Bayesian estimation of bandwidth in semiparametric kernel estimation of unknown probability mass and regression functions of count data

Tristan Senga Kiessé, Nabil Zougab, Célestin Kokonendji

Computational Statistics (revised version) manuscript No.
(will be inserted by the editor)

Bayesian estimation of bandwidth in semiparametric kernel estimation of unknown probability mass and regression functions of count data

Tristan Senga Kiessé · Nabil Zougab · Célestin C. Kokonendji

Received: date / Accepted: date

Abstract This work takes advantage of semiparametric modelling which improves significantly in many situations the estimation accuracy of the purely nonparametric approach. Herein for semiparametric estimations of probability mass function (pmf) of count data, and an unknown count regression function (crf), the kernel used is a binomial one and the bandiwdth selection is investigated by developing Bayesian approaches. About the latter, Bayes local and global bandwidth approaches are used to establish data-driven selection procedures in semiparametric framework. From conjugate beta prior distributions of the smoothing parameter and under the squared errors loss function, Bayes estimate for pmf is obtained in closed form. This is not available for the crf which is computed by the Markov Chain Monte Carlo technique. Simulation studies demonstrate that both proposed methods perform better than the classical cross-validation procedures, in particular the smoothing quality and execution times are optimized. All applications are made on real data sets.

Keywords Count regression function · Cross-validation · Discrete associated kernel · MCMC · Probability mass function

1 Introduction

The use of a discrete kernel is more suitable than the use of a continuous kernel for both estimations of the probability mass function (pmf) of a discrete variable, and of an unknown regression function of discrete explanatory variables. For small and moderate sample sizes, the discrete kernel approach needs a non-naive discrete kernel which is called *discrete associated kernel* as detailed in Kokonendji and

University of Nantes Angers Le Mans, GeM UMR 6183 CNRS
Chair of civil engineering and eco-construction, Saint-Nazaire, France
E-mail: tristan.sengakiesse@univ-nantes.fr
· University of Tizi-Ouzou and LAMOS Bejaia, Algeria
E-mail: nabilzougab@yahoo.fr
· University of Franche-Comté, LMB UMR 6623 CNRS-UFC, France
E-mail: celestin.kokonendji@univ-fcomte.fr

Senga Kiessé (2011). For a fixed point x in a discrete support $\mathbb{T}$ and a bandwidth parameter $h > 0$, a discrete associated kernel $K_{x,h}(\cdot)$ is defined to be a pmf with support $\mathbb{S}_x$ such that

$$\begin{aligned} x \in \mathbb{S}_x \quad & (A1), \\ \lim_{h\to 0} \mathbb{E}(\mathcal{K}_{x,h}) = x \quad & (A2), \\ \lim_{h\to 0} \mathbb{V}(\mathcal{K}_{x,h}) = 0 \quad & (A3), \end{aligned}$$

where $\mathcal{K}_{x,h}$ is the discrete random variable of pmf $K_{x,h}(\cdot)$. These conditions are gathered to obtain good asymptotic behaviours (similar to the Dirac or naive discrete kernel) of the corresponding estimators, as developped in Kokonendji and Zocchi (2010) for the family of discrete triangular kernels. However, for small and moderate sample sizes, the class of the so-called *standard discrete kernels*, satisfying only (A1) and (A4) - (A5) below, works also very well:

$$\begin{aligned} \mathbb{E}(\mathcal{K}_{x,h}) = x + h \quad & (A4), \\ \lim_{h\to 0} \mathbb{V}(\mathcal{K}_{x,h}) \in \mathcal{V}(0) \quad & (A5), \end{aligned}$$

with $\mathcal{V}(0)$ a set in the neighborhood of zero (Kokonendji and Senga Kiessé, 2011). Amongst standard discrete kernels, the most interesting is the *binomial kernel* denoted $B_{x,h}$ on the support $\mathbb{S}_x = \{0, 1, \ldots, x+1\}$ and with the bandwidth parameter in the interval $[0, 1]$:

$$B_{x,h}(y) = \frac{(x+1)!}{y!(x+1-y)!} \left(\frac{x+h}{x+1}\right)^y \left(\frac{1-h}{x+1}\right)^{x+1-y} \mathbb{1}_{\mathbb{S}_x}(y), \tag{1}$$

with $\mathbb{1}_A$ denoted the indicator function of any given event $A \subset \mathbb{T}$. Note that $B_{x,h}$ follows the binomial distribution $\mathscr{B}(x+1; (x+h)/(x+1))$ with a number of trials $x+1$ and a success probability in each trial $(x+h)/(x+1)$. In addition to discrete kernel choice, one of the main issues of the discrete kernel method is the bandwidth selection procedure. The classical cross-validation procedure is used for conducting bandwith choice in most of the works about discrete kernel estimations. That concerns the semiparametric estimation of count regression functions (crf) as in Abdous *et al.* (2012) and Cuny and Senga Kiessé (2014), nonparametric estimation of pmf in Kokonendji and Senga Kiessé (2011), semiparametric estimation of pmf in Kokonendji *et al.* (2009ab) and, Senga Kiessé and Cuny (2014).

Recently, Bayesian approaches using binomial kernel in equation (1) have been proposed as an alternative to cross-validation procedure for bandwidth selections but only for nonparametric estimations of pmf and crf (Zougab *et al.*, 2012, 2014a). This paper pursues the latter works in semiparametric framework since the semiparametric approach is demonstrated to significantly improve the purely nonparametric modelling in many situations (Abdous *et al.*, 2012). This work aims to take both advantages of semiparametric kernel estimation in comparison with nonparametric kernel estimation of count data and of Bayes approach in comparison with cross-validation procedure for bandwidth selection.

The first procedure in Section 2 is related to a pmf of count data formulated as a weighted Poisson distribution. The second method presented in Section 3 stduies an unknown regression function of count data considered as a weighted parametric regression model. The particular choice of binomial kernel in equation (1) is connected to Bayesian calculations: the bandwidth h will be treated as a parameter with a prior distribution, the posterior densitiy can be obtained via the likelihood cross validation and the prior distribution using Bayes rule. The performances of Bayesian bandwidth selection methods are compared to the cross-validation procedures for each of the estimators studied through simulated data sets in Section 4 and applications on real data sets in Section 5. Finally, Section 6 contains concluding remarks.

