On semiparametric regression for count explanatory variables

Belkacem Abdous, Célestin C. Kokonendji, Tristan Senga Kiessé

On semiparametric regression for count explanatory variables

Belkacem Abdous [a],

[a] *Université Laval, Médecine Sociale et Préventive*
Québec, Canada G1K 7P4

Célestin C. Kokonendji [b] and

[b] *Université de Franche-Comté, UFR Sciences et Techniques*
Laboratoire de Mathématiques de Besançon - UMR 6623 CNRS
16 route de Gray – 25030 Besançon cedex, France

Tristan Senga Kiessé [c,*]

[c] *Office National des Forêts*
54840 Velaine-en-Haye, France

Abstract

This work deals with a semiparametric estimation of a count regression function m that can be represented as a product of an unknown discrete parametric function r and an unknown discrete "smooth" function ω. We propose an estimation procedure in two steps: first, we construct an approximation $\widehat{r}$ of r, then we use a discrete associated kernel method to estimate non-parametrically the multiplicative correction factor $\omega = m/\widehat{r}$. The asymptotic and small-sample properties of the proposed estimator are investigated. Its comparison with the classical Nadaraya-Watson type count regression estimator shows that an improvement in terms of bias is achieved.

Key words: Asymptotic bias and variance, Discrete associated kernel, Discrete finite difference, Multiplicative correction factor, Nonparametric regression, Nadaraya-Watson estimator, Logistic regression.
2000 MSC: 62G08, 62J02, 62G99

* *Corresponding author*
Email addresses: belkacem.abdous@msp.ulaval.ca (Belkacem Abdous), celestin.kokonendji@univ-fcomte.fr (Célestin C. Kokonendji), tristan.senga-kiesse@onf.fr (Tristan Senga Kiessé).

1 Introduction

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

$$Y_i = m(X_i) + \epsilon_i, \tag{1}$$

where m is an unknown regression function and the ϵ_i's are assumed to have zero mean and finite variance. Parametric and nonparametric estimation of the regression function m are classical problems. The continuous case has received much more attention than the discrete one. Semiparametric approaches have been considered by several authors as well. For a recent account of the (multiplicative or additive) combinations of parametric and nonparametric kernel estimators, see, e.g., Martins-Filho et al. (2008), Su and Ullah (2008), Fan et al. (2009). While, nonparametric estimation of discrete functions such as probability mass or count regression functions is definitely less popular. An extension of the well known kernel technique has been investigated by Kokonendji et al. (2007, 2009b).

In this paper we focus on the semiparametric estimation of the unknown count regression function m given by (1). More precisely, we investigate the following form:

$$\begin{aligned} m(x) &= r(x;\beta)\omega(x) \\ &=: m_\omega(x;\beta), \text{ for } x \in \mathbb{N}^d, \end{aligned} \tag{2}$$

where $r(\cdot;\beta)$ is a parametric function that depends on unknown parameter $\beta = (\beta_1, \dots, \beta_p)^T$, and $\omega(\cdot)$ is a nonparametric multiplicative correction function. As shown later on, by splitting the estimation problem of the conditional mean function $m(x) = \mathbb{E}(Y|X = x)$ up into a parametric procedure for r and a kernel-based technique for ω, we improve the asymptotic bias of the proposed estimator.

The rest of this paper is organized as follows. Section 2 briefly recalls the nonparametric count regression estimator of m and investigates its asymptotic bias and variance expansions. In Section 3, we present our semiparametric estimation of (2) for $d = 1$. The proposed estimator is constructed in two steps: first, we use a generalized least squared criterion to find an estimate $\widehat{\beta}$ of the parameter β in $r(x;\beta)$; then a discrete associated kernel approach (Kokonendji et al. 2007, 2009a) is used to provide a nonparametric estimator of the multiplicative correction factor $\omega(x;\widehat{\beta}) = m(x)/r(x;\widehat{\beta})$. The asymptotic bias and variance of the obtained estimator are investigated and compared to those of a nonparametric discrete associated kernel estimator. The extension to higher dimension is tackled as well. Section 4 provides two applications on real data, while proofs of all results of Section 3 are postponed to Section 5.

