a performance criterion for the construction of the estimator (1), then the estimators of the optimal smoothing parameters a and b are those which jointly minimize the following criterion:

$$CV(a,b) = \hat{A} - 2\hat{B}, \tag{4}$$

where,

$$\begin{aligned}
\hat{A} &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} \hat{f}_{-i}\left(y/x_i\right)^2 dy, \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum\limits_{j\neq i}^{n}\sum\limits_{k\neq i} K_a\left(x_i - x_j\right) K_a\left(x_i - x_k\right)\int_{\mathbb{R}} K_b\left(y - y_j\right) K_b\left(y - y_k\right) dy}{\sum\limits_{j\neq i}^{n} K_a\left(x_i - x_j\right)}\right].
\end{aligned}$$

$$\begin{aligned}
\hat{B} &= \frac{1}{n}\sum_{i=1}^{n}\hat{f}_{-i}\left(y_i/x_i\right), \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum\limits_{j\neq i}^{n} K_a\left(x_i - x_j\right) K_b\left(y_i - y_j\right)}{\sum\limits_{j\neq i}^{n} K_a\left(x_i - x_j\right)}\right],
\end{aligned}$$

with $\hat{f}_{-i}$ represents the leave-on-out cross-validation estimator of the conditional density f computed from the set of points except the point (x_i, y_i).

In the case of the second estimator (3) Youndjé in [26] proposed to use the cross-validation technique to estimate the smoothing parameter associated with the estimator (3), the author showed that if we choice the ISE criterion to estimate the parameter h, so the optimal smoothing parameter using the cross-validation technique is the value h that minimizes the following expression:

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}}\left(\hat{f}_{-i}\left(y/X_i\right)\right)^2 W'(y)dyW(X_i) - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}\left(Y_i/X_i\right)W(X_i)W'(Y_i), \tag{5}$$

where, W and W' are beforehand fixed positive weight functions.

3. Principle of the global Bayesian approach

The multiplicity of the procedures used to select the smoothing parameter justifies that they are incomplete and suffer from many problems. To overcome this problem, recently the Bayesian approach has been proposed as an alternative to these classical methods.

At variance with classical methods which are often based on directly minimizing or maximizing a criterion, the Bayesian formalism is characterized by treating the smoothing parameter as a random variable follows a prior probability π, this distribution serves to compensate the lack of information when the sample size is small or medium.

If we choose the ISE as an error criterion, then the value of the smoothing parameter estimator using the Bayesian approach is equal to the posterior mean, our objective is to calculate this quantity. According to the complexity of the posterior, the Bayesian approach often leads to integration problems, practically there are several situations that is impossible to compute the posterior mean, so, it is impossible to directly obtain the Bayesian estimator of the smoothing parameter. To overcome the calculation problem of the posterior law we used the Monte Carlo Markov Chain (*MCMC*) method, exactly the random-walk of Metropolis-Hasting (*M-H*) algorithm, which proposed by Metropolis *et al.* [20] and generalized by Hastings [11].

Suppose that $x = (x_1, x_2, ...x_n)$ is the observation vector and $\theta^{(0)}$ is the starting value of the smoothing parameter to be estimated and $\theta^{(k)}$ is the candidate generated at the k^{th} step. The *M-H* algorithm is given as follows:

1 **Step1:** Initialize $\theta^{(0)}$.

2 **Step2:** For $k = 1, ..., N$

  a) Generate $\widetilde{\theta} \rightsquigarrow q(\widetilde{\theta}/\theta^{(k-1)})$.

  b) Calculate the acceptance probability $\rho = \min\left\{1, \frac{\pi(\widetilde{\theta}/x,y)q(\theta^{(k-1)}/\widetilde{\theta})}{\pi(\theta^{(k-1)}/x,y)q(\widetilde{\theta}/\theta^{(k-1)})}\right\}$.

  c) $\theta^{(k)} = \begin{cases} \widetilde{\theta} & \text{if } u < \rho \quad \text{with } \ u \sim U[0,1]. \\ \theta^{(k-1)} & \text{otherwise.} \end{cases}$

3 **Step3:** Set $k = k + 1$ and go to **Step2**.

4 **Step4:** Calculate $\hat{\theta} = \frac{1}{N-N_0} \sum\limits_{k=N_0+1}^{N} \theta^{(k)}$.