2 Poisson-weighted semiparametric binomial kernel estimation

2.1 Estimator and cross-validation procedure

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with an unknown pmf f on the set $\mathbb{T} = \mathbb{N}$ of non-negative integers. Let us assume that any pmf $f(\cdot) := \mathbb{P}(X_1 = \cdot)$ can be represented as a weighted count distribution with a given parametric part $p(\cdot\,;\mu)$ and its corresponding count weight function part $\omega^{[f]}(\cdot\,;\mu)$ (Kokonendji and Pérez-Casany, 2012). We consider the Poisson-weighted semiparametric estimator $\widehat{f}_{n,h}(\cdot) := p(\cdot\,;\widehat{\mu}) \times \widehat{\omega}^{[f]}_{n,h}(\cdot\,;\widehat{\mu})$ of f proposed by Kokonendji *et al.* (2009a) for the binomial kernel as follows:

$$\widehat{f}_{n,h}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(x;\widehat{\mu})}{p(X_i;\widehat{\mu})}\, B_{x,h}(X_i), \quad \forall x \in \mathbb{N}, \tag{2}$$

where $p(x;\mu) = \mu^x e^{-\mu}/x! > 0$ is the pmf of the Poisson distribution with mean parameter $\mu > 0$, $\widehat{\mu} = n^{-1}\sum_{i=1}^{n} X_i$ is the sample mean, and $B_{x,h}(\cdot)$ is the binomial kernel defined in (1) with $h = h(n) \in [0,1]$ an arbitrary sequence of smoothing parameters that fulfills $\lim_{n\to\infty} h(n) = 0$. The continuous version of semiparametric estimator in (2) for probability density function was discussed in Hjort and Glad (1995), Hjort and Jones (2004) or Naito (2004).

Remark 1 From Kokonendji and Senga Kiessé (2011), for purely nonparametric estimator from (2), we have $\widehat{f}_{n,h}(x) \in [0,1]$ for all $x \in \mathbb{N}$ and $\sum_{\mathbf{x}\in\mathbb{N}} \widehat{f}_{n,h}(x) = C_{n,h}$, where $C_{n,h}$ is not necessarly one but a positive and finite constant. In fact, it is easy to check that $C_{n,h} = 1$ for Aitchison-Aitken and Wang-vanRyzin kernels; see Examples 3 and 4 in Kokonendji and Senga Kiessé (2011). In general, we have $C_{n,h} \neq 1$ if we use binomial, Poisson and discrete triangular kernels; see Senga Kiessé (2008, p. 193) and Wansouwé et al. (2015) for numerical results. Thus, a normalization of $\widehat{f}_{n,h}$ is necessary for providing a pmf. Finally, we can investigate the behaviour of $C_{n,h}$ from 1 through the bias and variance of $\widehat{f}_{n,h}$ by

$$\mathbb{E}(C_{n,h}) = 1 + \sum_{x\in\mathbb{N}} Bias\left\{\widehat{f}_{n,h}(x)\right\}$$

and

$$Var(C_{n,h}) = \sum_{x\in\mathbb{N}} Var\left\{\widehat{f}_{n,h}(x)\right\}.$$

If the ratio $p(x;\widehat{\mu})/p(X_i;\widehat{\mu})$ from (2) is equal to 1 for all X_i and $x \in \mathbb{N}$, $\widehat{f}_{n,h}$ corresponds to the nonparametric estimator of pmf f with binomial kernel (Kokonendji and Senga Kiessé, 2011). The Poisson parametric part in (2) is generally retained because of its equidispersion property with respect to over- and under-dispersion phenomenon in the family of count distributions (Kokonendji, 2014).

Then, the classical cross-validation procedure is often used for finding an optimal bandwidth

$$h_{cv} = \arg\min_{h>0} CV_f(h)$$

minimizing the function

$$\begin{aligned} CV_f(h) &= \sum_{x\in\mathbb{N}} \widehat{f}^2_{n,h}(x) - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{n,h;-i}(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{p(X_i;\widehat{\mu})p(X_j;\widehat{\mu})} \sum_{x\in\mathbb{N}} p^2(x;\widehat{\mu})B_{x,h}(X_i)B_{x,h}(X_j) \\ &\quad - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq 1} B_{X_i,h}(X_j)\frac{p(X_i;\widehat{\mu}_{-i})}{p(X_j;\widehat{\mu}_{-i})}. \end{aligned}$$

The function $\widehat{f}_{n,h;-i}(x) = (n-1)^{-1}\sum_{j\neq i}^{n} B_{x,h}(X_j)$ is the leave-one-out kernel estimator of $\widehat{f}_n(x)$ and $\widehat{\mu}_{-i}$ is computed as $\widehat{\mu}$ by excluding X_i. The score function CV_f is an estimator asymptotically unbiased of the term depending on parameter $h > 0$ in the mean integrated squared error (MISE) of estimator $\widehat{f}_{n,h}(x)$.

2.2 Posterior estimate of local bandwidth

In this part we assume that the smoothing parameter $h \in [0,1]$ in (2) is a random quantity with a prior distribution $\pi(\cdot)$ in Bayesian framework. According to Zougab *et al.* (2012), we first define a discrete function f_h given by

$$f_h(x) = \sum_{y\in\mathbb{N}} f(y)B_{x,h}(y) = \mathbb{E}\{B_{x,h}(Y)\}, \ \ \forall x \in \mathbb{N},$$

with $B_{x,h}$ the binomial kernel in (1) and Y a random variable with pmf f. Under the assumptions (A4) - (A5), $f_h(x)$ and also $\widehat{f}_{n,h}(x)$ in equation (2) are close to $f(x)$ as n goes to ∞ and $h = h(n) \to 0$, as studied by Kokonendji and Senga Kiessé (2011) for $f_h(x)$, and Kokonendji *et al.* (2009a) for $\widehat{f}_{n,h}(x)$.

By considering h as a scale parameter for $f_h(x)$, the approach developed here consists of using $f_h(x)$, which is also estimated by $\widehat{f}_{n,h}(x)$ in equation (2), and constructing a local Bayesian estimator $\widehat{h}_n(x)$ for h at the point x. Indeed, let $\pi(h)$ be a prior distribution. Bayes theorem enables to express the posterior of h at the point of estimation x as follows:

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

Hence, an estimate of the posterior $\pi(h|x)$ is given by the posterior density

$$\widehat{\pi}(h|x, X_1, X_2 \ldots, X_n) = \frac{\widehat{f}_{n,h}(x)\pi(h)}{\int \widehat{f}_{n,h}(x)\pi(h)dh}, \tag{3}$$

where $\widehat{f}_{n,h}$ is a semiparametric estimator as in equation (2) of the unknown function f_h. Under the squared error loss, Bayes estimator of the smoothing parameter h is the mean of the previous posterior density given by

$$\widehat{h}_n(x) = \int h\,\widehat{\pi}(h|x, X_1, X_2, \ldots, X_n)\, dh. \tag{4}$$