2 Nonparametric count regression estimator

In this section we summarize some asymptotic expansions of the nonparametric count regression estimator introduced by Kokonendji et al. (2009b). Indeed, suppose that we are interested in a count regression function $m(x) = \mathbb{E}(Y|X = x)$ in which the r.v. Y and X belong to $\mathbb{R}$ and $\mathbb{N}$ respectively; without loss of generality, we here present the univariate case with $d = 1$ (see Section 3.3). Then, the well-known Nadaraya-Watson kernel estimate might be extended to the case of count regression as follows

$$\widetilde{m}_n(x) = \sum_{i=1}^{n} \frac{Y_i K_{x,h}(X_i)}{\sum_{j=1}^{n} K_{x,h}(X_j)}, \quad x \in \mathbb{N}, \tag{3}$$

where $h = h(n) > 0$ is an arbitrary sequence of smoothing parameters that fulfills $\lim_{n\to\infty} h(n) = 0$, while $K_{x,h}(\cdot)$ is a suitably chosen discrete associated kernel function (e.g. Kokonendji et al., 2007). The kernel function $K_{x,h}(\cdot)$ is itself a probability mass function (p.m.f.) with support S_x that does not depend on h and such that $x \in S_x$. Furthermore, we also impose the following two conditions:

$$\lim_{h\to 0} \mathbb{E}(\mathcal{K}_{x,h}) = x, \tag{4}$$

$$\lim_{h\to 0} Var(\mathcal{K}_{x,h}) = 0, \tag{5}$$

where $\mathcal{K}_{x,h}$ is the discrete random variable whose p.m.f. is $K_{x,h}(\cdot)$. See also Abdous and Kokonendji (2009) for a review of discrete associated kernels and asymptotic behaviour for discrete smoothing using (4) and (5).

To deal with derivatives of any count function $g : \mathbb{N} \to \mathbb{R}$, we will use finite differences instead of the usual differentiation on $\mathbb{R}$. For $k \in \mathbb{N} \setminus \{0\}$, we will put

$$g^{(k)}(x) = \{g^{(k-1)}(x)\}^{(1)} \text{ and } g^{(1)}(x) = \begin{cases} \{g(x+1) - g(x-1)\}/2 & \text{if } x \in \mathbb{N} \setminus \{0\} \\ g(1) - g(0) & \text{if } x = 0. \end{cases} \tag{6}$$

For instance, it follows that the finite difference of second order is given by

$$g^{(2)}(x) = \begin{cases} \{g(x+2) - 2g(x) + g(x-2)\}/4 & \text{if } x \in \mathbb{N} \setminus \{0, 1\} \\ \{g(3) - 3g(1) + 2g(0)\}/4 & \text{if } x = 1 \\ \{g(2) - 2g(1) + g(0)\}/2 & \text{if } x = 0. \end{cases} \tag{7}$$

Note that the finite differences $g^{(k)}$ are always defined no matter the smoothness of g.

The following theorem presents the asymptotic bias and variance of the nonparametric estimator $\widetilde{m}_n(x)$ in (3).