4. Global Bayesian approach in conditional density estimation

Similar to the work of Ladaouri and Cherfaoui [19] who applied the global Bayesian method to estimate the smoothing parameter associated with the estimator (3), in our work we will extend this idea to estimate the smoothing parameter associated with the estimator (1). Indeed, for comparative purposes we will follow the same steps and keep the same choice of parameters as [19]. To do this we have four steps to realize:

Step1. Estimator of the likelihood function:

In the Bayesian framework the likelihood function have crucial rule for calculate the posterior distribution. Let $(x_i, yi), i = 1, ..., n$ be i.i.d. bivariate observations, the likelihood approach function of the conditional density, under the cross-validation technique, is given by:

$$\hat{L}(a,b,Y/X)=\prod_{i=1}^{n}\hat{f}_{-i}(y_i/x_i)=\frac{1}{b^n}\prod_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}W_{j,a}(x_i)K\left(\frac{y_i-y_j}{b}\right).$$

Where $W_{j,a}(x)$ is given by (2).

Step2. Choice of the prior function:

The choice of the prior distribution is the most questionable and critical point in Bayesian analysis, where this function compensates for the lack of information when the sample comes in small or medium size. Note that the choice is not alone here, because practically the Bayesian estimator is not very sensitive to the choice of prior. Therefore, we must choose a prior that allows the *MCMC* methods to work properly.

In our case we consider that, the smoothing couple parameter (a,b) follows a Gamma bi-distribution with positive parameters (α_1,α_2), and (β_1,β_2). then, under the hypothesis that a and b are statistically independent, the prior is given by:

$$\begin{aligned}\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2)&=\frac{\beta_1^{\alpha_1}\beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\exp(-a\beta_1-b\beta_2)a^{\alpha_1-1}b^{\alpha_2-1},\\ &\propto a^{\alpha_1-1}b^{\alpha_2-1}\exp(-a\beta_1+b\beta_2),\end{aligned}$$

where $\propto$ denotes proportional.

Step3. Calculation of the posterior function:

Using the continuous version of Bayes theorem, the posterior density can be obtained via the likelihood cross-validation and the prior functions. Indeed, the joint posterior function of (a,b) is given by:

$$\begin{aligned}\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2,X,Y)&=\frac{L(a,b,Y/X)\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2)}{\int\int L(a,b,Y/X)\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2)dadb},\\ &\approx\frac{\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2)\prod_{i=1}^{n}\hat{f}_{-i}(y_i/x_i)}{\int\int \hat{L}(a,b,Y/X)\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2)dadb}.\end{aligned}$$

We note that the denominator does not depend on any components of the couple (a,b), so we can consider them as a normalizing constant for the posterior function. So, the posterior function becomes:

$$\begin{aligned}\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2,X,Y)&\propto\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2)\prod_{i=1}^{n}\hat{f}_{-i}(y_i/x_i),\\ &\propto\frac{\pi(a,b/\alpha_1,\beta_1,\alpha_2,\beta_2)}{b^n}\prod_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}W_{j,a}(x_i)K\left(\frac{y_i-y_j}{b}\right),\\ &\propto a^{\alpha_1-1}b^{\alpha_2-n-1}e^{(-a\beta_1+b\beta_2)}\prod_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}W_{j,a}(x_i)K\left(\frac{y_i-y_j}{b}\right).\end{aligned}\tag{6}$$

Step4. Global Bayesian estimator of the smoothing parameters:

The Bayesian estimator of (a, b) is the mean of the previous posterior (6), because the complexity of this expression, we propose to use the *MCMC* methods and exactly the M-H algorithm.

The candidate (a, b) generates from an instrumental density $q(a, b/(a^{(k)}, b^{(k)}), \gamma_1(k), \gamma_2(k))$ knowing that the two quantities $a^{(k)}$ and $b^{(k)}$ represent the candidates generates at the k^{th} step. The quantities γ_1 and γ_2 are adjustment parameters chosen so as to obtain an optimal acceptance rate, noted τ, its value is close to 0.5 (Gelman *et al.* [8]).

We propose to replace the instrumental function by a joint Gaussian distribution truncated on $[0; +\infty[\times[0; +\infty[$. Therefore:

$$q(a, b/(a^{(k)}, b^{(k)}), \gamma_1(k), \gamma_2(k)) = \frac{\phi\left(\frac{b^{(k)}}{\gamma_2(k)}\right)\phi\left(\frac{b^{(k)}}{\gamma_2(k)}\right)}{\gamma_1(k)\gamma_2(k)\sqrt{\phi\left(\frac{a^{(k)}}{\gamma_1(k)}\right)\phi\left(\frac{a-a^{(k)}}{\gamma_1(k)}\right)}},$$

where $\phi(.)$ is a standard normal distribution function.