Now let us apply estimate of h in equation (4) to Poisson-weighted semiparametric binomial estimation in equation (2) for finding an explicit local bandwidth for given $x \in \mathbb{N}$. A natural conjugate prior density of $h \in [0,1]$ is the well-known beta prior distribution with positive parameters α and β given by

$$\pi(h) = \frac{1}{b(\alpha,\beta)} h^{\alpha-1}(1-h)^{\beta-1}\mathbb{1}_{[0,1]}(h), \tag{5}$$

with mean $\alpha/(\alpha+\beta)$ and

$$b(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt, \quad \alpha,\beta > 0.$$

First, let us consider the numerator of $\widehat{\pi}$ in equation (3): from equations (1), (2) and the binomial expansion $(x+h)^{X_i} = \sum_{k=0}^{X_i} \frac{X_i!}{k!(X_i-k)!} x^k h^{X_i-k}$, we have

$$\begin{aligned}
\widehat{f}_{n,h}(x)\pi(h) &= \frac{1}{nb(\alpha,\beta)} \sum_{i=1}^{n} \frac{p(x;\widehat{\mu})}{p(X_i;\widehat{\mu})} B_{x,h}(X_i) \times h^{\alpha-1}(1-h)^{\beta-1} \\
&= \frac{1}{nb(\alpha,\beta)} \sum_{i=1}^{n} \frac{p(x;\widehat{\mu})}{p(X_i;\widehat{\mu})} \frac{(x+1)!}{X_i!(x+1-X_i)!} \left(\frac{x+h}{x+1}\right)^{X_i} \left(\frac{1-h}{x+1}\right)^{x+1-X_i} \\
&\quad \times h^{\alpha-1}(1-h)^{\beta-1} \\
&= \frac{1}{nb(\alpha,\beta)} \sum_{i=1}^{n} \frac{p(x;\widehat{\mu})}{p(X_i;\widehat{\mu})} \frac{(x+1)!}{X_i!(x+1-X_i)!} \times \frac{\sum_{k=0}^{X_i} X_i/\{k!(X_i-k)!\} x^k h^{X_i-k}}{(x+1)^{X_i}} \\
&\quad \times \frac{(1-h)^{x+1-X_i}}{(x+1)^{x+1-X_i}} \times h^{\alpha-1}(1-h)^{\beta-1} \\
&= \frac{1}{nb(\alpha,\beta)} \sum_{i=1}^{n} \sum_{k=0}^{X_i} \frac{p(x;\widehat{\mu})}{p(X_i;\widehat{\mu})} \frac{(x+1)!}{(x+1-X_i)!k!(X_i-k)!} \frac{x^k}{(x+1)^{x+1}} \\
&\quad \times h^{X_i+\alpha-k-1}(1-h)^{x+\beta-X_i}.
\end{aligned}$$

Then, the denominator of $\widehat{\pi}$ in (3) is written by

$$\begin{aligned}
\int \widehat{f}_{n,h}(x)\pi(h)dh &= \frac{1}{nb(\alpha,\beta)} \sum_{i=1}^{n} \sum_{k=0}^{X_i} \frac{p(x;\widehat{\mu})}{p(X_i;\widehat{\mu})} \frac{(x+1)!}{(x+1-X_i)!k!(X_i-k)!} \frac{x^k}{(x+1)^{x+1}} \\
&\quad \times \int h^{X_i+\alpha-k-1}(1-h)^{x+\beta-X_i}dh \\
&= \frac{1}{nb(\alpha,\beta)} \sum_{i=1}^{n} \sum_{k=0}^{X_i} \frac{p(x;\widehat{\mu})}{p(X_i;\widehat{\mu})} \frac{(x+1)!}{(x+1-X_i)!k!(X_i-k)!} \frac{x^k}{(x+1)^{x+1}} \\
&\quad \times b(X_i+\alpha-k, x+\beta-X_i+1).
\end{aligned}$$

Hence, taking the ratio of the previous calculus results, we easily deduce the posterior density:

$$\widehat{\pi}(h|x, X_1, X_2, \ldots, X_n) = \frac{\sum_{i=1}^{n}\sum_{k=0}^{X_i} p(x;\widehat{\mu})x^k/\{(x+1-X_i)!k!(X_i-k)!p(X_i;\widehat{\mu})\}}{\sum_{i=1}^{n}\sum_{k=0}^{X_i} p(x;\widehat{\mu})x^k/\{(x+1-X_i)!k!(X_i-k)!p(X_i;\widehat{\mu})\}} \times \frac{h^{X_i+\alpha-k-1}(1-h)^{x+\beta-X_i}}{b(X_i+\alpha-k, x+\beta-X_i+1)}.$$

Therefore, the local bandwidth $\widehat{h}_n(x)$ of (4) is expressed as

$$\widehat{h}_n(x) = \frac{\sum_{i=1}^{n}\sum_{k=0}^{X_i} p(x;\widehat{\mu})x^k/\{(x+1-X_i)!k!(X_i-k)!p(X_i;\widehat{\mu})\}}{\sum_{i=1}^{n}\sum_{k=0}^{X_i} p(x;\widehat{\mu})x^k/\{(x+1-X_i)!k!(X_i-k)!p(X_i;\widehat{\mu})\}} \times \frac{b(X_i+\alpha-k+1, x+\beta-X_i+1)}{b(X_i+\alpha-k, x+\beta-X_i+1)}, \tag{6}$$

with $X_i \leq x+1$.

The following theorem shows the convergence of $\widehat{h}_n(x)$ to zero as $n \to \infty$ for judicious choices of prior parameters α and β.

Theorem 1 *For given $x \in \mathbb{N}$, Bayesian bandwidth $\widehat{h}_n(x)$ in (6) converges to zero almost surely as $n \to \infty$ for prior parameter sequences satisfying $\alpha > 0$ and $\beta = \beta_n$ with $\lim_{n\to\infty} \beta_n = \infty$.*

Proof Following the proof of Theorem 2.1 in Zougab *et al.*(2012), we can easily show that

$$\frac{\alpha}{\beta_n+\alpha+x+1} \leq \widehat{h}_n(x) \leq \frac{x+\alpha+1}{\beta_n}, \quad \forall x \in \mathbb{N}.$$

Therefore, we deduce that $\widehat{h}_n(x)$ almost surely goes to 0 when $\beta_n \to \infty$. □

Remark 2 The choice of β depending on sample size n is important for the consistency of the Poisson-weighted semiparametric binomial kernel estimator and for its smoothing quality. Based on the proof of previous theorem, we can control the rate of convergence of the bandwidth $\widehat{h}_n(x)$ as the sample size increases for judicious choices of $\beta = \beta_n$. We can refer to Zougab *et al.* (2012, Remark 3).

