Ake: An R Package for Discrete and Continuous Associated Kernel Estimations

by Wanbitching E. Wansouwé, Sobom M. Somé and Célestin C. Kokonendji

Abstract Kernel estimation is an important technique in exploratory data analysis. Its utility relies on its ease of interpretation, especially based on graphical means. The **Ake** package is introduced for univariate density or probability mass function estimation and also for continuous and discrete regression functions using associated kernel estimators. These associated kernels have been proposed due to their specific features of variables of interest. The package focuses on associated kernel methods appropriate for continuous (bounded, positive) or discrete (count, categorical) data often found in applied settings. Furthermore, optimal bandwidths are selected by cross-validation for any associated kernel and by Bayesian methods for the binomial kernel. Other Bayesian methods for selecting bandwidths with other associated kernels will complete this package in its future versions; particularly, a Bayesian adaptive method for gamma kernel estimation of density functions is developed. Some practical and theoretical aspects of the normalizing constant in both density and probability mass functions estimations are given.

Introduction

Kernel smoothing methods are popular tools for revealing the structure of data that could be missed by parametric methods. For real datasets, we often encounter continuous (bounded, positive) or discrete (count, categorical) data types. The classical kernels methods assume that the underlying distribution is unbounded continuous, which is frequently not the case; see, for example, Duong (2007) for multivariate kernel density estimation and discriminant analysis. A solution is provided for categorical data sets by Hayfield and Racine (2008). In fact, they used kernels well adapted for these categorical sets (Aitchison and Aitken, 1976). Throughout the present paper, the unidimensional support $\mathbb{T}$ of the variable of interest can be $\{0, 1, \ldots, N\}$, $[a, b]$ or $[0, \infty)$ for a given integer N and reals $a < b$.

The recently developed **Ake** package, implements associated kernels that seamlessly deal with continuous (bounded, positive) and discrete (categorical, count) data types often found in applied settings; see, for example, Libengué (2013) and Kokonendji and Senga Kiessé (2011). These associated kernels are used to smooth probability density functions (p.d.f.), probability mass functions (p.m.f.) or regression functions. The coming versions of this package will contain, among others, p.d.f. estimation of heavy tailed data (e.g., Ziane et al., 2015) and the estimation of other functionals. The bandwidth selection remains crucial in associated kernel estimations of p.d.f., p.m.f. or regression functions. Some methods have been investigated for selecting bandwidth parameters but the commonly used is the least squared cross-validation. A Bayesian approach has been also recently introduced by Zougab et al. (2012) in the case of a binomial kernel. This method can be extended to various associated kernels with other functionals. Despite the great number of packages implemented for nonparametric estimation in continuous cases with unbounded kernels, to the best of our knowledge, the R packages to estimate p.m.f. with categorical or count variables, p.d.f. with bounded or positive datasets, and regression functions have been far less investigated.

The rest of the paper is organized as follows. In Section Non-classical associated kernels, we briefly describe the definition of associated kernels and then illustrate examples in both continuous and discrete cases which are discussed. Then, the associated kernel estimator for p.d.f. or p.m.f. is presented and illustrated with some R codes in Section Density or probability mass function estimations. In particular, three bandwidth selection methods are available: cross-validation for any (continuous or discrete) associated kernel, the Bayesian local method for the binomial kernel and also a new theoretical Bayesian adaptive method for the gamma kernel. Also, some practical and theoretical aspects of the normalizing constant in both p.d.f. and p.m.f. estimations are given. Section Bandwidth selection for kernel regression involving associated kernels investigates the case of regression functions with two bandwidth selection techniques: cross-validation and also the Bayesian global method for the binomial kernel. Section Summary and final remarks concludes.

Non-classical associated kernels

Recall that the support $\mathbb{T}$ of the p.m.f., p.d.f. or regression function, to be estimated, is any set $\{0, 1, \ldots, N\}$, $[a, b]$ or $[0, \infty)$ for a given integer N and reals $a < b$. The associated kernel in both continuous and discrete cases is defined as follows.

Definition 34.2.1. *(Kokonendji and Senga Kiessé, 2011; Libengué, 2013) Let $\mathbb{T}$ ($\subseteq \mathbb{R}$) be the support of the p.m.f., p.d.f. or regression function, to be estimated, $x \in \mathbb{T}$ a target and h a bandwidth. A parametrized p.m.f. (respectively p.d.f.) $K_{x,h}(\cdot)$ of support $S_{x,h}$ ($\subseteq \mathbb{R}$) is called "associated kernel" if the following conditions are satisfied:*

$$x \in S_{x,h}, \tag{1}$$
$$\mathbb{E}\left(\mathcal{Z}_{x,h}\right) = x + a(x, h), \tag{2}$$
$$\mathbb{V}\left(\mathcal{Z}_{x,h}\right) = b(x, h), \tag{3}$$

where $\mathcal{Z}_{x,h}$ denotes the random variable with p.m.f. (respectively p.d.f.) $K_{x,h}$ and both $a(x, h)$ and $b(x, h)$ tend to 0 as h goes to 0.

Remark 34.2.2. *This definition has the following interesting interpretations:*

(i) *The function $K_{x,h}(\cdot)$ is not necessary symmetric and is intrinsically linked to x and h.*

(ii) *The support $S_{x,h}$ is not necessary symmetric around x; it can depend or not on x and h.*

(iii) *The condition (1) can be viewed as $\cup_{x \in \mathbb{T}} S_{x,h} \supseteq \mathbb{T}$ and it implies that the associated kernel takes into account the support $\mathbb{T}$ of the density f, to be estimated.*

(iv) *If $\cup_{x \in \mathbb{T}} S_{x,h}$ does not contain $\mathbb{T}$ then this is the well-known problem of boundary bias.*

(v) *Both conditions (2) and (3) indicate that the associated kernel is more and more concentrated around x as h goes to 0. This highlights the peculiarity of the associated kernel which can change its shape according to the target position.*

In order to construct an associated kernel $K_{x,h}(\cdot)$ from a parametric (discrete or continuous) probability distribution K_θ, $\theta \in \Theta \subset \mathbb{R}^d$ on the support S_θ such that $S_\theta \cap \mathbb{T} \neq \emptyset$, we need to establish a correspondence between $(x, h) \in \mathbb{T} \times (0, \infty)$ and $\theta \in \Theta$; see Kokonendji and Senga Kiessé (2011). In what follows, we will call $K \equiv K_\theta$ the *type of kernel* to make a difference from the classical notion of a continuous symmetric (e.g., Gaussian) kernel. In this context, the choice of the associated kernel becomes important as well as that of the bandwidth. Moreover, we distinguish the associated kernels said sometimes of "second order" of those said of "first order" which verify the two first conditions (1) and (2). The rest of this section is devoted to discuss examples of associated kernels in both discrete and continuous cases.

Discrete associated kernels

Among the discrete associated kernels found in literature, we here use the best in sense of Definition 34.2.1. Negative binomial and Poisson kernels are respectively overdispersed (i.e., $\mathbb{V}(\mathcal{Z}_{x,h}) > E(\mathcal{Z}_{x,h})$) and equidispersed (i.e., $\mathbb{V}(\mathcal{Z}_{x,h}) = \mathbb{E}(\mathcal{Z}_{x,h})$) and thus are not recommended; see Kokonendji and Senga Kiessé (2011) for further details. The first associated kernel listed below, namely the binomial kernel, is the best of the first order or *standard* kernels which satisfies

$$\lim_{h \to 0} \mathbb{V}(\mathcal{Z}_{x,h}) \in \mathcal{V}(0), \tag{4}$$

where $\mathcal{V}(0)$ is a neighborhood of 0 which does not depend on x. The two other discrete associated kernels satisfy all conditions of Definition 34.2.1.

- The binomial (bino) kernel is defined on the support $S_x = \{0, 1, \ldots, x + 1\}$ with $x \in \mathbb{T} := \mathbb{N} = \{0, 1, \ldots\}$ and then $h \in (0, 1]$:

$$B_{x,h}(u) = \frac{(x+1)!}{u!(x+1-u)!} \left(\frac{x+h}{x+1}\right)^u \left(\frac{1-h}{x+1}\right)^{x+1-u} \mathbb{1}_{S_x}(u),$$

  where $\mathbb{1}_A$ denotes the indicator function of any given event A. Note that $B_{x,h}$ is the p.m.f. of the binomial distribution $\mathcal{B}(x + 1; (x + h)/(x + 1))$ with its number of trials $x + 1$ and its success probability in each trial $(x + h)/(x + 1)$. It is appropriate for count data with small or moderate sample sizes and, also, it satisfies (4) rather than (3); see Kokonendji and Senga Kiessé (2011) and also Zougab et al. (2012) for a bandwidth selection by Bayesian method.

- The following class of symmetric discrete triangular kernels has been proposed in Kokonendji et al. (2007). The support $\mathbb{T}$ of the p.m.f. f to be estimated, can be unbounded (*e.g.,* $\mathbb{N}, \mathbb{Z}$) or finite (*e.g.,* $\{0, 1, \ldots, N\}$). Then, suppose that h is a given bandwidth parameter and a is an arbitrary and fixed integer. For fixed arm $a \in \mathbb{N}$, the discrete triangular (DTra) kernel is defined on $\mathbb{S}_{x,a} = \{x, x \pm 1, \ldots, x \pm a\}$ with $x \in \mathbb{T} = \mathbb{N}$:

$$DT_{x,h;a}(u) = \frac{(a+1)^h - |u - x|^h}{P(a, h)} \mathbb{1}_{\mathbb{S}_{x,a}}(u),$$

where $P(a, h) = (2a + 1)(a + 1) - 2 \sum_{k=0}^{a} k^h$ is the normalizing constant. It is symmetric around the target x, satisfying Definition 34.2.1 and suitable for count variables; see Kokonendji and Zocchi (2010) for an asymmetric version. Note that $h \to 0$ gives the Dirac kernel.