In order to reduce the bias of the Bayesian estimator which is due to the effect of the initial values, the calculation starts after a certain rank of iteration N_0, where the number of iterations before N_0 is ignored, we call N_0 "the burn-in period". So, the smoothing parameter is the average of the values which we obtained after the iteration N_0. So, the estimator of the smoothing parameter a and b given as follows:

$$\hat{a} = \frac{1}{N - N_0}\sum_{k=N_0}^{N} a^{(k)}, \ \ and \ \ \hat{b} = \frac{1}{N - N_0}\sum_{k=N_0}^{N} b^{(k)}.$$

The updates formula for adjustment parameters are given by(Garthwaite *et al.* [7]):

$$\gamma_1(k+1) = \begin{cases} \gamma_1(k) + \frac{\gamma_1(k)}{(k)\tau} & \text{if } \widetilde{a} \text{ is accepted.} \\ \gamma_1(k) - \frac{\gamma_1(k)}{(k)(1-\tau)} & \text{if } \widetilde{a} \text{ is rejected.} \end{cases}$$

$$\gamma_2(k+1) = \begin{cases} \gamma_2(k) + \frac{\gamma_2(k)}{(k)\tau} & \text{if } \widetilde{b} \text{ is accepted.} \\ \gamma_2(k) - \frac{\gamma_2(k)}{(k)(1-\tau)} & \text{if } \widetilde{b} \text{ is rejected.} \end{cases}$$

5. NUMERICAL APPLICATION:

In this section, we have carried out a comparative numerical application based on simulated samples of different sizes, using the Gaussian kernel, with the aim of, on the one hand, we focused on using the global Bayesian approach to estimate the smoothing parameter in the construction of the estimator (1), in order to identify the impact of Bayesian inference.

On the other hand, we took the global Bayesian approach in comparison with the Cross-Validation method in the estimation of the kernel conditional density using the version (1), which has been proposed under the hypothesis that the two smoothing parameters a and b are different $(h = a \neq b)$. More precisely, we analyze the impact of each of these smoothing parameter selection methods on the performance of the estimator in question, in the sense of the average of their ISE and the average time calculation required for their implementation. We also highlighted the effect of the two hypotheses "H_1" and "H_2" on the performances (the average of the ISE and the calculation time) of the estimators in question by comparing the results obtained in this work with those obtained by Ladaouri and Cherfaoui in [19].

For a simplification reason, we use the following notations in the rest of this document:

- (a_{Bays}, b_{Bays}): The smoothing parameter obtained by the Bayesian alternative associated with the first estimator (1).
- h_{Bays}: The smoothing parameter obtained by the Bayesian alternative associated with the second estimator (3).
- (a_{UCV}, b_{UCV}): The smoothing parameter obtained by the unbiased cross-validation method associated with the first estimator (1).
- h_{UCV}: The smoothing parameter obtained by the unbiased cross-validation method associated with the second estimator (3).

5.1. **Description of application parameters:** For answer to our problem, we implement a simulator under the programming language Matlab. To realize this simulator, and for calculation reasons, we propose to discretize the ISE, and approximate it by the following criteria:

$$AISE = \frac{\Delta}{n}\sum_{j=1}^{N}\sum_{i=1}^{n}\left[\hat{f}(y_j^{'}/X_i) - f(y_j^{'}/X_i)\right]^2, \tag{7}$$

where $(X_i, Y_i), i = 1, ..., n$ an i.i.d sample from the joint density of (X, Y) and $y^{'} = \left(y_1^{'}, y_2^{'}, ..., y_J^{'}\right)$ is a vector of equidistant points in the space of Y and $\Delta = y_{j+1}^{'} - y_j^{'}, \forall j \in \{1, 2, ..., J-1\}$.

So, the optimal smoothing parameters in the sense of the average ISE are the quantities which minimizing the following expression:

$$\overline{ISE} = \frac{1}{m}\sum_{l=1}^{m} AISE(a, b, X^{(l)}, Y^{(l)}, y', f), \tag{8}$$

where $(X^{(l)}, Y^{(l)}), l = 1, ..., m$ is an i.i.d sample from the density $g(X, Y)$.