3 Semiparametric binomial kernel estimator of regression function

This section is concerned with the global Bayesian bandwidth selection procedure h in a semiparametric count regression estimation using binomial kernel in equation (1). The estimation of the unknown parameter σ^2 of the variance Gaussian error is also investigated. Note that the Bayesian bandwidth selection approach in kernel regression estimation has been far less investigated in comparison with kernel density estimation. To our knowledge, two works have proposed the Bayseian bandwidth procedure in purely nonparametric kernel regression estimation using the MCMC sampling algorithm; see Zhang *et al.* (2009) for continuous case using the Gaussian kernel and Zougab *et al.* (2012) for discrete case using the binomial kernel.

3.1 Estimator and cross-validation method

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a sequence of i.i.d. random variables defined on $\mathbb{T} \times \mathbb{R}$ and such that

$$Y_i = m(X_i) + \epsilon_i, \tag{7}$$

where $m(\cdot) := \mathbb{E}(Y_1|X_1 = \cdot)$ is an unknown regression function of count data and the ϵ_i's are assumed to have zero mean and finite variance $\sigma^2 > 0$. According to Abdous *et al.* (2012), the semiparametric estimator of m is defined by:

$$\widehat{m}_{n,h}(x) = r(x; \widehat{\theta}) \times \sum_{i=1}^{n} \frac{Y_i B_{x,h}(X_i)}{r(X_i; \widehat{\theta}) \sum_{j=1}^{n} B_{x,h}(X_j)}, \;\; \forall x \in \mathbb{T} = \mathbb{N}, \tag{8}$$

where $r(\cdot\,; \theta)$ is a parametric count function that depends on unknown parameter $\theta = (\theta_1, \ldots, \theta_s)^\top$, $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_s)^\top$ is an estimate of θ constructed in the previous step (by generalized least squared method for example), $B_{x,h}$ is the binomial kernel in equation (1) with $h = h(n) \in [0, 1]$ an arbitrary sequence of smoothing parameters that fulfills $\lim_{n\to\infty} h(n) = 0$. In practice, when X is a binary variable or observations of X are count data, generalized linear models (GLM) studied by McCullagh and Nelder (1989) for these cases can serve as parametric start regression models $r(\cdot; \theta)$. Then, the estimation accuracy of chosen parametric regression model can be improved by the nonparametric correction term which is the second factor in the right side of equality in equation (8). One can refer to Glad (1998) and Fan *et al.* (2009) as both related references for continuous version of estimator in equation (8).

Let us remark that if ratio $r(x; \widehat{\theta})/r(X_i; \widehat{\theta})$ in equation (8) is equal to 1 for all X_i and $x \in \mathbb{N}$, $\widehat{m}_{n,h}$ is the nonparametric estimator of m with binomial kernel. The cross-validation procedure for bandwidth selection is adapted to semiparametric count regression estimator by Cuny and Senga Kiessé (2014). Similar to semiparametric estimation of pmf, it consists of finding an optimal value h_{cv} by minimizing the score function

$$CV_m(h) = \frac{1}{n} \sum_{i=1}^{n} \{Y_i - \widehat{m}_{n,h;-i}(X_i)\}^2,$$

where $\widehat{m}_{n,h;-i}$ is computed as $\widehat{m}_{n,h}$ of (8) by excluding X_i.

3.2 Bayesian bandwidth estimation

3.2.1 Likelihood and posterior

Let $(x_i, y_i), i = 1, 2, ..., n$ be i.i.d. bivariate observations and let ϵ_i be i.i.d. residuals following the Gaussian distribution $\mathcal{N}(0, \sigma^2)$ with mean zero and constant variance σ^2. Thus, the model in equation (7) can be expressed as

$$Y_i - m(X_i) \sim \mathcal{N}(0, \sigma^2).$$

Then, treating h and σ^2 as parameters, the likelihood cross-validation function of the observed data is

$$\mathrm{LCV}(y_1, y_2, \ldots, y_n | h, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\{y_i - \widehat{m}_{n,h;-i}(x_i)\}^2\right],$$

with $\widehat{m}_{n,h;-i}(\cdot)$ the leave-one-out binomial kernel estimator of $\widehat{m}_{n,h}(x)$ computed by excluding X_i.

Concerning the prior distributions of σ^2 and h, $\pi(h)$ is already defined in equation (5) and $\pi(\sigma^2)$ is assumed to be an inverted Gamma with positive parameters a and b denoted as $\mathrm{IG}(a, b)$ with its density function being

$$\pi(\sigma^2) = \frac{b^a}{\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left(-\frac{1}{\sigma^2}\right) \mathbb{1}_{[0,\infty)}(\sigma^2),$$

with

$$\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)dt, \quad a > 0.$$

Thus, the joint posterior density function of (h, σ^2) given the data is

$$\begin{aligned}\pi(h, \sigma^2 | y_1, y_2, \ldots, y_n) &\propto \frac{\pi(h)\pi(\sigma^2)}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\{y_i - \widehat{m}_{n,h;-i}(x_i)\}^2\right] \\ &\propto h^{\alpha-1}(1-h)^{\beta-1}\left(\frac{1}{\sigma^2}\right)^{(n+2a)/2+1} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}\{y_i - \widehat{m}_{n,h;-i}(x_i)\}^2 + 2b\right]\right),\end{aligned}$$

where $\propto$ denotes proportional. It ensues that the posterior density function of σ^2 given h and data is

$$\sigma^2 | h, y_1, y_2, \ldots, y_n \sim \mathrm{IG}\left(\frac{n+2a}{2}, \frac{1}{2}\sum_{i=1}^{n}\{y_i - \widehat{m}_{n,h;-i}(x_i)\}^2 + b\right).$$

Also, the posterior density function of h given data is derived from the expression $\pi(h, \sigma^2 | y_1, y_2, \ldots, y_n)$ by integrating out σ^2 as follows:

$$\pi(h | y_1, y_2, \ldots, y_n) \propto h^{\alpha-1}(1-h)^{\beta-1}\left[\frac{1}{2}\sum_{i=1}^{n}\{y_i - \widehat{m}_{n,h;-i}(x_i)\}^2 + b\right]^{-(n+2a)/2}. \quad (9)$$

Since the posterior distribution (9) cannot be simulated directly, the next subsection sets up the MCMC method and Gibbs sampling procedure to generate samples from h and to simulate the Gaussian error σ^2, respectievly. Thus, the estimates of h and σ^2 are provided by their ergodic averages.