- A discrete kernel estimator for categorical data has been introduced in Aitchison and Aitken (1976). Its asymmetric discrete associated kernel version that we here label DiracDU (DirDU) as "Dirac Discrete Uniform" has been deduced in Kokonendji and Senga Kiessé (2011) as follows. For fixed $c \in \{2, 3, \ldots\}$ the number of categories, we define $\mathbb{S}_c = \{0, 1, \ldots, c - 1\}$ and

$$DU_{x,h;c}(u) = (1 - h) \mathbb{1}_{\{x\}}(u) + \frac{h}{c-1} \mathbb{1}_{\mathbb{S}_c \setminus \{x\}}(u),$$

where $h \in (0, 1]$ and $x \in \mathbb{T} = \mathbb{S}_c$. In addition, the target x can be considered as the reference point of f to be estimated; and, the smoothing parameter h is such that $1 - h$ is the success probability of the reference point. This DiracDU kernel is symmetric around the target, satisfying Definition 34.2.1 and appropriated for categorical set $\mathbb{T}$. See, *e.g.,* Racine and Li (2004) for some uses. Note that $h = 0$ provides the Dirac kernel.

Continuous associated kernels

One can find several continuous associated kernels in literature among the Birnbaum-Saunders of Jin and Kawczak (2003). Here, we present seven associated kernels well adapted for the estimations of density or regression functions on any compact or nonnegative support of datasets. All these associated kernels satisfy Definition 34.2.1.

- The extended beta (BE) kernel is defined on $\mathbb{S}_{x,h,a,b} = [a, b] = \mathbb{T}$ with $a < b < \infty$, $x \in \mathbb{T}$ and $h > 0$ such that

$$BE_{x,h,a,b}(u) = \frac{(u - a)^{(x-a)/\{(b-a)h\}} (b - u)^{(b-x)/\{(b-a)h\}}}{(b - a)^{1+h^{-1}} B\left(1 + (x - a)/(b - a)h, 1 + (b - x)/(b - a)h\right)} \mathbb{1}_{\mathbb{S}_{x,h,a,b}}(u),$$

where $B(r, s) = \int_0^1 t^{r-1}(1 - t)^{s-1} dt$ is the usual beta function with $r > 0$, $s > 0$; see Libengué (2013). For $a = 0$ and $b = 1$, it corresponds to the beta kernel (Chen, 1999) which is the p.d.f. of the beta distribution with shape parameters $1 + x/h$ and $(1 - x)/h$. The extended beta kernel is appropriate for any compact support of observations.

- The gamma (GA) kernel is given on $\mathbb{S}_{x,h} = [0, \infty) = \mathbb{T}$ with $x \in \mathbb{T}$ and $h > 0$:

$$GA_{x,h}(u) = \frac{u^{x/h}}{\Gamma(1 + x/h) h^{1+x/h}} \exp\left(-\frac{u}{h}\right) \mathbb{1}_{[0,\infty)}(u),$$

where $\Gamma(v) = \int_0^\infty s^{v-1} \exp(-s) ds$ is the classical gamma function with $v > 0$; see Chen (2000). It is the p.d.f. of the gamma distribution $\mathcal{GA}(1 + x/h, h)$ with scale parameter $1 + x/h$ and shape parameter h. It is suitable for the non-negative real set $\mathbb{T} = [0, \infty)$.

- The lognormal (LN) kernel is defined on $\mathbb{S}_{x,h} = [0, \infty) = \mathbb{T}$ with $x \in \mathbb{T}$ and $h > 0$ such that

$$LN_{x,h}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{1}{h}\log(\frac{u}{x}) - h\right)^2\right\} \mathbb{1}_{\mathbb{S}_{x,h}}(u);$$

see Libengué (2013) and also Igarashi and Kakizawa (2015). It is the p.d.f. of the classical lognormal distribution with mean $\log(x) + h^2$ and standard deviation h.

- The reciprocal inverse Gaussian (RIG) kernel is given on $\mathbb{S}_{x,h} = (0, \infty) = \mathbb{T}$ with $x \in \mathbb{T}$ and $h > 0$:

$$RIG_{x,h}(u) = \frac{1}{\sqrt{2\pi h u}} \exp\left\{-\frac{(x^2 + xh)^{1/2}}{2h}\left(\frac{u}{(x^2 + xh)^{1/2}} - 2 + \frac{(x^2 + xh)^{1/2}}{u}\right)\right\} \mathbb{1}_{\mathbb{S}_{x,h}}(u);$$

(a)

(b)
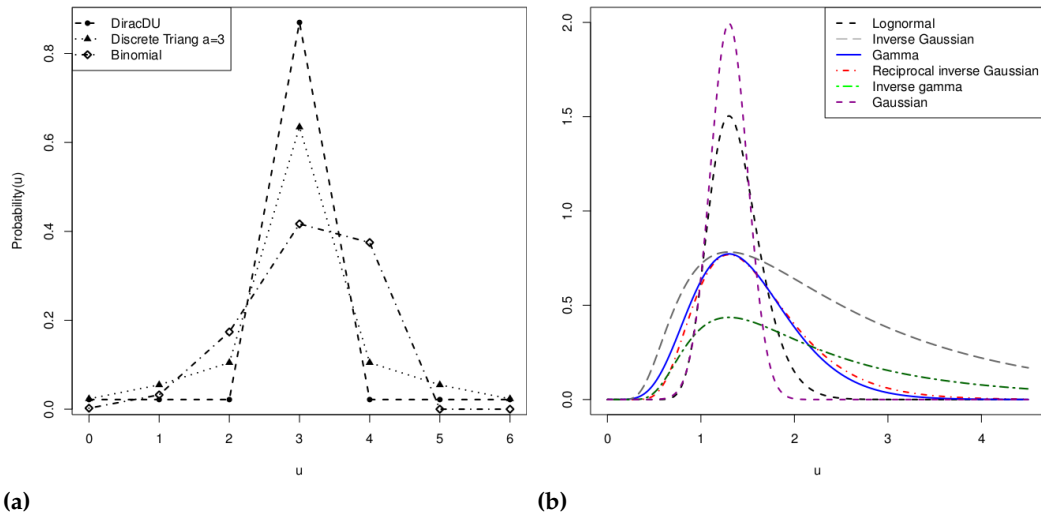
Figure 1: Shapes of univariate (discrete and continuous) associated kernels: (a) DiracDU, discrete triangular $a = 3$ and binomial with same target $x = 3$ and bandwidth $h = 0.13$; (b) lognormal, inverse Gaussian, gamma, reciprocal inverse Gaussian, inverse gamma and Gaussian with same target $x = 1.3$ and $h = 0.2$.

see Scaillet (2004), Libengué (2013) and also Igarashi and Kakizawa (2015). It is the p.d.f. of the classical reciprocal inverse Gaussian distribution with mean $1/\sqrt{x^2 + xh}$ and standard deviation $1/h$.

Remark 34.2.3. *The three continuous associated kernels inverse gamma, inverse Gaussian and Gaussian are not adapted for density estimation on supports $[0, \infty)$ and thus are not included in the* **Ake** *package; see Part (b) of Figure 1.*

Indeed:

- The inverse gamma (IGA) kernel, defined on $\mathbb{S}_{x,h} = (0, \infty) = \mathbb{T}$ with $x \in (0, 1/h)$ and $h > 0$ such that

$$IGA_{x,h}(u) = \frac{h^{1-1/(xh)}}{\Gamma\left(-1 + 1/(xh)\right)} u^{-1/(xh)} \exp\left(-\frac{1}{hu}\right) \mathbb{1}_{(0,\infty)}(u)$$

  (Libengué, 2013), is graphically the worst since it does not well concentrate on the target x. Note that it is the p.d.f. of the inverse gamma distribution with scale parameter $-1 + 1/(xh)$ and scale parameter $1/h$.

- Also, the inverse Gaussian (IG) kernel, defined on $\mathbb{S}_{x,h} = (0, \infty) = \mathbb{T}$ with $x \in (0, 1/3h)$ and $h > 0$ by

$$IG_{x,h}(u) = \frac{1}{\sqrt{2\pi h u}} \exp\left\{ -\frac{(1-3xh)^{1/2}}{2h} \left(\frac{u}{(1-3xh)^{1/2}} - 2 + \frac{(1-3xh)^{1/2}}{u} \right) \right\} \mathbb{1}_{\mathbb{S}_{x,h}}(u)$$

  (Scaillet, 2004; Libengué, 2013), has the same graphical properties as the inverse gamma. Note that it is the p.d.f. of the inverse Gaussian distribution $\mathcal{IG}(1 + x/h, h)$ with scale parameter $x/(1-3xh)^{1/2}$ and shape parameter $1/h$.

- From
  ated version (Gaussian) on $\mathbb{S}_{x,h} = \mathbb{R}$ with $x \in \mathbb{T} := \mathbb{R}$ and $h > 0$:

$$K^G_{x,h}(u) = \frac{1}{h\sqrt{2\pi}} \exp\left\{ \frac{1}{2} \left(\frac{u-x}{h} \right)^2 \right\} \mathbb{1}_{\mathbb{R}}(u).$$

  It has the same shape at any target and thus is well adapted for continuous variables with unbounded supports but not for $[0, \infty)$ or compact set of $\mathbb{R}$; see also Epanechnikov (1969) for another example of a continuous symmetric kernel.

Figure 1 shows some forms of the above-mentioned univariate associated kernels. The plots highlight the importance given to the target point and around it in discrete (a) and continuous (b) cases. Furthermore, for a fixed bandwidth h, the Gaussian keeps its same shape along the support; however, they change according to the target for the other non-classical associated kernels. This

Arguments	Description
x	The target.
t	The single or the grid value where the function is computed.
h	The bandwidth or smoothing parameter.
ker	The associated kernel.
a0,a1	The left and right bounds of the support for the extended beta kernel.
a	The arm of the discrete triangular kernel. Default value is 1.
c	The number of categories in DiracDU kernel. Default value is 2.
Result	**Description**
	Returns a single value of the associated kernel function.