For the numerical application, we took the two examples presented by Ladaouri and Cherfaoui [19], namely:

Model 1: We considered the following model:

$$Y_i = 1 + \frac{1}{2}X_i + \epsilon_i,$$

where: X_i and ϵ_i are two independent random variables, from normal distribution: $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, 0.2)$ respectively.

Model 2: Let:

$$Y = 2\sin(\pi X) + \epsilon,$$

with X and ϵ are two random variables, where X follows the uniform distribution on $[0, 2]$, and $\epsilon/X = WU + (1 - W)V$ with W is an equiprobable binary random variable $(P(W = 0) = P(W = 1) = 0.5)$, U is a random variable from the distribution $\mathcal{N}(X, 0.3)$ and V is a random variable that follows $\mathcal{N}(0, 0.3)$.

In the rest of the application, we propose to fix the parameters of simulation as follows:

(1) The Gaussian kernel for the construction of estimator in question.
(2) The discretization of $Y : y'$ varied between -20 and 20 with a step 0.1.
(3) The sample size: $n \in \{5, 10, 15, 20, 25, 50, 100, 150, 200, 250, 500, 1000\}$
(4) 30 repeat experiments were conducted for each sample size.
(5) The results for the *MCMC* method are obtained for 15 chains, for each chain we simulate 10000 iterations and we assumed that the burn-in period is 1000.

5.2. **Numerical and graphical results:** The results obtained in our application are arranged in Table 1 and Table 2, and presented in the Figures 1, 2, 3 and 4. For a comparative reason, we will present our results with those obtained in the work of Ladaouri and Cherfaoui [19] in figures 5 and 6.

	UCV			*Bayes*		
n	(a_{ucv}, b_{ucv})	$\overline{ISE}$	*time* (mn)	(a_{Bay}, b_{Bay})	$\overline{ISE}$	*time* (mn)
5	(0.3580, 0.3735)	0.7787	0.0114	(0.1904, 0.3973)	0.5180	1.3675
10	(0.3502, 0.2438)	0.5837	0.0234	(0.1871, 0.3074)	0.3750	1.7267
15	(0.2404, 0.2017)	0.9038	0.0371	(0.1739, 0.2666)	0.3023	2.0911
20	(0.2959, 0.1684)	0.6427	0.0502	(0.1643, 0.2364)	0.3029	2.4725
25	(0.2674, 0.1529)	0.7828	0.0596	(0.1640, 0.2179)	0.2919	2.8457
50	(0.2568, 0.1425)	0.4610	0.1184	(0.1675, 0.1979)	0.2242	4.9221
100	(0.1794, 0.1343)	0.3923	0.2901	(0.1548, 0.1522)	0.1524	9.5711
150	(0.2363, 0.1103)	0.3080	0.4511	(0.1665, 0.1342)	0.1261	15.2139
200	(0.1923, 0.1119)	0.2613	0.6688	(0.1534, 0.1301)	0.1082	21.5722
250	(0.2219, 0.1014)	0.2935	0.8443	(0.1570, 0.1262)	0.0978	28.9978
500	(0.1584, 0.1030)	0.0965	2.7944	(0.1424, 0.1029)	0.0650	84.3428
1000	(0.1587, 0.0794)	0.1501	7.0323	(0.1280, 0.0935)	0.0472	221.7033

Table 1: Variation of $\overline{ISE}$ and the mean calculation time according to the sample size n, case of **Model 1**.

	UCV			$Bayes$		
n	(a_{ucv}, b_{ucv})	$\overline{ISE}$	$time$ (mn)	(a_{Bay}, b_{Bay})	$\overline{ISE}$	$time$ (mn)
5	(3.5445, 1.1333)	0.3691	0.0128	(0.1740, 0.8918)	0.2619	1.3197
10	(1.8877, 0.7784)	0.3165	0.0253	(0.1609, 0.7639)	0.2551	1.6756
15	(1.4416, 0.6034)	0.4167	0.0389	(0.1449, 0.6909)	0.2504	2.0419
20	(0.1078, 0.5102)	0.2748	0.0470	(0.1346, 0.5670)	0.2230	2.4154
25	(0.1033, 0.5317)	0.2620	0.0599	(0.1122, 0.5738)	0.2056	2.7925
50	(0.0641, 0.4433)	0.1737	0.1314	(0.0773, 0.4116)	0.1474	4.8780
100	(0.0542, 0.3285)	0.1260	0.3129	(0.0551, 0.3184)	0.1173	9.6825
150	(0.0493, 0.2853)	0.1009	0.5199	(0.0482, 0.2606)	0.0984	15.1243
200	(0.0440, 0.2569)	0.0895	0.7690	(0.0441, 0.2457)	0.0846	21.5179
250	(0.0422, 0.2520)	0.0807	0.9974	(0.0417, 0.2347)	0.0775	28.7020
500	(0.0357, 0.2139)	0.0532	3.0828	(0.0337, 0.1943)	0.0529	86.7742

Table 2: Variation of $\overline{ISE}$ and the mean calculation time according to the sample size n, case of **Model 2**.