3.2.2 An MCMC algorithm

Following Zougab *et al.* (2014a) for purely nonparametric count regression, we pursue in the same manner for the semiparametric count regression function $\widehat{m}_{n,h;-i}$ in equation (9). Indeed, we use the MCMC sampling algorithm such as the random-walk Metropolis algorithm and the Gibbs sampling procedure to generate draws of h and σ^2 using an arbitrary initial value $h^{(0)}$ in $(0, 1)$. After a burn-in period N_0 and a total number of iterations N sufficiently large, the Markov chain converges to the interest density. The iterations N_0 are not used in the computing of the estimatros. Then, the estimators of the bandwidth h and the variance Gausian error σ^2 are given as follows:

$$\widehat{h} = \frac{1}{N - N_0} \sum_{t=N_0+1}^{N} h^{(\mathbf{t})} \quad \text{and} \quad \widehat{\sigma^2} = \frac{1}{N - N_0} \sum_{t=N_0+1}^{N} \{\sigma^2\}^{(\mathbf{t})}.$$

The random-walk Metropolis-Hastings algorithm is based on a new proposing point using candidate (or proposal) distribution $q(\widetilde{h}|h) = q(\tilde{h} - h)$. In this work, we propose again to use the symmetric candidate distribution such as

$$q(\tilde{h} - h) = \frac{1}{\tau\sqrt{2\pi}} \exp\left\{ \frac{(\tilde{h} - h)^2}{2\tau^2} \right\}.$$

The value of τ will be chosen for obtaining an acceptance rate which is close to 0.5 (Gelman *et al.*, 1996). A sequence of draws from the random-walk Metropolis-Hastings algorithm and Gibbs sampling procedure is obtained as follows:

Step 1. initialize $h^{(0)} \in (0, 1)$
Step 2. for $t \in \{1, \cdots, N\}$
(a) generate $\tilde{h} \sim \mathscr{N}(h^{(t)}, t^2)$
(b) $\tilde{h} = |\tilde{h}|$ and if $\tilde{h} > 1$, return to (a)
(c) calculate the acceptance probability $\rho = \min\{1, \pi(\tilde{h}|y_1, y_2, \cdots, y_n)/\pi(h^t|y_1, y_2, \cdots, y_n)\}$

$$h^{(t+1)} = \begin{cases} \tilde{h}, & \text{if } \; u < \rho, u \sim \mathscr{U}_{[0,1]} \\ h^{(t)}, & \text{otherwise} \end{cases}$$

(d) generate $\{\sigma^2\}^{(t)}$ from $\mathrm{IG}\big((n + 2a)/2, (1/2) \sum_{i=1}^{n} \{y_i - \widehat{m}_{n,h;-i}(x_i)\}^2 + b\big)$
Step 3. $t = t + 1$ and goes to 2.

In order to avoid numerical underflow in practice, we modified the acceptance probability ρ as follows (see Brewer, 1998):

$$\hat{\rho} = \begin{cases} \min\{\gamma, 0\} & \text{if } \pi(h^{(t)}|y_1, y_2, \ldots, y_n) > 0 \\ 0 & \text{if } \pi(h^{(t)}|y_1, y_2, \ldots, y_n) = 0, \end{cases}$$

with

$$\gamma = \log\{\pi(\tilde{h}|y_1, y_2, \ldots, y_n)\} - \log\{\pi(h^{(t)}|y_1, y_2, \ldots, y_n)\}.$$

4 Monte Carlo simulation studies

This section gives two parts of simulation studies of proposed bandwidth selections in semiparametric estimations by using binomial kernel in equation (1). The first part treats three models for the Poisson-weighted semiparametric estimation and the second part presents one model for semiparametric count regression estimation. For Bayesian approaches, we do not need any sensitivity analysis because we will fix the beta prior parameters $\alpha = \alpha_0$ and $\beta = \sqrt{n}$. The choice of β which depends on the sample size n is important for the consistency of both semiparametric kernel estimations of pmf and regression function of count data. Further, this choice enables the convergence of Bayesian bandwidth estimator to zero with the same rate as that of the MISE optimal bandwidth (see Remark 2 of Section 2 and Zougab *et al.*, 2012, 2014a).

4.1 Poisson-weighted semiparametric binomial estimation

In this subsection, we evaluate the performance of Bayes local approach and the classical cross-validation technique for bandwidth choice in Poisson-weighted semiparametric binomial estimation. Simulated count data from Poisson models with parameters $\mu = 2$, $\mu = 5$ and $\mu = 8$ are used. The sample sizes $n = 10$, 25, 50, 100, 200 and 500 and the number of replications $N_{sim} = 100$ are considered for this study. Thus, the performance of $\widehat{f}_{n,h}(x)$ is evaluated by using the Integrated Squared Error (ISE):

$$\text{ISE} = \sum_{x\in\mathbb{N}} \left\{\widehat{f}_{n,h}(x) - f(x)\right\}^2,$$

where $\widehat{f}_{n,h}$ is the global or the local semiparametric binomial kernel estimator.

Table 1 reports the average ISE denoted by $\overline{\text{ISE}}$. Note that the results from Bayesian local approach were obtained using the conjugate beta prior with parameters values $\alpha = 0.5$ and $\beta = \beta_n = \sqrt{n}$ as discussed in Remark 2 of Section 2. The execution times are also given in Table 2. From Tables 1 and 2, we can see that:

i) the values $\overline{\text{ISE}}$ decrease as sample size increases for the two methods;
ii) Bayesian local approach works better than the cross-validation method in term of $\overline{\text{ISE}}$; and,
iii) Bayesian local approach performs better than the cross-validation in the sense of Central Processing Unit (CPU) times.

4.2 Semiparametric binomial kernel regression

Now we investigate the performance of the proposed Bayesian estimator to bandwidth selection in semiparametric binomial kernel regression via a simulation study. We consider bivariate observations $(x_i, y_i)_{i=1,\cdots,n}$ such that

$$y_i = m(x_i) + \epsilon_i, \quad \text{where} \quad m(x) = \frac{2^x}{x!}, \quad x \in \mathbb{N}. \tag{10}$$

Table 1 Some expected values of $\overline{\text{ISE}}$ and their standard errors in parentheses based on 100 replications for the Poisson models with parameter $\mu = 2$, 5 and 8.

f	n	$\overline{\text{ISE}}_{Bayes}$	$\overline{\text{ISE}}_{cv}$
	10	0.0232(0.0173)	0.0410(0.0246)
	25	0.0100(0.0063)	0.0178(0.0109)
Poisson($\mu = 2$)	50	0.0054(0.0043)	0.0090(0.0072)
	100	0.0027(0.0018)	0.0044(0.0030)
	200	0.0015(0.0011)	0.0022(0.0012)
	500	0.0005(0.0004)	0.0009(0.0006)
	10	0.0188(0.0149)	0.0297(0.0175)
	25	0.0097(0.0066)	0.0155(0.0088)
Poisson($\mu = 5$)	50	0.0045(0.0028)	0.0070(0.0037)
	100	0.0027(0.0020)	0.0038(0.0021)
	200	0.0011(0.0006)	0.0017(0.0009)
	500	0.0005(0.0003)	0.0007(0.0004)
	10	0.0202(0.0113)	0.0344(0.0217)
	25	0.0093(0.0113)	0.0143(0.0096)
Poisson($\mu = 8$)	50	0.0046(0.0029)	0.0063(0.0033)
	100	0.0025(0.0029)	0.0039(0.0023)
	200	0.0011(0.0007)	0.0018(0.0009)
	500	0.0004(0.0002)	0.0007(0.0003)