Table 1: Summary of arguments and results of `kern.fun`.

explains the inappropriateness of the Gaussian kernel for density or regression estimation in any bounded interval and of the DiracDU kernel for count regression estimation; see Part (b) of Figure 1. From Part (*v*) of Remark 34.2.2, the inverse gamma and inverse Gaussian are the worst since they do not well concentrate on the target *x*; see Remark 34.2.3. These previous associated kernels can be applied to various functionals.

We have implemented in R the method `kern.fun` for both discrete and continuous associated kernels. Seven possibilities are allowed for the kernel function. We enumerate the arguments and results of the default `kern.fun.default` function in Table 1. The `kern.fun` is used as follows for the binomial kernel:

```
R> x <- 5
R> h <- 0.1
R> y <- 0:10
R> k_b <- kern.fun(x, y, h, "discrete", "bino")
```

Density or probability mass function estimations

The p.d.f. or p.m.f. estimation is an usual application of the associated kernels. Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables with an unknown p.d.f. (respectively p.m.f.) f on $\mathbb{T}$. An associated kernel estimator $\widehat{f}_n$ of f is simply:

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{T}. \tag{5}$$

Here, we point out some pointwise properties of the estimator (5) in both discrete and continuous cases.

Proposition 34.3.1. *(Kokonendji and Senga Kiessé, 2011; Libengué, 2013) Let $X_1, X_2, \dots, X_n$ be an n random sample i.i.d. from the unknown p.m.f. (respectively p.d.f.) f on $\mathbb{T}$. Let $\widehat{f}_n = \widehat{f}_{n,h,K}$ be an estimator (5) of f with an associated kernel. Then, for all $x \in \mathbb{T}$ and $h > 0$, we have*

$$\mathbb{E}\{\widehat{f}(x)\} = \mathbb{E}\{f(\mathscr{Z}_{x,h})\},$$

where $\mathscr{Z}_{x,h}$ is the random variable associated to the p.m.f. (respectively p.d.f.) $K_{x,h}$ on $\mathbb{S}_{x,h}$. Furthermore, for a p.m.f. (respectively p.d.f.), we have respectively $\widehat{f}_n(x) \in [0,1]$ (respectively $\widehat{f}_n(x) > 0$) for all $x \in \mathbb{T}$ and

$$\int_{x \in \mathbb{T}} \widehat{f}_n(x) \nu(dx) = C_n, \tag{6}$$

where $C_n = C(n; h, K)$ is a positive and finite constant if $\int_{\mathbb{T}} K_{x,h}(t) \nu(dx) < \infty$ for all $t \in \mathbb{T}$, and ν is a count or Lebesgue measure on $\mathbb{T}$.

It is easy to see that $C_n = 1$ for the estimators (5) with DiracDU kernel or any classical (symmetric)

associated kernel. Indeed, for the DiracDU kernel estimation we have

$$\sum_{x=0}^{c-1} \widehat{f}_n(x) = \sum_{x=0}^{c-1} \left\{ (1-h)\, \mathbb{1}_{\{x\}}(X_1) + \frac{h}{c-1}\mathbb{1}_{S_c \setminus \{x\}}(X_1) \right\}$$

$$= (1-h) + \frac{h}{c-1}(c-1)$$

$$= 1.$$

In general we have $C_n \neq 1$ for other discrete and also continuous associated kernels, but it is always close to 1. In practice, we compute C_n depending on observations before normalizing $\widehat{f}_n$ to be a p.m.f. or a p.d.f. The following code helps to compute the normalizing constant, e.g., for gamma kernel estimation:

```
R> data("faithful", package = "datasets")
R> x <- faithful$waiting
R> f <- dke.fun(x, ker = "GA", 0.1)
R> f$C_n
```

```
[1] 0.9888231
```

Without loss of generality, we study $x \mapsto \widehat{f}_n(x)$ up to a normalizing constant which is used at the end of the density estimation process. Notice that that non-classical associated kernel estimators $\widehat{f}_n$ are improper density estimates or as kind of "balloon estimators"; see Sain (2002). There are two ways to normalize these estimators (5). The first method is the *global* normalization using C_n of (6):

$$\widetilde{f}_n(x) = \frac{\widehat{f}_n(x)}{\displaystyle\int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \widehat{f}_n(x)\nu(dx)}, \quad x \in \mathbb{T}. \tag{7}$$

Another alternative is to use an *adaptive* normalization of (5) according to each target x:

$$\widetilde{\widetilde{f}}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{K_{x,h}(X_i)}{\displaystyle\int_{inf(\mathbb{T})}^{sup(\mathbb{T})} K_{x,h}(X_i)\nu(dx)}, \quad x \in \mathbb{T},$$

but this approach, with similar results than (7), is not used here. The representations are done with the global normalization (7). In the package, we also compute the normalizing constant (6) for any data set.

In discrete cases, the integrated squared error (ISE) defined by

$$ISE_0 = \sum_{x \in \mathbb{N}} \{\widetilde{f}_n(x) - f_0(x)\}^2,$$

is the criteria used to measure numerically the discrete smoothness of $\widetilde{f}_n$ from (7) with the empirical or naive p.m.f. f_0 such that $\sum_{x \in \mathbb{N}} f_0(x) = 1$; see, e.g., Kokonendji and Senga Kiessé (2011). Concerning the continuous variables, the histogram gives a graphical measure of comparison with $\widetilde{f}_n$; see, for example, Figure 2.

Some theoretical aspects of the normalizing constant

In this section, we present some theoretical aspects of the normalizing constant C_n of (6) and two examples in the continuous and discrete cases. We first recall the following result on pointwise properties of the estimator (5).

Lemma 34.3.2. (*Kokonendji and Senga Kiessé, 2011; Libengué, 2013*) *Let $x \in \mathbb{T}$ be a target and $h \equiv h_n$ a bandwidth. Assuming f is in the class $\mathscr{C}^2(\mathbb{T})$ in the continuous case, then*

$$Bias\left\{ \widehat{f}_n(x) \right\} = A(x,h)f'(x) + \frac{1}{2}\left\{ A^2(x,h) + B(x,h) \right\} f''(x) + o(h^2). \tag{8}$$

Similar expressions (8) hold in the discrete case, except that f' and f'' are finite differences of the first and second order respectively.

Furthermore, for the continuous case, if f is bounded on $\mathbb{T}$ then there exists $r_2 = r_2\left(K_{x,h}\right) > 0$ the largest

real number such that $\left\| K_{x,h} \right\|_2^2 := \int_{S_{x,h}} K_{x,h}^2(u) du \leq c_2(x) h^{-r_2}$, $0 \leq c_2(x) \leq \infty$ *and*

$$\mathbb{V}\left\{ \widehat{f}_n(x) \right\} = \frac{1}{n} f(x) \left\| K_{x,h} \right\|_2^2 + o\left(\frac{1}{nh^{r_2}} \right). \tag{9}$$

For discrete situations, the result (9) *becomes*

$$\mathbb{V}\left\{ \widehat{f}_n(x) \right\} = \frac{1}{n} f(x) [\{\mathbb{P}(\mathcal{Z}_{x,h} = x)\}^2 - f(x)],$$

where $\mathcal{Z}_{x,h}$ *denotes the discrete random variable with p.m.f.* $K_{x,h}$.

It is noticeable that the bias (8) is bigger than the one with symmetric kernels and thus can be reduced; see, e.g., Zhang (2010), Zhang and Karunamuni (2010) and Libengué (2013).

Proposition 34.3.3. *Following notations in Lemma* 34.3.2, *the mean and variance of* C_n *of* (6) *are respectively:*

$$\mathbb{E}(C_n) \simeq 1 + \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left\{ A(x,h) f'(x) + \frac{1}{2} \left[A^2(x,h) + B(x,h) \right] f''(x) \right\} \nu(dx), \tag{10}$$

$$\mathbb{V}(C_n) \simeq
\begin{cases}
\dfrac{1}{n} \displaystyle\int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left(f(x) \left\| K_{x,h} \right\|_2^2 \right) dx & \text{if } \mathbb{T} \text{ is continuous,} \\[2em]
\dfrac{1}{n} \displaystyle\sum_{x \in \mathbb{T}} \left(f(x) [\{\mathbb{P}(\mathcal{Z}_{x,h} = x)\}^2 - f(x)] \right) & \text{if } \mathbb{T} \text{ is discrete,}
\end{cases} \tag{11}$$

where $inf(\mathbb{T})$ *and* $sup(\mathbb{T})$ *are respectively the infimum and supremum of* $\mathbb{T}$, *the measure* ν *is Lesbesgue or count on the support* $\mathbb{T}$, *and where* "$\simeq$" *stands for approximation.*

Proof. From Lemma 34.3.2 and the Fubini theorem, we successively show (10) as follows:

$$\mathbb{E}(C_n) = \mathbb{E}\left(\int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \widehat{f}_n(x) \nu(dx) \right) = \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \mathbb{E}\left(\widehat{f}_n(x) \right) \nu(dx)$$

$$= \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left(Bias\left\{ \widehat{f}_n(x) \right\} + f(x) \right) \nu(dx)$$

$$\simeq \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left\{ A(x,h) f'(x) + \frac{1}{2} \left[A^2(x,h) + B(x,h) \right] f''(x) + f(x) \right\} \nu(dx)$$

$$\simeq \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} f(x) \nu(dx) + \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left\{ A(x,h) f'(x) + \frac{1}{2} \left[A^2(x,h) + B(x,h) \right] f''(x) \right\} \nu(dx)$$

$$\simeq 1 + \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left\{ A(x,h) f'(x) + \frac{1}{2} \left[A^2(x,h) + B(x,h) \right] f''(x) \right\} \nu(dx).$$