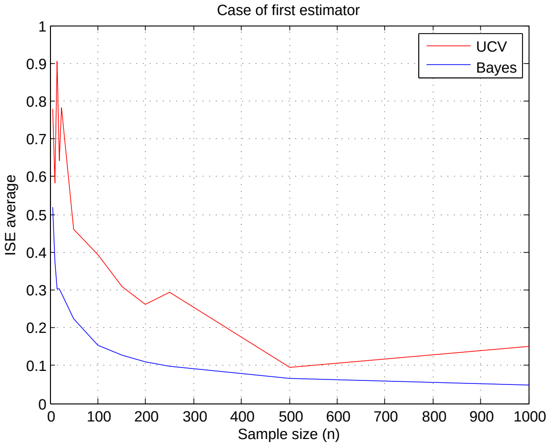

FIGURE 1. Variation of $\overline{ISE}$ according to the sample size n, case of **Model 1**.

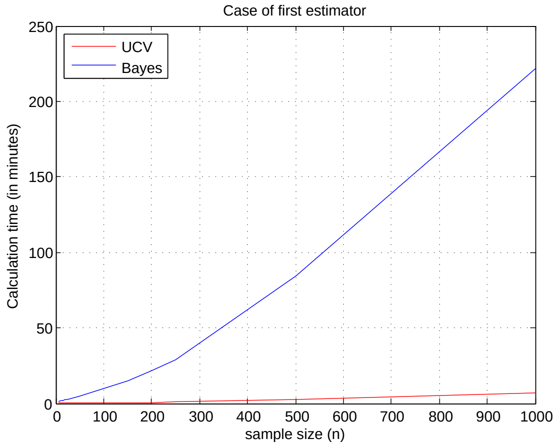

FIGURE 2. Variation of calculation Time according to the sample size n, case of **Model 1**.

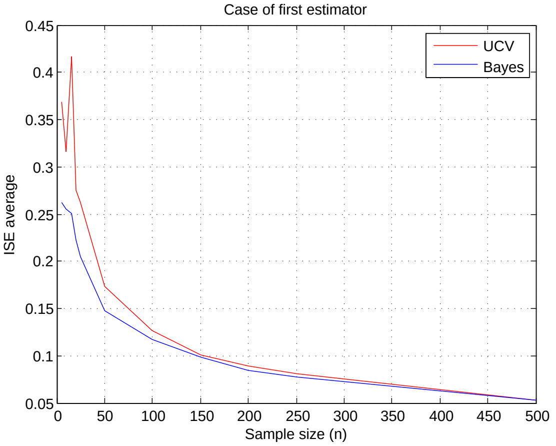

FIGURE 3. Variation of $\overline{ISE}$ according to the sample size n, case of **Model 2**.

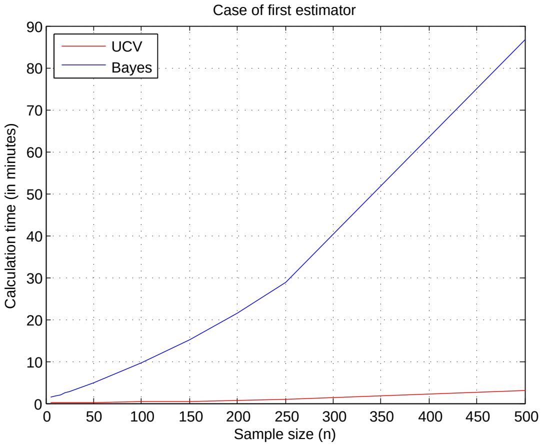

FIGURE 4. Variation of calculation Time according to the sample size n, case of **Model 2**.