Table 2 Comparison of execution times (in seconds) for one replication.

f	n	t_{Bayes}	t_{cv}
	10	0.03	0.29
	25	0.12	0.36
Poisson($\mu = 8$)	50	0.21	0.81
	100	0.48	1.43
	200	1.10	2.71
	500	2.70	12.99

We simulate y_i using the model in equation (10) and the Gaussian model error with zero mean and fixed variance, i.e., $\epsilon_i \sim \mathscr{N}(0, \sigma^2)$ with $\sigma^2 = 0.1^2$. We also consider the small and moderate sample sizes $n \in \{10, 25, 50, 100, 200\}$. Our aim is to estimate $m(\cdot)$ using the semiparametric regression model given in equation (8) and the variance of Gaussian model error σ^2. Note that we used a GLM (McCullagh and Nelder, 1989) as start function in equation (8) (Abdous *et al.*, 2012). The used GLM represents a normal model for the response variable y_i with a logarithmic link and it is given by

$$y_i = \theta_1 + \theta_2 x_i + \theta_3 \log x_i + e_i, \quad x_i \in \mathbb{N}.$$

The estimators $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)^\top$ of $\theta = (\theta_1, \theta_2, \theta_3)^\top$ can be obtained by the generalized least squared method. However, the performance of the discrete semiparametric estimator $\widehat{m}_{n,h}(\cdot)$ depends crucially on the choice of bandwidth h.

Thus, we apply the MCMC and Gibbs sampler algorithm (given in Section 3) based on 5,000 iterations of the burn-in period and 15,000 as the total number of iterations for deriving global Bayesian bandwidth and estimating the variance model error σ^2. The performance of Bayesian bandwidth estimator depends on the choice of parameters of beta prior (5) with mean $\alpha/(\alpha + \beta)$. Hence, we propose

$\alpha = 1.5 > 0$ and $\beta = \beta_n = \sqrt{n}$. Under this latter choice, we can observe that the prior of bandwidth h is concentrated at 0 for large sample size n. This choice of β which depends of the sample size n is necessary for the consistency of semi-parametric binomial kernel regression estimator in equation (8). Furthermore, the parameters for inverted Gamma prior of σ^2 are fixed as $a = 1$ and $b = 0.05$. For such choice, one can also refer to Kim *et al.* (1998) and Zhang *et al.* (2014).

4.2.1 MCMC convergence

The convergence of the MCMC method is examined by the Batch-Mean Standard Error (BMSE) and the Simulation Inefficiency Factor (SIF). These indicators have been intensively employed in the literature (see Zhang *et al.*, 2009, 2014, Hu *et al.*, 2012, or Zougab *et al.*, 2014a). To illustrate the mixing performance of Bayesian MCMC technique, we present in Tables 3 and 4 the values of BMSE and SIF indicators. The obtained results for each simulated parameter h and σ^2 shown that the sampler had achieved a reasonable mixing performance.

Table 3 Estimated parameter and associated statistics of global Bayesian bandwidth estimator.

n	Estimate	Std	BMSE	SIF	Acceptance rate
10	0.2187	0.1533	0.0023	2.32	0.5108
25	0.1610	0.1171	0.0017	2.17	0.4870
50	0.1010	0.0755	0.0018	5.76	0.5524
100	0.0586	0.0430	0.0010	5.77	0.4619
200	0.0284	0.0218	0.0004	3.47	0.4864

Table 4 Estimated parameter and associated statistics of variance model error σ^2.

n	Estimate	Std	BMSE	SIF
10	0.4605^2	0.1081	0.0010	0.9270
25	0.3181^2	0.0303	0.0002	0.4954
50	0.1824^2	0.0068	6.9775e-05	1.0486
100	0.1697^2	0.0041	4.1472e-05	1.0137
200	0.1261^2	0.0015	2.0193e-05	1.5955

4.2.2 Bayesian approach versus cross-validation method

In this paragraph, we investigate the performance of proposed Bayesian approach and cross-validation technique using the Average Squared Error (ASE) defined by

$$\mathrm{ASE} = \frac{1}{n}\sum_{i=1}^{n}\{\widehat{m}_{n,h}(x_i) - m(x_i)\}^2.$$

For each dataset the number of replications is $N_{sim} = 50$. The presented results in Table 5 shown that Bayesian approach performed better than the cross-validation technique for small sample sizes (n = 10, 25 and 50). However, the performance were quite similar for moderate sample sizes (n = 100 and 200). The mean and the standard deviation of the global Bayseian bandwidth (h_{Bayes}) and the global cross-validation bandwidth (h_{CV}) based on 50 replications were also reported in Table 6. The h_{CV}-values were very small in comparison with h_{Bayes}-values.

Table 5 Some expected values of $\overline{\text{ASE}}$ and their standard errors in parentheses based on 50 replications.

Model error	n	$\overline{\text{ASE}}_{CV}$	$\overline{\text{ASE}}_{Bayes}$
	10	0.0960(0.2882)	0.0842(0.0531)
	25	0.0501(0.1062)	0.0377(0.0489)
$\mathscr{N}(0, 0.1^2)$	50	0.0324(0.0869)	0.0208(0.0256)
	100	0.0133(0.0144)	0.0133(0.0079)
	200	0.0076(0.0048)	0.0077(0.0049)

Table 6 Mean and standard deviation (sd) in parentheses for global bandwidth (Bayes and CV) based on 50 replications.

Model error	n	$\overline{h}_{CV}$	$\overline{h}_{Bayes}$
	10	0.0539(0.0580)	0.2406(0.0513)
	25	0.0404(0.0693)	0.1652(0.0315)
$\mathscr{N}(0, 0.1^2)$	50	0.0035(0.0034)	0.0917(0.0254)
	100	0.0031(0.0000)	0.0526(0.0140)
	200	0.0031(0.0000)	0.0271(0.0050)

From this simulation study, Bayesian estimation of bandwidth in the semi-parametric kernel estimations of pmf and crf provides better results than cross-validation in the sense of chosen accuracy measure (ISE or ASE), for all sample sizes considered.