The variance (11) is trivial from Lemma 34.3.2. □

Example 1. Let f be an exponential density with parameter $\gamma > 0$. Thus, one has:

$$f(x) = \gamma \exp(-\gamma x),$$
$$f'(x) = -\gamma^2 \exp(-\gamma x),$$
$$f''(x) = \gamma^3 \exp(-\gamma x).$$

Consider the lognormal kernel with $A(x,h) = x(\exp(3h^2/2) - 1)$, $B(x,h) = x^2 \exp(3h^2)(\exp(h^2) - 1)$ and $\left\| LN_{x,h} \right\|_2^2 = 1/(2\pi h x^{1/2})$; see Libengué (2013). Then, using the Taylor formula around h, the expressions of $\mathbb{E}(C_n)$ and $\mathbb{V}(C_n)$ are:

$$\mathbb{E}(C_n) \simeq 1 + \left[\left(-1 + \exp\frac{3h^2}{2} \right) + \frac{1}{2}\left\{ \left(-1 + \exp\frac{3h^2}{2} \right)^2 + \exp 3h^2 \left(-1 + \exp h^2 \right) \right\} \right] \simeq 1 - \frac{h^2}{2}$$

and $\mathbb{V}(C_n) \simeq \gamma (2nh\sqrt{\pi})^{-1} \int_0^\infty z^{-1} \exp(-z) dz$ with $\int_0^\infty z^{-1} \exp(-z) dz \approx 16.2340$ by computation with R. Thus, the quantity C_n cannot be equal to 1.

Example 2. Let f be a Poisson p.m.f. with parameter λ and thus $f(x) = \lambda^x \exp(-\lambda)/x!$. The finite differences $f^{(k)}(x)$ of order $k \in \{1, 2, \ldots\}$ at $x \in \mathbb{N}$ are given by the recursive relation:

$$f^{(k)}(x) = \{f^{(k-1)}(x)\}^{(1)} \text{ with } f^{(1)}(x) =
\begin{cases}
\{f(x+1) - f(x-1)\}/2, & \text{if } x \in \mathbb{N} \setminus \{0\}, \\
f(1) - f(0), & \text{if } x = 0,
\end{cases}$$

and

$$f^{(2)}(x) = \begin{cases} \{f(x+2) - 2f(x) + f(x-2)\}/4, & \text{if } x \in \mathbb{N} \setminus \{0,1\}, \\ \{f(3) - 3f(1) + f(0)\}/4, & \text{if } x = 1, \\ \{f(2) - 2f(1) + f(0)\}/2, & \text{if } x = 0. \end{cases}$$

Considering the binomial kernel with $A(x,h) = x+h$, $B(x,h) = (x+h)(1-h)/(x+1)$, we successively obtain

$$\mathbb{E}(C_n) \simeq 1 + \frac{h}{4}\left(f(3) + 3f(2) - 3f(1) - 2f(0) \right.$$

$$+ \sum_{x=2}^{\infty} 2\{f(x+1) - f(x-1)\} + \frac{(x+3)\{f(x+2) - 2f(x) + f(x-2)\}}{x+1} \Bigg)$$

$$+ \left(\frac{3f(3) + 8f(2) - 17f(1) + 3f(0)}{2} \right.$$

$$+ \sum_{x=2}^{\infty} \frac{x\{f(x+1) - f(x-1)\}}{2} + \frac{(x^3 + x^2 + 1)\{f(x+2) - 2f(x) + f(x-2)\}}{x+1} \Bigg)$$

$$\simeq 1 + \frac{h\exp(-\lambda)}{4}\left[\frac{\lambda^3}{3!} + 3\frac{\lambda^2}{2!} - 3\lambda - 2 \right.$$

$$+ \sum_{x=2}^{\infty} \left\{ \frac{2\lambda^{x+1}}{(x+1)!} - \frac{2\lambda^{x-1}}{(x-1)!} + \frac{x+3}{x+1}\left(\frac{\lambda^{x+2}}{(x+2)!} - 2\frac{\lambda^x}{x!} + \frac{\lambda^{x-2}}{(x-2)!} \right) \right\} \Bigg]$$

$$+ \left[\frac{\lambda^3}{4} + 2\lambda^2 - \frac{17\lambda}{2} + 3 \right.$$

$$+ \sum_{x=2}^{\infty} \left\{ \frac{x}{2}\left(\frac{\lambda^{x+1}}{(x+1)!} - \frac{\lambda^{x-1}}{(x-1)!} \right) + \frac{x^3 + x^2 + 1}{x+1}\left(\frac{\lambda^{x+2}}{(x+2)!} - 2\frac{\lambda^x}{x!} + \frac{\lambda^{x-2}}{(x-2)!} \right) \right\} \Bigg]$$

and

$$\mathbb{V}(C_n) \simeq \frac{\exp(-\lambda)}{n} \sum_{x \in \mathbb{T}} \left(\frac{\lambda^x}{x!} \left[\left\{ (1+h)\left(\frac{x+h}{x+1} \right)^x \right\}^2 - \frac{\lambda^x \exp(-\lambda)}{x!} \right] \right).$$

Bandwidth selection

Now, we consider the bandwidth selection problems which are generally crucial in nonparametric estimation. Several methods already existing for continuous kernels can be adapted to the discrete case as the classical least-squares cross-validation method; see, for example, Bowman (1984), Marron (1987) and references therein. Here, we simply propose three procedures for the bandwidth selection: cross-validation, Bayesian local for binomial and adaptive for the gamma kernel. Also, a review of bayesian bandwidth selection methods is presented. Each time, the smoothing parameter selection is done with the non-normalized version $\widehat{f}_n$ of the estimator (5) before the global normalization $\widetilde{f}_n$ of (7).

Cross-validation for any associated kernel

For a given associated kernel $K_{x,h}$ with $x \in \mathbb{T}$ and $h > 0$, the optimal bandwidth h_{cv} of h is obtained by cross-validation as $h_{cv} = \arg\min_{h>0} CV(h)$ with

$$CV(h) = \int_{x \in \mathbb{T}} \left\{ \widehat{f}_n(x) \right\}^2 \nu(dx) - \frac{2}{n} \sum_{i=1}^{n} \widehat{f}_{n,-i}(X_i),$$

where $\widehat{f}_{n,-i}(X_i) = (n-1)^{-1} \sum_{j \neq i} K_{X_i,h}(X_j)$ is being computed as $\widehat{f}_n(X_i)$ by excluding the observation X_i and ν is the Lebesgue or count measure. This method is applied to all estimators (5) with associated kernels cited in this paper, independently on the support $\mathbb{T}$ of f to be estimated.

Table 2 gives the arguments and results of the cross-validation function hcvc.fun defined for continuous data are below. The hcvd.fun is the corresponding function for discrete data. The hcvc.fun is performed with the Old Faithful geyser data described in Azzalini and Bowman (1990) and Härdle (2012). The dataset concerns waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. The following codes and Figure 2 give smoothing density estimation with various associated kernels of the waiting time variable.

Arguments	Description
Vec	The positive continuous data sample.
seq.bws	The sequence of bandwidths where to compute the cross-validation function.
ker	The associated kernel.
a0,a1	The bounds of the support of extended beta kernel. Default values are respectively 0 and 1.
a	The arm of the discrete triangular kernel. Default value is 1.
c	The number of categories in DiracDU kernel. Default value is 2.

Results	Description
hcv	The optimal bandwidth obtained by cross-validation.
seq.h	The sequence of bandwidths used to compute hcv.
CV	The values of the cross-validation function.

Table 2: Summary of arguments and results of hcvc.fun.
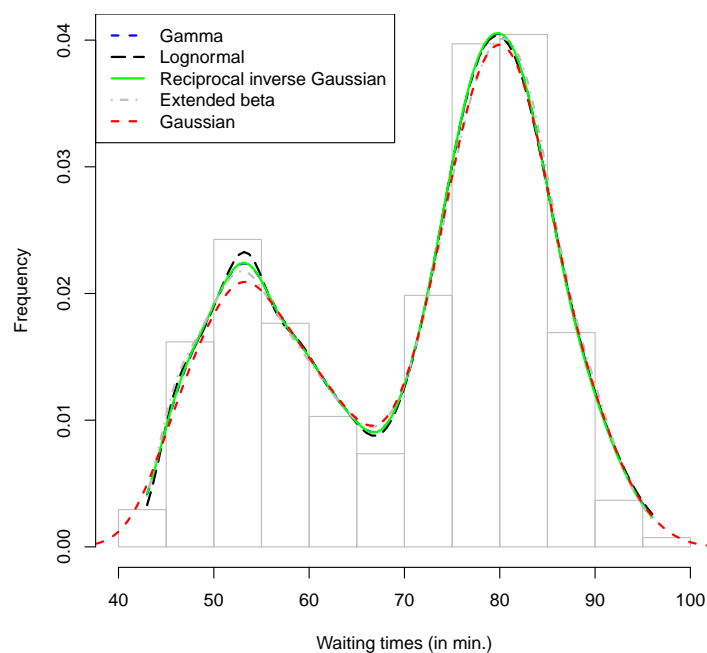
Figure 2: Smoothing density estimation of the Old Faithful geyser data (Azzalini and Bowman, 1990) by some continuous associated kernels with the support of observations $[43, 96] = \mathbb{T}$.

```
R> data("faithful", package = "datasets")
R> x <- faithful$waiting
R> f1 <- dke.fun(x, 0.1, "continuous", ker = "GA")
R> f2 <- dke.fun(x, 0.036, "continuous", ker = "LN")
R> f3 <- dke.fun(x, 0.098, "continuous", ker = "RIG")
R> f4 <- dke.fun(x, 0.01, "continuous", ker = "BE", a0 = 40, a1 = 100)
R> t <- seq(min(x), max(x), length.out = 100)
R> hist(x, probability = TRUE, xlab = "Waiting times (in min.)",
+    ylab = "Frequency", main = "", border = "gray")
R> lines(t, f1$fn, lty = 2, lwd = 2, col = "blue")
R> lines(t, f2$fn, lty = 5, lwd = 2, col = "black")
R> lines(t, f3$fn, lty = 1, lwd = 2, col = "green")
R> lines(t, f4$fn, lty = 4, lwd = 2, col = "grey")
R> lines(density(x, width = 12), lty = 8, lwd = 2, col = "red")
R> legend("topleft", c("Gamma", "Lognormal", "Reciprocal inverse Gaussian",
+    "Extended beta", "Gaussian"), col = c("blue", "black", "green", "grey", "red"),
+    lwd = 2, lty = c(2, 5, 1, 4, 8), inset = .0)
```

A review of Bayesian bandwidth selection

Bayesian inference grows out of the simple formula known as Bayes rule. Assume we have two random variables A and B. A principle rule of probability theory known as the chain rule allows us to specify the joint probability of A and B taking on particular values a and b, $P(a, b)$, as the product of the conditional probability that A will take on value a given that B has taken on value b, $P(a|b)$, and the marginal probability that B takes on value b, $P(b)$. Which gives us:

Joint probability = Conditional Probability $\times$ Marginal Probability.