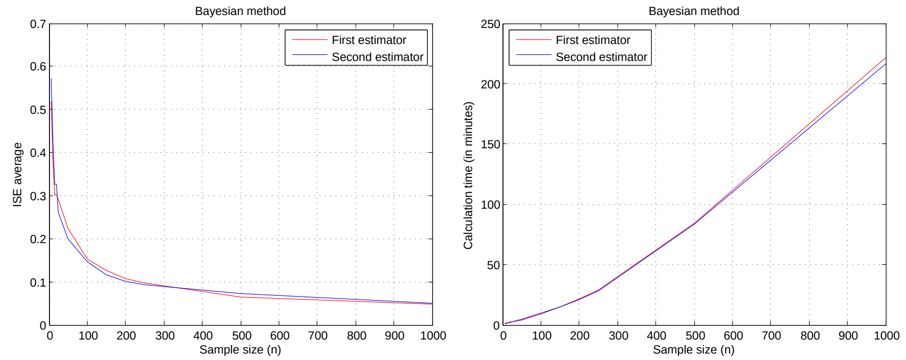

Figure 5. Variation of $\overline{ISE}$ and calculation Time according to the sample size n, using Bayesian method case of **Model 1**.

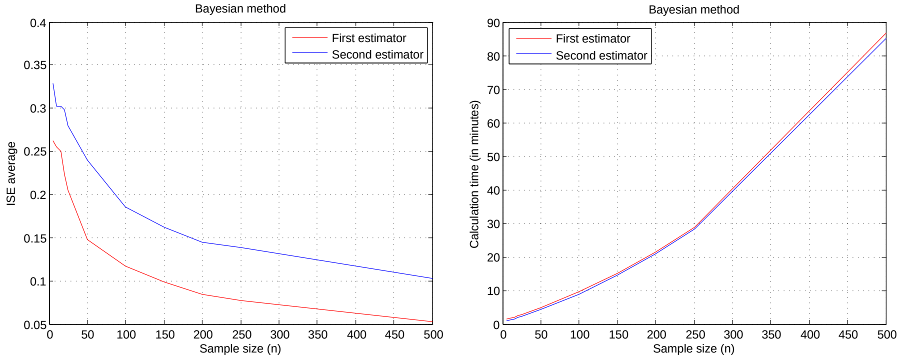

Figure 6. Variation of $\overline{ISE}$ and calculation Time according to the sample size n, using Bayesian method case of **Model 2**.

5.3. **Discussion of the results.** Taking into account the numerical and graphical results obtained in the present paper and compared with those in the paper [19], we note:

- In the two methods (Bayesian and UCV) and in the case of the two models, the estimator (1) converges according to the sample size n.
- The Bayesian approach is better than classical methods in the sense of the asymptotic mean integrated square error, this applies to both proposed models.
- In some cases, for a median and large sample size, the Bayesian approach and the UCV method give us estimators with practically the same ISE (see figure 3).
- According to the graphs, we note that in some cases and especially in median and large sample size situations, the investment of calculation time to improve the quality of the estimate is not interesting, because the contribution of an additional calculation time in the quality of the estimator is very minimal. Indeed, for example, in the case of model 2, first estimator, when the size of the sample $n = 500$, we clearly observe in Figure 3 that in order to obtain a precision only of order 10^{-3} in the sense of the ISE for the Bayesian estimator compared to the UCV estimator, we note that the calculation time requires less than two minutes in the UCV method,

while the use of the Bayesian approach requires approximately an hour and a half (see Table 2).

- For both models, the calculation time consumed in the application of the estimator (1) is more considerable in the case of the Bayesian estimator than in the UCV estimator, the difference is very large between the two methods (see figure 2 and 4).
- The quality of the estimates in the sense of the ISE depends on the model and the estimator considered. Indeed, in the case of the first model, the choice of estimator's versions (1) and (3) does not seem to have great importance, and this fact that the two versions provide us with estimators having practically the same ISE (Figure 5). In the second model, the most efficient estimators in the sense of the average ISE, are obtained when we consider the estimator defined in (3) and this whatever the size of the sample (see Figure 6).
- For both models and whatever method used in the calculation of the estimator, the calculation time is more considerable in the case of the first estimator (1) than the second one (3), but this difference is not large in the two models, almost the two estimators have the same calculation time.

6. Conclusion

The aim of the present paper is to extend the idea of the Bayes rule to compute the global optimal smoothing parameter in the framework of the kernel conditional density estimation, and to analyze the performance of this estimator as a function of the sample size and the hypothesis imposed on the smoothing parameters (The smoothing parameters in the X direction and the Y direction are independent and the opposite case $(a \neq b)$), and to compare their performance with the cross-validation method.