5 Applications to real count datasets

In this section Bayesian approach is applied in comparison with cross-validation procedure for semiparametric estimation and regression of count data using binomial kernel. First we estimate the pmf of longevity of adult insects, and then we estimate the semiparametric regression of average daily fat data.

5.1 Poisson-weighted semiparametric binomial estimation

We consider the count dataset related to the development of an insect parasite called the spiraling whitefly and observed in Republic of Congo (Senga Kiessé and

Mizère, 2012). This insect pest plant causes some damages as sucking the sap, decreasing photosynthesis activity and drying up the leaves. The congolese biologists are searching for a suitable modeling by studying some count data characterizing the growth of spiraling whitefly such as the longevity of the adult insect (Table 7).

Table 7 Data of longevity of adult insects observed in days (Senga Kiessé and Mizère, 2012)

Days	1	2	3	4	5	6	7	8	9
Observed frequencies	29	16	22	8	2	4	0	0	1

The performance of $\widehat{f}_n$ with respect to each method of bandwidth selection was evaluated by

$$\mathrm{ISE}_0 = \sum_{x\in\mathbb{N}} \left\{\widehat{f}_{n,h}(x) - f_0(x)\right\}^2,$$

where $f_0(x)$ is the empirical frequency estimate of observations. We provide the ISE_0 values and the execution times for Bayesian and cross-validation approaches in Table 8. The performance of Bayesian approach and cross-validation technique were quite similar in the sense of ISE_0 but Bayesian approach performed better than the cross-validation method in term of execution times. Note that the result with parametric model $p(x;\hat{\mu})$ are also given in the Table 8. The best results are obtained with semiparametric approach for these count dataset.

Table 8 Comparison of ISE_0 and execution times (in second) for cross-validation and Bayes approaches for semiparametric estimation of data in Table 7

Method	Semi-Parametric with CV	Semi-Parametric with Bayes	Parametric
Criterion ISE_0	0.0308	0.0313	0.0385
Execution times	0.2500	0.1100	

5.2 Semiparametric binomial kernel regression

The real count dataset concerns the study of average daily fat (kg/day) yields from the milk of a single cow for each of the 35 first weeks denoted x_i. The quantity of fat in the milk increases during the first 14 weeks and decreases thereafter, see Table 9. These data have been analyzed by Senga Kiessé and Rivoire (2011) using the discrete semiparametric regression and, Kokonendji *et al.*(2009b) and Zougab *et al.*(2014a) using the discrete nonparametric regression.

We applied the GLM and the semiparametric binomial kernel regression estimator to approximate m for the considered average daily fat data. The GLM was employed as start parametric model for the discrete semiparametric binomial kernel estimator. We set the parameters of the priors as $(\alpha, \beta) = (1.5, \sqrt{n})$ and $(a, b) = (1, 0.05)$ and we used the MCMC method with Gibbs sampler based on 5,000 iterations of the burn-in period and 15,000 as the total number of iterations

Table 9 Average daily fat data (Kokonendji *et al.*, 2009b)

x_i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	0.31	0.39	0.50	0.58	0.59	0.64	0.68	0.66	0.67	0.70	0.72	0.68
x_i	13	14	15	16	17	18	19	20	21	22	23	24
y_i	0.65	0.64	0.57	0.48	0.46	0.45	0.31	0.33	0.36	0.30	0.26	0.34
x_i	25	26	27	28	29	30	31	32	33	34	35	
y_i	0.29	0.31	0.29	0.20	0.15	0.18	0.11	0.07	0.06	0.01	0.01	

for estimating the global bandwidth h and the variance of model error σ^2. The MCMC convergence was examined using the BMSE and the SIF. The bandwidth selection was also investigated by the cross-validation technique in comparison with Bayesian approach. For evaluating the performance of the estimators, we used the Root Mean Square Error (RMSE) defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\{y_i - \widehat{m}_{n,h}(x_i)\}^2},$$

where $\widehat{m}_{n,h}(x_i)$ is the adjustment of the ith observation x_i.

Table 10 presents the results corresponding to average daily fat data. For the discrete semiparametric binomial kernel regression estimator, the obtained optimal bandwidths were $h_{cv} = 0.0031$ and $h_{Bayes} = 0.1554$. The cross-validation method had smaller RMSE than Bayesian approach. However, the difference was not significant since $\text{RMSE}_{Bayes} - \text{RMSE}_{CV} = 0.0008$. Note that Bayesian approach gave also the estimate of the variance of model error ($\hat{\sigma}^2 = 0.0663^2$) but not the cross-validation technique.

Figure 1 shown that the discrete semiparametric binomial kernel regression estimator with Bayesian bandwidth improved the GLM which tends to underestimate or overestimate the y-values.

Table 10 Estimated parameters and their associated statistics of Bayesain bandwidth estimation with Gaussian model error on data in Table 9. Comparison using RMSE

Method	Parameter	estimate	Std	BMSE	SIF	Acceptance rate	RMSE
GLM							0.0541
Semip-Bayes	σ^2	0.0663^2	0.0010	9.5211e-06	0.7616		
	h	0.1554	0.1093	0.0027	6.3513	0.5143	0.0228
Semip-CV	h	0.0031					0.0220

6 Concluding remarks

In this work, Bayesian estimations of smoothing parameter in discrete nonparametric kernel estimation has been extended to discrete semiparametric estimation of pmf and crf by using binomial kernel in equation (1). Simulation studies

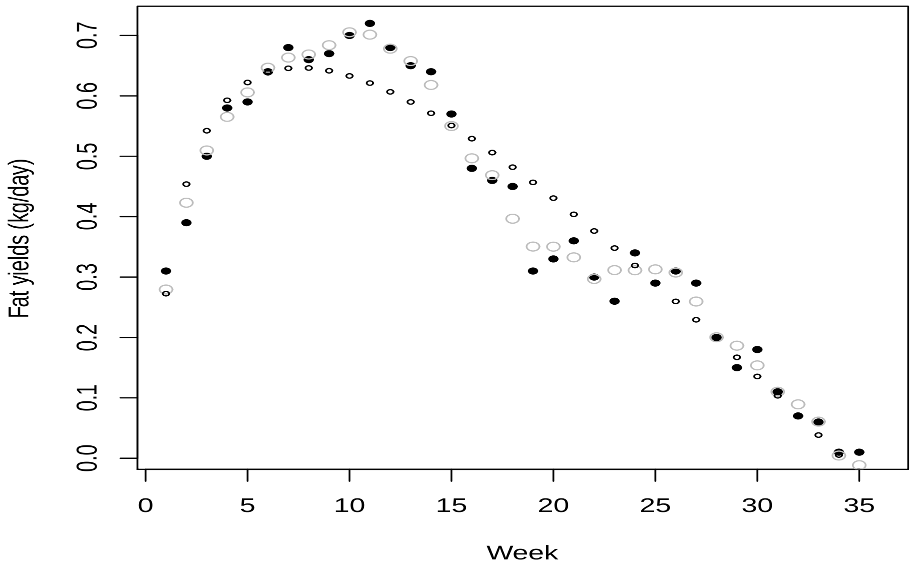

Fig. 1 Semiparametric regression estimation for average daily fat data. The points in black represent the observed data, the circles in grey represent the semiparametric binomial kernel regression estimator with Bayesian bandwidth and the circles in black represent the GLM.