Thus we have: $P(a, b) = P(a|b)P(b)$.

This expression (Bayes rule) indicates that we can compute the conditional probability of a variable A given the variable B from the conditional probability of B given A. This introduces the notion of prior and posterior knowledge.

Prior and posterior knowledge. A prior probability is the probability available to us beforehand, and before making any additional observations. A posterior probability is the probability obtained from the prior probability after making additional observation to the prior knowledge available.

Summarizing the Bayesian approach. The Bayesian approach to parameter estimation works as follows:

1. Formulate our knowledge about a situation.
2. Gather data.
3. Obtain posterior knowledge that updates our beliefs.

How do we formulate our knowledge about a situation?

a. Define a distribution model which expresses qualitative aspects of our knowledge about the situation. This model will have some unknown parameters, which will be dealt with as random variables.
b. Specify a prior probability distribution which expresses our subjective beliefs and subjective uncertainty about the unknown parameters, before seeing the data.

After gathering the data, how do we obtain posterior knowledge?

c. Compute posterior probability distribution which estimates the unknown parameters using the rules of probability and given the observed data, presenting us with updated beliefs.

Zougab et al. (2013) proposed a Bayesian approach based upon a likelihood cross-validation approximation and a Markov chain Monte Carlo (MCMC) method for deriving the global optimal bandwidth using the famous binomial kernel. However, a global bandwidth does not generally provide a good estimator for complex p.m.f.'s, in particular for small and moderate sample sizes. Generally, the global discrete associated kernel estimator tends to simultaneously under- and oversmooth $f(x)$.

In order to improve the global discrete associated kernel estimator, in particular for complex count data with small and moderate sample sizes, Zougab et al. (2012) and Zougab et al. (2013) adapted two versions of variable bandwidths for discrete associated kernel estimator and proposed Bayesian approaches for selecting these variable bandwidths. Note that these two versions are originally proposed for kernel density estimation (see, e.g., Sain and Scott 1996; Breiman et al. 1977; Abramson 1982; Brewer 2000; Zhang et al. 2006; Zougab et al. 2014a and Zhang et al. 2016).

Recently, Zougab et al. (2012) have considered the local discrete associated kernel estimator (balloon estimator in discrete case) and have derived the closed form of the variable bandwidth at each point x for which the p.m.f. is estimated by considering the binomial kernel estimator and locally treating the bandwidth as a random quantity with a beta prior distribution. This approach outperforms existing classical global methods, namely, MISE and CV in particular for small and moderate sample sizes. Zougab et al. (2013) have also proposed the adaptive discrete associated kernel estimator (sample-point estimator in discrete situation), which replaces h by h_i for each observation x_i with $i = 1, \ldots, n$, and then employs the Bayesian approach for estimating the adaptive bandwidths h_i. The authors have considered the binomial kernel and the beta prior for each variable bandwidth h_i, and have shown that this approach performs better than the popular classical global selectors.

Bayesian estimation of localized bandwidth for the binomial kernel

An alternative to the cross-validation for bandwidth selection is by using Bayesian methods. These methods have been investigated with three different procedures: local, global and adaptive; see respectively Zougab et al. (2012, 2013, 2014b). In terms of integrated squared error and execution times, the local Bayesian outperforms the other Bayesian procedures. In the local Bayesian framework, the variable bandwidth is treated as parameter with prior $\pi(\cdot)$. Under squared error loss function, the Bayesian bandwidth selector is the posterior mean; see Zougab et al. (2012).

First, as we have mentioned above, $f(x)$ can be approximated by

$$f(x|h) = f_h(x) = \sum_{u \in \mathbb{T}} f(u) B_{x,h}(u) = \mathbb{E}\{B_{x,h}(X)\},$$

where $B_{x,h}$ is the binomial kernel and X is a random variable with p.m.f. f. Now, considering h as a scale parameter for $f_h(x)$, the local approach consists of using $f_h(x)$ and constructing a Bayesian estimator for $h(x)$.

Indeed, let $\pi(h)$ denote the beta prior density of h with positive parameters α and β. By the Bayes theorem, the posterior of h at the point of estimation x takes the form

$$\pi(h|x) = \frac{f_h(x)\pi(h)}{\int f_h(x)\pi(h)dh}.$$

Since f_h is unknown, we use $\widehat{f}_h$ as natural estimator of f_h, and hence we can estimate the posterior by

$$\pi(h|x, X_1, X_2, \dots, X_n) = \frac{\widehat{f}_h(x)\pi(h)}{\int \widehat{f}_h(x)\pi(h)dh}.$$

Under the squared error loss, the Bayes estimator of the smoothing parameter $h(x)$ is the posterior mean and is given by $\widehat{h}_n(x) = \int h\widehat{\pi}(h|x, X_1, X_2, \dots, X_n)dh$. Exact approximation is

$$\widehat{h}_n(x) = \frac{\displaystyle\sum_{i=0}^{n}\sum_{k=0}^{X_i} \frac{x^k}{(x+1-X_i)!k!(X_i-k)!} B(X_i + \alpha - k + 1, x + \beta + 1 - X_i)}{\displaystyle\sum_{i=0}^{n}\sum_{k=0}^{X_i} \frac{x^k}{(x+1-X_i)!k!(X_i-k)!} B(X_i + \alpha - k, x + \beta + 1 - X_i)}, \forall x \in \mathbb{N} \text{ with } X_i \leq x + 1,$$

where $B(\cdot, \cdot)$ is the beta function; see Zougab et al. (2012) for more details.

Bayesian estimation of adaptive bandwidth for the gamma kernel

The bandwidth h in the gamma kernel density estimation can be allowed to be adaptive. This approach gives a variable bandwidth h_i for each observation X_i in place of the initial fixed bandwidth h. Following Zougab et al. (2014a), we suggest using Bayesian methods to estimate such adaptive bandwidths or variable bandwidths $h_i, i = 1, \dots, n$. Thus, we treat h_i as a random variable with a prior distribution $\pi(\cdot)$. The estimator (5) with gamma kernel of Section Continuous associated kernels and variable bandwidths are reformulated as

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} GA_{x,h_i}(X_i). \tag{12}$$

The leave-one-out kernel estimator of $f(X_i)$ deduced from (12) is

$$\widehat{f}(X_i \mid \{X_{-i}\}, h_i) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} GA_{X_i,h_i}(X_j), \tag{13}$$

where $\{X_{-i}\}$ denotes the set of observations excluding X_i. The posterior distribution for each variable bandwidth h_i given X_i provided from the Bayesian rule is expressed as follow

$$\pi(h_i \mid X_i) = \frac{\widehat{f}(X_i \mid \{X_{-i}\}, h_i)\pi(h_i)}{\int_0^{\infty} \widehat{f}(X_i \mid \{X_{-i}\}, h_i)\pi(h_i)dh_i}. \tag{14}$$

We obtain the Bayesian estimator $\widetilde{h}_i$ of h_i by using the quadratic loss function

$$\widetilde{h}_i = \mathbb{E}(h_i \mid X_i). \tag{15}$$

In the following, we assume that each $h_i = h_i(n)$ has an inverse gamma prior distribution $\mathcal{IGA}(\alpha, \beta)$ with the shape parameter $\alpha > 0$ and scale parameter $\beta > 0$. The density of $\mathcal{IGA}(a, b)$ with $a, b > 0$ is defined as

$$\Phi_{a,b}(z) = \frac{b^a}{\Gamma(a)} z^{-a-1} \exp(-b/z) \mathbb{1}_{(0,\infty)}(z). \tag{16}$$

This allows us to obtain the closed form of the posterior density and the Bayesian estimator given by the following result.

Theorem 34.3.4. *For fixed $i \in \{1, 2, \ldots, n\}$, consider each observation X_i with its corresponding bandwidth h_i. Using the gamma kernel estimator (12) and the inverse gamma prior distribution $\mathcal{IGA}(\alpha, \beta)$ given in (16) with $\alpha > 1/2$ and $\beta > 0$ for each h_i, then:*

(i) The posterior density (14) is the following weighted sum of inverse gamma

$$\pi(h_i \mid X_i) = \frac{1}{D_{ij}} \sum_{j=1, j \neq i}^{n} \left\{ A_{ij} \Phi_{\alpha+1/2, B_{ij}}(h_i) \mathbb{1}_{(0,\infty)}(X_i) + C_j \Phi_{\alpha+1, X_j+\beta}(h_i) \mathbb{1}_{\{0\}}(X_i) \right\}$$

with $A_{ij} = [\Gamma(\alpha + 1/2)] / (\beta^\alpha X_i^{1/2} \sqrt{2\pi} B_{ij}^{\alpha+1/2})$, $B_{ij} = X_i \log X_i - X_i \log X_j + X_j - X_i + \beta$, $C_j = \Gamma(\alpha + 1) / [\beta^{-\alpha}(X_j + \beta)^{\alpha+1}]$ and $D_{ij} = \sum_{j=1, j \neq i}^{n} \left\{ A_{ij} \mathbb{1}_{(0,\infty)}(X_i) + C_j \mathbb{1}_{\{0\}}(X_i) \right\}$.