The results obtained, through this work and comparing them with those obtained in the work Ladaouri and Cherfaoui [19], indicate that, for the small and medium sample size, the choice of the global smoothing parameter using the Bayesian approach is most efficient than the use of classical method and it is better to impose the independence between the two smoothing parameters than the equality between them, in the sense of the average ISE, but this efficient decreases as the sample size increases. However, if we are interested in reducing the calculation time it is better to use the UCV method and to impose the equality of the two smoothing parameters.

Authors' Contributions. All authors have read and approved the final version of the manuscript. The authors contributed equally to this work.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] A.W. Bowman, An Alternative Method of Cross-Validation for the Smoothing of Density Estimates, Biometrika 71 (1984), 353. https://doi.org/10.2307/2336252.

[2] M.J. Brewer, A Modelling Approach for Bandwidth Selection in Kernel Density Estimation, in: COMPSTAT, Physica-Verlag HD, Heidelberg, 1998, pp. 203-208. https://doi.org/10.1007/978-3-662-01131-7_22.

[3] S.X. Chen, Beta Kernel Estimators for Density Functions, Comput. Stat. Data Anal. 31 (1999), 131–145. https://doi.org/10.1016/s0167-9473(99)00010-9.

[4] S.X. Chen, Probability Density Function Estimation Using Gamma Kernels, Ann. Inst. Stat. Math. 52 (2000), 471–480. https://doi.org/10.1023/a:1004165218295.

[5] M.S. de Lima, G.S. Atuncar, A Bayesian Method to Estimate the Optimal Bandwidth for Multivariate Kernel Estimator, J. Nonparametric Stat. 23 (2011), 137–148. https://doi.org/10.1080/10485252.2010.485200.

[6] J. Fan, A Crossvalidation Method for Estimating Conditional Densities, Biometrika 91 (2004), 819–834. https://doi.org/10.1093/biomet/91.4.819.

[7] P.H. Garthwaite, Y. Fan, S.A. Sisson, Adaptive Optimal Scaling of Metropolis–Hastings Algorithms Using the Robbins–Monro Process, Commun. Stat. - Theory Methods 45 (2016), 5098–5111. https://doi.org/10.1080/03610926.2014.936562.

[8] A. Gelman, G.O. Roberts, W.R. Gilks, Efficient Metropolis Jumping Rules, Bayesian Stat. 5 (1996), 599–608. https://doi.org/10.1093/oso/9780198523567.003.0038.

[9] P. Hall, J. Racine, Q. Li, Cross-Validation and the Estimation of Conditional Probability Densities, J. Am. Stat. Assoc. 99 (2004), 1015–1026. https://doi.org/10.1198/016214504000000548.

[10] B.E. Hansen, Nonparametric Conditional Density Estimation, Technical Report, University of Wisconsin, (2004).

[11] W.K. Hastings, Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika 57 (1970), 97. https://doi.org/10.2307/2334940.

[12] J.D.F. Habbema, J. Hermans, K. van den Broek, A Stepwise Discriminant Analysis Program Using Density Estimation, in: Compstat: Proceedings in Computational Statistics, pp. 101–110, (1974). https://api.semanticscholar.org/CorpusID:123794032.

[13] R.J. Hyndman, D.M. Bashtannyk, G.K. Grunwald, Estimating and Visualizing Conditional Densities, J. Comput. Graph. Stat. 5 (1996), 315–336. https://doi.org/10.1080/10618600.1996.10474715.

[14] T.S. Kiesse, N. Zougab, C.C. Kokonendji, Bayesian Estimation of Bandwidth in Semiparametric Kernel Estimation of Unknown Probability Mass and Regression Functions of Count Data, Comput. Stat. 31 (2015), 189–206. https://doi.org/10.1007/s00180-015-0627-1.

[15] C.C. Kokonendji, T.S. Kiesse, Discrete Associated Kernels Method and Extensions, Stat. Methodol. 8 (2011), 497–516. https://doi.org/10.1016/j.stamet.2011.07.002.

[16] C.C. Kokonendji, T. Senga Kiesse, S.S. Zocchi, Discrete Triangular Distributions and Non-Parametric Estimation for Probability Mass Function, J. Nonparametric Stat. 19 (2007), 241–254. https://doi.org/10.1080/10485250701733747.