and applications to real count data shown that Bayesian approaches are interesting and valuable alternatives to the classical cross-validation method. Indeed, the posterior estimate of local bandwidth outperforms the bandwidth selected through cross-validation for all small and moderate sample sizes considered. More precisely, Bayesian approaches are better or quite similar to cross-validation method in terms of smoothing quality and execution times, for semiparametric estimation of pmf as well as for semiparametric estimation of crf. Furthermore, the cross-validation method has the inconvenience that it is not always consistent depending on the discrete kernel and data used. Finally, it would be interesting to investigate Bayesian estimation of bandwidth by using other discrete kernels and adaptive approach (Zougab *et al.*, 2014b).

Acknowledgments

The research and education chair of civil engineering and eco-construction is financed by the Chamber of Trade and Industry of Nantes and Saint-Nazaire cities, the CARENE (urban agglomeration of Saint-Nazaire), Charier, Architectes Ingénieurs Associés, Vinci construction, the Regional Federation of Buildings, and the Regional Federation of Public Works. The authors wish also to thank these partners for their patronage.

The authors would like to thank two anonymous referees for their careful readings and helpful comments that led to a considerable improvement of the paper.

References

1. Abdous, B., Kokonendji, C.C., Senga Kiessé, T., 2012. On semiparametric regression for count explanatory variables. *Journal of Statistical Planning and Inference* 142, 1537–1548.
2. Brewer, M.J., 1998. A modelling approach for bandwidth selection in kernel density estimation. In: *Proceedings of COMPSTAT* 1998, 203–208, Heidelberg: Physica Verlag.
3. Cuny, H.E., Senga Kiessé, T., 2014. On modeling wood formation using parametric and semiparametric regressions for count data. *Communications in Statistics - Simulation and Computation*, DOI:10.1080/03610918.2013.875570.
4. Fan, J., Wu, Y., Feng, Y., 2009. Local quasi-likelihood with a parametric guide. *The Annals of Statistics* 37, 4153–4183.
5. Gelman, A., Roberts, G.O., Gilks, W.R., 1996. Efficient Metropolis jumping rules. *Bayesian Statistics* 5, 599–608.
6. Glad, I.K., 1998. Parametrically Guided Non-parametric Regression. *Scandinavian Journal of Statistics* 25, 649–668.
7. Hjort, N.L., Glad, I.K., 1995. Nonparametric density estimation with a parametric start. *Annals of Statistics* 23, 882–904.
8. Hjort, N.L., Jones, M.C., 1996. Locally parametric nonparametric density estimation. *Annals of Statistics* 24, 1619–1647.
9. Hu, S., Poskitt, D.S., Zhang, X., 2012. Bayesian adaptive bandwidth kernel density estimation of irregular multivariate distributions. *Computational Statistics and Data Analysis* 56, 732–740.
10. Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies* 65, 361–393.
11. Kokonendji, C.C., 2014. Over- and underdispersion models. In *The Wiley Encyclopedia of Clinical Trials - Methods and Applications of Statistics in Clinical Trials, Vol. 2 - Planning, Analysis, and Inferential Methods*, Chap. 30, pp. 506-526, (Editor N. Balakrishnan), John Wiley & Sons, Newark.
12. Kokonendji, C.C., Pérez-Casany, M., 2012. A note on weighted count distributions. *Journal of Statistical Theory and Applications* 11, 337–352.
13. Kokonendji, C.C., Senga Kiessé, T., 2011. Discrete associated kernels method and extensions. *Statistical Methodology* 8, 497–516.
14. Kokonendji, C.C., Zocchi, S.S., 2010. Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions. *Statistics and Probability Letters* 80, 1655–1662.
15. Kokonendji, C.C., Senga Kiessé, T., Balakrishnan, N., 2009a. Semiparametric estimation for count data through weighted distributions. *Journal of Statistical Planning and Inference* 139, 3625–3638.
16. Kokonendji, C.C., Senga Kiessé, T., Demétrio, C.G.B., 2009b. Appropriate kernel regression on a count explanatory variable and applications. *Advances and Applications in Statistics* 12, 99–125.
17. McCullagh, P. and Nelder, J. A., 1989. *Generalized Linear Models*, 2nd edn., London: Chapman and Hall.
18. Naito, K., 2004. Semiparametric density estimation by local L2-fitting. *Annals of Statistics* 32, 1162–1191.
19. Senga Kiessé, T., 2008. Nonparametric approach by discrete associated kernel for count data. Ph.D. of University of Pau, France. http://tel.archives-ouvertes.fr/tel-00372180/fr/.
20. Senga Kiessé, T., Cuny, H.E., 2014. Discrete triangular associated kernel and bandwidth choices in semiparametric estimation for count data. *Journal of Statistical Computation and Simulation* 84, 1813–1829.
21. Senga Kiessé, T., Mizère, D., 2012. Weighted Poisson and semiparametric kernel models applied for a parasite growth. *Australian and New Zealand Journal of Statistics* 55, 1–13.
22. Senga Kiessé, T., Rivoire, M., 2011. Discrete semiparametric regression models with associated kernel and applications. *Journal of Nonparametric Statistics* 23, 927–941.
23. Wansouwé, W.E., Kokonendji, C.C., Kolyang, D.T., 2015. Disake: Discrete associated kernel estimators. URL http://cran.r-project.org/web/packages/Disake/.

24. Zhang, X., Brooks, R.D., King, M.L., 2009. A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation. *Journal of Econometrics* 153, 21–32.
25. Zhang, X., King, M.L., Shang, H.L., 2014. A sampling algorithm for bandwidth estimation in a nonparametric regression model with a flexible error density. *Computational Statistics and Data Analysis* 78, 218–234.
26. Zougab, N., Adjabi, S., Kokonendji, C.C., 2012. Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics* 24, 783–795.
27. Zougab, N., Adjabi, S., Kokonendji, C.C., 2014a. Bayesian approach in nonparametric count regression with binomial kernel. *Communications in Statistics - Simulation and Computation* 43, 1052–1063.
28. Zougab, N., Adjabi, S., Kokonendji, C.C., 2014b. Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. *Computational Statistics and Data Analysis* 75, 28–38.