(ii) The Bayesian estimator $\widetilde{h}_i$ of h_i, given in (15), is

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

according to the previous notations of A_{ij}, B_{ij}, C_j and D_{ij}.

Proof. (i) Let us represent $\pi(h_i \mid X_i)$ of (14) as the ratio of $N(h_i \mid X_i) := \widehat{f}(X_i \mid \{X_{-i}\}, h_i) \pi(h_i)$ and $\int_0^\infty N(h_i \mid X_i) dh_i$. From (13) and (16) the numerator is, first, equal to

$$N(h_i \mid X_i) = \left(\frac{1}{n-1} \sum_{j=1, j \neq i}^{n} GA_{X_i, h_i}(X_j) \right) \left(\frac{\beta^\alpha}{\Gamma(\alpha)} h_i^{-\alpha-1} \exp(-\beta/h_i) \right)$$

$$= \frac{[\Gamma(\alpha)]^{-1}}{(n-1)} \sum_{j=1, j \neq i}^{n} \frac{GA_{X_i, h_i}(X_j)}{\beta^{-\alpha} h_i^{\alpha+1}} \exp(-\beta/h_i). \tag{17}$$

Following Chen (2000), we assume that for all $X_i \in (0, \infty)$ one has $1 + (X_i/h_i) \to \infty$ as $n \to \infty$. Using the Stirling formula $\Gamma(z + 1) \simeq \sqrt{2\pi} z^{z+1/2} \exp(-z)$ as $z \to \infty$, the term of the sum in (17) can be successively calculated as

$$\frac{GA_{X_i, h_i}(X_j)}{\beta^{-\alpha} h_i^{\alpha+1}} \exp(-\beta/h_i) = \frac{X_j^{(X_i/h_i)} \exp(-X_j/h_i)}{h_i^{1+(X_i/h_i)} \Gamma[1 + (X_i/h_i)] \beta^{-\alpha} h_i^{\alpha+1}} \exp(-\beta/h_i)$$

$$= \frac{\exp[-(X_j + \beta - X_i \log X_j)/h_i]}{\beta^{-\alpha} h_i^{(X_i/h_i)+\alpha+2} \sqrt{2\pi} \exp(-X_i/h_i)(X_i/h_i)^{(X_i/h_i)+1/2}}$$

$$= \frac{\Gamma(\alpha + 1/2)}{\beta^{-\alpha} X_i^{1/2} \sqrt{2\pi} B_{ij}^{\alpha+1/2}} \times \frac{B_{ij}^{\alpha+1/2} \exp[-B_{ij}/h_i]}{h_i^{\alpha+3/2} \Gamma(\alpha + 1/2)}$$

$$= A_{ij} \Phi_{\alpha+1/2, B_{ij}}(h_i), \tag{18}$$

with $B_{ij} = X_i \log X_i - X_i \log X_j + X_j - X_i + \beta$, $A_{ij} = [X_j^{-1} \Gamma(\alpha + 1/2)] / (\beta^{-\alpha} X_i^{-1/2} \sqrt{2\pi} B_{ij}^{\alpha+1/2})$ and $\Phi_{\alpha+1/2, B_{ij}}(h_i)$ is given in (16).

Also, for $X_i = 0$, the term of the sum (17) can be expressed as follows

$$\frac{GA_{0,h_i}(X_j)}{\beta^{-\alpha}h_i^{\alpha+1}}\exp(-\beta/h_i) = \frac{\exp(-X_j/h_i)}{\beta^{-\alpha}h_i^{\alpha+2}}\exp(-\beta/h_i)$$

$$= \frac{\Gamma(\alpha+1)}{\beta^{-\alpha}(X_j+\beta)^{\alpha+1}} \times \frac{(X_j+\beta)^{\alpha+1}\exp[-(X_j+\beta)/h_i]}{h_i^{\alpha+2}\Gamma(\alpha+1)}$$

$$= C_j\,\Phi_{\alpha+1,X_j+\beta}(h_i), \tag{19}$$

with $C_j = \Gamma(\alpha+1)/[\beta^{-\alpha}(X_j+\beta)^{\alpha+1}]$ and $\Phi_{\alpha+1,X_j+\beta}(h_i)$ is given in (16). Combining (18) and (19), the expression of $N(h_i \mid X_i)$ in (17) becomes

$$N(h_i \mid X_i) = \frac{[\Gamma(\alpha)]^{-1}}{(n-1)}\sum_{j=1,j\neq i}^{n}\left\{A_{ij}\,\Phi_{\alpha+1/2,B_{ij}}(h_i)\mathbb{1}_{(0,\infty)}(X_i) + C_j\,\Phi_{\alpha+1,X_j+\beta}(h_i)\mathbb{1}_{\{0\}}(X_i)\right\}. \tag{20}$$

From (20), the denominator is successively computed as follows:

$$\int_0^\infty N(h_i \mid X_i)\,dh_i = \frac{[\Gamma(\alpha)]^{-1}}{(n-1)}\sum_{j=1,j\neq i}^{n}\left(A_{ij}\int_0^\infty \Phi_{\alpha+1/2,B_{ij}}(h_i)\mathbb{1}_{(0,\infty)}(X_i)\,dh_i\right.$$

$$\left.+ C_j\int_0^\infty \Phi_{\alpha+1,X_j+\beta}(h_i)\mathbb{1}_{\{0\}}(X_i)\,dh_i\right)$$

$$= \frac{[\Gamma(\alpha)]^{-1}}{(n-1)}\sum_{j=1,j\neq i}^{n}\left\{A_{ij}\mathbb{1}_{(0,\infty)}(X_i) + C_j\mathbb{1}_{\{0\}}(X_i)\right\}$$

$$= \frac{[\Gamma(\alpha)]^{-1}}{(n-1)}D_{ij}, \tag{21}$$

with $D_{ij} = \sum_{j=1,j\neq i}^{n}\left(A_{ij}\mathbb{1}_{(0,\infty)}(X_i) + C_j\mathbb{1}_{\{0\}}(X_i)\right)$. Finally, the ratio of (20) and (21) leads to the result of Part (i).

(ii) Let us remember that the mean of the inverse gamma distribution $\mathcal{IG}(\alpha,\beta)$ is $\beta/(\alpha-1)$. Thus, the expression of $\pi(h_i \mid X_i)$ in (14) is given by

$$\pi(h_i \mid X_i) = \frac{1}{D_{ij}}\sum_{j=1,j\neq i}^{n}\left\{A_{ij}\,\Phi_{\alpha+1/2,B_{ij}}(h_i)\mathbb{1}_{(0,\infty)}(X_i) + C_j\,\Phi_{\alpha+1,X_j+\beta}(h_i)\mathbb{1}_{\{0\}}(X_i)\right\}$$

and, therefore, $\widetilde{h}_i = \mathbb{E}(h_i \mid X_i) = \int_0^\infty h_i\pi(h_i \mid X_i)\,dh_i$ is finally

$$\widetilde{h}_i = \mathbb{E}(h_i \mid X_i) = \frac{1}{D_{ij}}\sum_{j=1,j\neq i}^{n}\left\{\frac{A_{ij}B_{ij}}{\alpha-1/2}\mathbb{1}_{(0,\infty)}(X_i) + \frac{(X_j+\beta)C_j}{\alpha}\mathbb{1}_{\{0\}}(X_i)\right\}.$$

$\square$

This new method of selecting bandwidth by the Bayesian adaptive procedure will be implemented in a future version of the **Ake** package.

Bandwidth selection for kernel regression involving associated kernels

One of the most often encountered models in nonparametric statistics is the regression model. The function that provides the best prediction of a dependent variable y in terms of an independent variable x is the conditional expectation $\mathbb{E}(y/x) = m(x)$. This is called regression function and its estimation from a sequence of n pairs (x_i, y_i), $i = 1, \ldots, n$ is a problem in statistics. We will consider the case $(x,y) \in \mathbb{T} \times \mathbb{R}$. For simplicity, we take $\mathbb{T} = \mathbb{R}$ if x is a continuous variable and $\mathbb{T} = \mathbb{N}$ in the discrete case. The classical non parametric regression model between two variables y and x is

$$y_i = m(x_i) + \epsilon_i, \tag{22}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ is a response vector, $\mathbf{x} = (x_1, \ldots, x_n)$ is an explanatory vector, $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ is the error following a Gaussian distribution with zero mean and finite variance σ^2, i.e., $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and $m : \mathbb{T} \mapsto \mathbb{R}$ is the unknown regression function. Several methods have been proposed to estimate the regression function in the continuous case. We cite for example the histograms introduced by Tukey (1961) and studied by Geoffroy (1980) and Lecoutre (1990), the spline method which can be found in Reinsch (1967), Silverman (1985) and Wahba (1990) and also the regression using partition

proposed by Breiman et al. (1984).