[17] K.B. Kulasekera, W.J. Padgett, Bayes Bandwidth Selection in Kernel Density Estimation with Censored Data, J. Nonparametric Stat. 18 (2006), 129–143. https://doi.org/10.1080/10485250600556744.

[18] C. Kuruwita, K. Kulasekera, W. Padgett, Density Estimation Using Asymmetric Kernels and Bayes Bandwidths with Censored Data, J. Stat. Plan. Inference 140 (2010), 1765–1774. https://doi.org/10.1016/j.jspi.2010.01.001.

[19] L.N. El-hayet, M. Cherfaoui, Bayesian Approach in the Nonparametric Conditional Density Estimation, Stud. Eng. EXACT Sci. 5 (2024), e11284. https://doi.org/10.54021/seesv5n2-592.

[20] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of State Calculations by Fast Computing Machines, J. Chem. Phys. 21 (1953), 1087–1092. https://doi.org/10.1063/1.1699114.

[21] M. Rosenblatt, Conditional Probability Density and Regression Estimators, in: P.R. Shnaiah (Ed.), Multivariate Analysis II, Academic Press, New York, pp. 25–31, (1969).

[22] M. Rudemo, Empirical Choice of Histogram and Kernel Density Estimators, Scand. J. Stat. 9 (1982), 65–78. https://www.jstor.org/stable/4615859.

[23] D.W. Scott, G.R. Terrell, Biased and Unbiased Cross-Validation in Density Estimation, J. Am. Stat. Assoc. 82 (1987), 1131–1146. https://doi.org/10.1080/01621459.1987.10478550.

[24] B. Silverman, Density Estimation for Statistics and Data Analysis, Routledge, 1986. https://doi.org/10.1007/978-1-4899-3324-9.

[25] S.M. Somé, C.C. Kokonendji, F.G.B.L. Dobélé-Kpoka, An Effective Estimation of Multivariate Density Functions Using Extended-Beta Kernels with Bayesian Adaptive Bandwidths, arXiv:2502.05366 (2025). https://doi.org/10.48550/arxiv.2502.05366.

[26] E. Youndjé, Validation Croisee pour l'Estimation Non-Parametrique de la Densite Conditionnelle, Ann. ISUP, 38 (1994), 57–80. https://hal.science/hal-03659839v1.

[27] E. Youndjé, Propriètés de Convergence de l'Estimateur à Noyau de la Densité Conditionnelle, Rev. Roum. Math. Pures Appl. 41 (1996), 535–566.

[28] X. Zhang, M.L. King, R.J. Hyndman, A Bayesian Approach to Bandwidth Selection for Multivariate Kernel Density Estimation, Comput. Stat. Data Anal. 50 (2006), 3009–3031. https://doi.org/10.1016/j.csda.2005.06.019.

[29] X. Zhang, R.D. Brooks, M.L. King, A Bayesian Approach to Bandwidth Selection for Multivariate Kernel Regression with an Application to State-Price Density Estimation, J. Econ. 153 (2009), 21–32. https://doi.org/10.1016/j.jeconom.2009.04.004.

[30] N. Zougab, S. Adjabi, C.C. Kokonendji, Binomial Kernel and Bayes Local Bandwidth in Discrete Function Estimation, J. Nonparametric Stat. 24 (2012), 783–795. https://doi.org/10.1080/10485252.2012.678847.

[31] N. Zougab, S. Adjabi, C.C. Kokonendji, A Bayesian Approach to Bandwidth Selection in Univariate Associate Kernel Estimation, J. Stat. Theory Pract. 7 (2013), 8–23. https://doi.org/10.1080/15598608.2013.756286.

[32] N. Zougab, S. Adjabi, C.C. Kokonendji, Adaptive Smoothing in Associated Kernel Discrete Functions Estimation Using Bayesian Approach, J. Stat. Comput. Simul. 83 (2013), 2219–2231. https://doi.org/10.1080/00949655.2012.686615.

[33] N. Zougab, Approche Bayésienne dans l'Estimation Non Paramétrique de la Densité de Probabilité et la Courbe de régression de la Moyenne, Doctoral Thesis, University of Bejaia, Algerie, (2013).

[34] N. Zougab, S. Adjabi, C.C. Kokonendji, Bayesian Approach in Nonparametric Count Regression with Binomial Kernel, Commun. Stat. - Simul. Comput. 43 (2013), 1052–1063. https://doi.org/10.1080/03610918.2012.725145.