As for the density or probability mass function, the estimate of the regression function by the kernel method is the most used because of its good asymptotic properties and interest in practice. Introduced initially for continuous density estimation by Rosenblatt (1956) and Parzen (1962), this method was adopted by Nadaraya (1964) and Watson (1964) for estimating the continuous regression function. It was also applied to smooth the discrete regression function m for $x \in \mathbb{N}$. Some studies have been done to estimate the discrete regression function, using the Dirac-type kernel (naive estimator) or discrete kernels of Aitchison and Aitken (1976). However, the naive estimator is appropriate only when the sample size is large, and the discrete kernel of Aitchison and Aitken (1976) is only suitable for categorical data; see Hayfield and Racine (2008) and also Hayfield and Racine (2014). Kokonendji et al. (2009) adapted the Nadaraya (1964) and Watson (1964) kernel to the discrete unknown function m, using the discrete associated kernels. In their work, using the integrated mean square error and the coefficient of determination R^2, they showed that the binomial or discrete triangular kernels are better compared to the optimal Epanechnikov kernel. In this section we present the theoretical foundations of the estimated regression function with continuous and discrete associated kernels.

Both in continuous and discrete cases, consider the relation between a response variable Y and an explanatory variable x given by

$$Y = m(x) + \epsilon, \tag{23}$$

where m is an unknown regression function from $\mathbb{T} \subseteq \mathbb{R}$ to $\mathbb{R}$ and ϵ the disturbance term with null mean and finite variance. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sequence of i.i.d. random variables on $\mathbb{T} \times \mathbb{R}(\subseteq \mathbb{R}^2)$ with $m(x) = \mathbb{E}(Y|X = x)$ of (23). Using (continuous or discrete) associated kernels, the Nadaraya (1964) and Watson (1964) estimator $\widehat{m}_n$ of m is

$$\widehat{m}_n(x; h) = \sum_{i=1}^{n} \frac{Y_i K_{x,h}(X_i)}{\sum_{i=1}^{n} K_{x,h}(X_i)} = \widehat{m}_n(x), \quad \forall x \in \mathbb{T} \subseteq \mathbb{R}, \tag{24}$$

where $h \equiv h_n$ is the smoothing parameter such that $h_n \to 0$ as $n \to \infty$.

Besides the criterion of kernel support, we retain the root mean squared error (RMSE) and also the practical coefficient of determination given respectively by

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \{Y_i - \widehat{m}_n(X_i)\}^2}$$

and

$$R^2 = \frac{\sum_{i=1}^{n} \{\widehat{m}_n(X_i) - \overline{y}\}^2}{\sum_{i=1}^{n} (Y_i - \overline{y})^2},$$

with $\overline{y} = n^{-1}(Y_1 + \ldots + Y_n)$.

In discrete cases, the reg.fun function for (24) is used with the binomial kernel on milk data as follows. This dataset is about average daily fat (kg/day) yields from milk of a single cow for each of the first 35 weeks.

```
R> data("milk", package = "Ake")
R> x <- milk$week
R> y <- milk$yield
R> h <- reg.fun(x, y, "discrete", "bino", 0.1)
R> h

Bandwidth h:0.1        Coef_det=0.9726

Number of points: 35;        Kernel = Binomial

data             y
Min.   : 1.0   Min.   :0.0100
1st Qu.: 9.5   1st Qu.:0.2750
Median :18.0   Median :0.3600
Mean   :18.0   Mean   :0.3986
3rd Qu.:26.5   3rd Qu.:0.6150
Max.   :35.0   Max.   :0.7200
eval.points        m_n
Min.   : 1.0   Min.   :0.01542
1st Qu.: 9.5   1st Qu.:0.27681
```

Arguments	Description
Vec	The explanatory data sample can be discrete or continuous.
y	The response variable.
ker	The associated kernel.
h	The sequence of bandwidths where to compute the optimal bandwidth.
a0,a1	The bounds of the support of extended beta kernel. Default values are respectively 0 and 1.
a	The arm of the discrete triangular kernel. Default value is 1.
c	The number of categories in DiracDU kernel. Default value is 2.

Results	Description
kernel	The associated kernel.
hcv	The optimal bandwidth obtained by cross-validation.
CV	The values of the cross-validation function.
seqbws	The sequence of bandwidths used to compute hcv.

Table 3: Summary of arguments and results of hcvreg.fun.

```
Median :18.0   Median :0.35065
Mean   :18.0   Mean   :0.39777
3rd Qu.:26.5   3rd Qu.:0.60942
Max.   :35.0   Max.   :0.70064
```

The above reg.fun is also used for continuous cases; see Figure 3 and Table 4 for the motorcycle impact data of Silverman (1985).

Bandwidth selection

We present two bandwidth selection methods for the regression: the well-known cross-validation for any associated kernel and the Bayesian global for the binomial kernel.

Cross-validation for any associated kernel

For a given associated kernel, the optimal bandwidth parameter is $\widehat{h}_{cv} = \arg\min_{h>0} \mathrm{LSCV}(h)$ with

$$\mathrm{LSCV}(h) = \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i - \widehat{m}_{-i}(X_i) \right\}^2, \tag{25}$$

where $\widetilde{m}_{-i}(X_i)$ is computed as $\widehat{m}_n$ of (24) excluding X_i; see, e.g., Kokonendji et al. (2009). The hcvreg.fun function to compute this optimal bandwidth is described in Table 3.

The following code helps to compute the bandwidth parameter by cross-validation on milk data. The associated kernel used is the discrete triangular kernel with arm $a = 1$.

```
R> data("milk", package = "Ake")
R> x <- milk$week
R> y <- milk$yield
R> f <- hcvreg.fun(x, y, type_data = "discrete", ker = "triang", a = 1)
R> f$hcv
```

```
[1] 1.141073
```

When we consider the continuous associated kernel, one needs to set the type of data parameter to "continuous" in the hcvreg.fun function. Thus, the hcvreg.fun and reg.fun functions are used with gamma, lognormal, reciprocal inverse Gaussian and Gaussian kernel on the motor cycle impact data described in Silverman (1985). The observations consist of accelerometer reading taken through time in an experimentation on the efficiency of crash helmets. The results in Table 4 agree with the shapes of continuous associated kernels of Part (b) of Figure 1; see also Figure 3. In fact, since the lognormal kernel is well concentrated around the target x, it gives the best R^2 which is 75.9%. The gamma and the reciprocal inverse Gaussian kernels give similar R^2 in the order 73%. Although the Gaussian kernel is well concentrated on the target, it gives the lower result of $R^2 = 70.90\%$. This is mainly due to the symmetry of the kernel which cannot change its shapes according to the target.

	Gamma	Lognormal	Rec. Inv. Gaussian	Gaussian
R^2	0.7320	0.7591	0.7328	0.7090

Table 4: Some expected values of R^2 of nonparametric regressions of the motor cycle impact data (Silverman, 1985) by some continuous associated kernels.
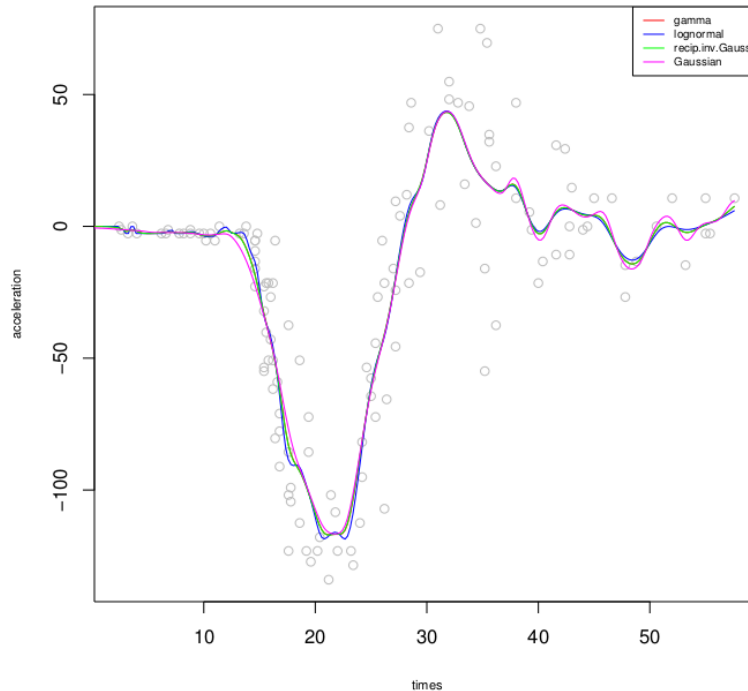
Figure 3: Nonparametric regressions of the motors cycle impact data (Silverman, 1985) by some continuous associated kernels.

Bayesian global for binomial kernel

Using Bayes theorem, the joint posterior distribution of h given the observations is

$$\pi(h|X_1, X_2, \ldots, X_n) \propto h^{\alpha-1}(1-h)^{\beta-1}\left(\frac{1}{2}\sum_{i=1}^{n}\{y_i - \widehat{m}_{-i}(X_i)\}^2 + b\right)^{-(n+2a)/2},$$

where $\propto$ denotes proportional, the reals a and b are the parameters of the inverse gamma distribution $\texttt{IG}(a,b)$, and α and β those of the beta distribution $\mathcal{B}e(\alpha, \beta)$. The estimate $\widehat{h}_{bay}$ of the smoothing parameter h is given by Markov chain Monte Carlo (MCMC) techniques with Gibbs sampling:

$$\widehat{h}_{bay} = \frac{1}{N - N_0}\sum_{N_0+1}^{N} h^{(t)},$$

where N_0 is the burn-in period and N the number of iterations; see Zougab et al. (2014b) for further details. It will be implemented in a future version of the **Ake** package.

Summary and final remarks

The **Ake** package offers easy tools for R users whose research involves kernel estimation of density functions and/or regression functions through associated kernels that are capable of handling all categorical, count and real positive datasets. Figure 1 shows the importance of the associated kernel choice as well as the bandwidth selection. In fact, symmetric (e.g., Gausssian) kernel estimators (respectively empirical estimators) are not suitable for bounded or positive continuous datasets (respectively discrete small samples). We then need an appropriate associated kernel. The binomial kernel is suitable for small size count data while the discrete triangular or the naive kernel are more indicated for large sample sizes. In continuous cases, the lognormal and gamma kernels give the best

estimation for positive data while the extended beta is suitable for any compact support.

This package includes various continuous and discrete associated kernels. It also contains functions to handle the bandwidth selection problems through cross-validation, local and global Bayesian procedures for binomial kernel and also the adaptive Bayesian procedure for the gamma kernel. In general, Bayesian choices of smoothing parameters will be better than their cross-validation counterparts. Future versions of the package will contain Bayesian methods with other associated kernels. Also, these associated kernels are useful for heavy tailed data p.d.f. estimation and can be added later in the package; see, e.g., Ziane et al. (2015). The case of multivariate data needs to be taken in consideration; see Kokonendji and Somé (2015) for p.d.f. estimation and Somé and Kokonendji (2016) for regression. We think that the **Ake** package can be of interest to nonparametric practitioners of different applied settings.

Acknowledgments

Bibliography

I. S. Abramson. On bandwidth variation in kernel estimates – A square root law. *The Annals of Statistics*, 10(4):1217–1223, 1982. doi: 10.1214/aos/1176345986. [p267]

J. Aitchison and C. G. Aitken. Multivariate binary discrimination by the kernel method. *Biometrika*, 63 (3):413–420, 1976. doi: 10.2307/2335719. [p258, 260, 271]

A. Azzalini and A. W. Bowman. A look at some data on the old faithful geyser. *Applied Statistics*, 39(3): 357–365, 1990. doi: 10.2307/2347385. [p265, 266]

A. W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360, 1984. doi: 10.1093/biomet/71.2.353. [p265]

L. Breiman, W. Meisel, and E. Purcell. Variable kernel estimates of multivariate densities. *Technometrics*, 19(2):135–144, 1977. doi: 10.2307/1268623. [p267]

L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and Regression Trees*. CRC Press, 1984. [p271]

M. J. Brewer. A Bayesian model for local smoothing in kernel density estimation. *Statistics and Computing*, 10(4):299–309, 2000. doi: 10.1023/a:1008925425102. [p267]

S. X. Chen. Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2): 131–145, 1999. doi: 10.1016/s0167-9473(99)00010-9. [p260]

S. X. Chen. Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, 52(3):471–480, 2000. doi: 10.1023/a:1004165218295. [p260, 269]

T. Duong. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, 21(7):1–16, 2007. doi: 10.18637/jss.v021.i07. [p258]

V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158, 1969. doi: 10.1137/1114019. [p261]

G. L. Geoffroy. Synthesis, molecular dynamics, and reactivity of mixed-metal clusters. *Accounts of Chemical Research*, 13(12):469–476, 1980. doi: 10.1021/ar50156a006. [p270]

W. Härdle. *Smoothing Techniques: With Implementation in S*. Springer-Verlag, 2012. doi: 10.1007/978-1-4612-4432-5. [p265]

T. Hayfield and J. S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5):1–32, 2008. doi: 10.18637/jss.v027.i05. [p258, 271]

T. Hayfield and J. S. Racine. *np: Nonparametric Kernel Smoothing Methods for Mixed Datatypes*, 2014. URL https://CRAN.R-project.org/package=np. R package version 0.60-2. [p271]

G. Igarashi and Y. Kakizawa. Bias corrections for some asymmetric kernel estimators. *Journal of Statistical Planning and Inference*, 159:37–63, 2015. doi: 10.1016/j.jspi.2014.11.003. [p260, 261]

X. Jin and J. Kawczak. Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, 4:103–124, 2003. [p260]

C. C. Kokonendji and T. Senga Kiessé. Discrete associated kernels method and extensions. *Statistical Methodology*, 8(6):497–516, 2011. doi: 10.1016/j.stamet.2011.07.002. [p258, 259, 260, 262, 263]

C. C. Kokonendji and S. M. Somé. On multivariate associated kernels for smoothing some density function. *arXiv:1502.01173*, 2015. URL https://arxiv.org/abs/1502.01173. [p274]

C. C. Kokonendji and S. S. Zocchi. Extensions of discrete triangular distributions and boundary bias in kernel estimation for discrete functions. *Statistics & Probability Letters*, 80(21):1655–1662, 2010. doi: 10.1016/j.spl.2010.07.008. [p260]

C. C. Kokonendji, T. Senga Kiessé, and S. S. Zocchi. Discrete triangular distributions and non-parametric estimation for probability mass function. *Journal of Nonparametric Statistics*, 19(6-8): 241–254, 2007. doi: 10.1080/10485250701733747. [p260]

C. C. Kokonendji, T. Senga Kiessé, and C. G. B. Demétrio. Appropriate kernel regression on a count explanatory variable and applications. *Advances and Applications in Statistics*, 12(1):99–125, 2009. doi: 10.1007/s00180-015-0627-1. [p271, 272]

J.-P. Lecoutre. Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statistics & Probability Letters*, 10(2):145–149, 1990. doi: 10.1016/0167-7152(90)90010-5. [p270]

F. G. Libengué. *Méthode Non-Paramétrique des Noyaux Associés Mixtes et Applications*. PhD thesis, Besançon, 2013. [p258, 259, 260, 261, 262, 263, 264]

J. Marron. A comparison of cross-validation techniques in density estimation. *The Annals of Statistics*, pages 152–162, 1987. doi: 10.1214/aos/1176350258. [p265]

E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964. doi: 10.1137/1109020. [p271]

E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962. doi: 10.1214/aoms/1177704472. [p271]

J. Racine and Q. Li. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1):99–130, 2004. doi: 10.1016/s0304-4076(03)00157-x. [p260]

C. H. Reinsch. Smoothing by spline functions. *Numerische Mathematik*, 10(3):177–183, 1967. doi: 10.1007/bf02162161. [p270]

M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956. doi: 10.1214/aoms/1177728190. [p271]

S. R. Sain. Multivariate locally adaptive density estimation. *Computational Statistics & Data Analysis*, 39 (2):165–186, 2002. doi: 10.1016/s0167-9473(01)00053-6. [p263]

S. R. Sain and D. W. Scott. On locally adaptive density estimation. *Journal of the American Statistical Association*, 91(436):1525–1534, 1996. doi: 10.2307/2291578. [p267]

O. Scaillet. Density estimation using inverse and reciprocal inverse Gaussian kernels. *Nonparametric Statistics*, 16(1–2):217–226, 2004. doi: 10.1080/10485250310001624819. [p261]

B. W. Silverman. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(1):1–52, 1985. doi: 10.2307/2982831. [p270, 272, 273]

S. M. Somé and C. C. Kokonendji. Effects of associated kernels in nonparametric multiple regressions. *Journal of Statistical Theory and Practice*, 10(2):456–471, 2016. doi: 10.1080/15598608.2016.1160010. [p274]

J. W. Tukey. Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics*, 3(2):191–219, 1961. doi: 10.1080/00401706.1961.10489940. [p270]

G. Wahba. *Spline Models for Observational Data*, volume 59. SIAM, 1990. doi: 10.1137/1.9781611970128. [p270]

G. S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372, 1964. [p271]

S. Zhang. A note on the performance of the gamma kernel estimators at the boundary. *Statistics & Probability Letters*, 80(7):548–557, 2010. doi: 10.1016/j.spl.2009.12.009. [p264]

S. Zhang and R. J. Karunamuni. Boundary performance of the beta kernel estimators. *Journal of Nonparametric Statistics*, 22(1):81–104, 2010. doi: 10.1080/10485250903124984. [p264]

X. Zhang, M. L. King, and R. J. Hyndman. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics & Data Analysis*, 50(11):3009–3031, 2006. doi: 10.1016/j.csda.2005.06.019. [p267]

X. Zhang, M. L. King, and H. L. Shang. Bayesian bandwidth selection for a nonparametric regression model with mixed types of regressors. *Econometrics*, 4(2):24, 2016. doi: 10.3390/econometrics4020024. [p267]

Y. Ziane, S. Adjabi, and N. Zougab. Adaptive Bayesian bandwidth selection in asymmetric kernel density estimation for nonnegative heavy-tailed data. *Journal of Applied Statistics*, 42(8):1645–1658, 2015. doi: 10.1080/02664763.2015.1004626. [p258, 274]

N. Zougab, S. Adjabi, and C. Kokonendji. Binomial kernel and Bayes local bandwidth in discrete function estimation. *Journal of Nonparametric Statistics*, 24(3):783–795, 2012. doi: 10.1080/10485252.2012.678847. [p258, 259, 267, 268]

N. Zougab, S. Adjabi, and C. C. Kokonendji. A Bayesian approach to bandwidth selection in univariate associate kernel estimation. *Journal of Statistical Theory and Practice*, 7(1):8–23, 2013. doi: 10.1080/15598608.2013.756286. [p267, 268]

N. Zougab, S. Adjabi, and C. C. Kokonendji. Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. *Computational Statistics & Data Analysis*, 75:28–38, 2014a. doi: 10.1016/j.csda.2014.02.002. [p267, 268]

N. Zougab, S. Adjabi, and C. C. Kokonendji. Bayesian approach in nonparametric count regression with binomial kernel. *Communications in Statistics – Simulation and Computation*, 43(5):1052–1063, 2014b. doi: 10.1080/03610918.2012.725145. [p268, 273]

Wanbitching E. Wansouwé
The University of Maroua
Higher Teachers'Training College
Department of Computer Science
P.O. Box 55 Maroua, Cameroon
ericwansouwe@gmail.com

Sobom M. Somé
Université Bourgogne Franche-Comté
Laboratoire de Mathématiques de Besançon
16 route de Gray, 25030 Besançon cedex, France
sobom.some@univ-fcomte.fr

Célestin C. Kokonendji
Université Bourgogne Franche-Comté
Laboratoire de Mathématiques de Besançon
16 route de Gray, 25030 Besançon cedex, France
celestin.kokonendji@univ-fcomte.fr