Theorem 1 *Let f be a p.m.f. of a discrete r.v. X defined on $\mathbb{N}$. Assume that $f(x) = \Pr(X = x) > 0$ for a given $x \in \mathbb{N}$. Furthermore, suppose that the bandwidth $h = h(n) > 0$ satisfies $\lim_{n\to\infty} h = 0$ and that the discrete kernel $K_{x,h}(\cdot)$ fulfills assumptions (4)-(5). Then, the bias and variance of $\widetilde{m}_n(x)$ admit the following expansions*

$$Bias\{\widetilde{m}_n(x)\} = \left\{m^{(2)}(x) + 2m^{(1)}(x)\left(\frac{f^{(1)}}{f}\right)(x)\right\}\frac{Var(\mathcal{K}_{x,h})}{2} + O\left(\frac{1}{n}\right) + o(h), \tag{8}$$

$$Var\{\widetilde{m}_n(x)\} = \frac{Var(Y|X = x)}{nf(x)}\{\Pr(\mathcal{K}_{x,h} = x)\}^2 + o\left(\frac{1}{n}\right), \tag{9}$$

where $f^{(1)}$, $m^{(1)}$ and $m^{(2)}$ are finite differences as defined in (6) and (7).

The proof is given in Section 5.

3 Semiparametric count regression estimator

Following Kokonendji et al. (2009a) for p.m.f., we here adapt the semiparametric approach of count regression function m in (1) using expression in (2) with a parametric start $r(\cdot;\beta)$ and a nonparametric correction function $\omega(\cdot)$. In fact, given a sample (X_i, Y_i), $i = 1, \ldots, n$, it uses any parametric estimation procedure to provide an estimation of β involved in $r(x;\beta)$, and it serves the nonparametric discrete associated kernel procedure to estimate the correction function $\omega(x) = m(x)/r(x;\beta)$ by the following Nadaraya-Watson type estimator

$$\widetilde{\omega}_n(x;\widehat{\beta}) = \sum_{i=1}^{n} \frac{Y_i K_{x,h}(X_i)}{r(X_i;\widehat{\beta})\sum_{j=1}^{n} K_{x,h}(X_j)}, \quad x \in \mathbb{N}, \tag{10}$$

where $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^T$ is an estimate of β constructed in the previous step (by generalized least squared method for example). Upon combining these two estimation steps, we end up with the following semiparametric estimator $\widehat{m}_n$ of m

$$\widehat{m}_n(x) = r(x;\widehat{\beta})\widetilde{\omega}_n(x;\widehat{\beta}) = r(x;\widehat{\beta})\sum_{i=1}^{n} \frac{Y_i K_{x,h}(X_i)}{r(X_i;\widehat{\beta})\sum_{j=1}^{n} K_{x,h}(X_j)}. \tag{11}$$

Next, let us examine the asymptotic bias and variance of the proposed estimator (11) and compare them to those of the traditional estimator (3). These asymptotics will be tackled under two settings: known and unknown parametric functionr $r(x;\beta)$. Besides, we will investigate the multivariate case as well.

3.1 Known parametric start

Suppose that based on a goodness-of-fit test or any a priori knowledge about m we decide to fix the parametric start in (2) and put $r_0(x) = r(x;\beta_0)$. Then, by letting $m(x) = r_0(x)\omega(x)$ and by using the estimation procedure described above, it is easy to see that a semiparametric estimate of m is given by

$$\widehat{m}_n(x) = r_0(x)\widetilde{\omega}_n(x) = \sum_{i=1}^{n} \frac{Y_i K_{x,h}(X_i)}{\sum_{j=1}^{n} K_{x,h}(X_j)} \times \frac{r_0(x)}{r_0(X_i)}, \quad x \in \mathbb{N}. \tag{12}$$

The following theorem exhibits the asymptotic expansions of $\widehat{m}_n(x)$'s bias and variance

Theorem 2 *Let x be a given point in $\mathbb{N}$ satisfying $f(x) = \Pr(X = x) > 0$. Assume that the regression function satisfies $m(x) = r_0(x)\omega(x)$ with $r_0(x) = r(x;\beta_0)$ being a fixed start. Then, provided that $h = h(n) \to 0$ for $n \to \infty$, the estimator $\widehat{m}_n(x)$ verifies*

$$Bias\{\widehat{m}_n(x)\} = \left\{ r_0(x)\omega^{(2)}(x) + 2r_0(x)\omega^{(1)}(x)\left(\frac{f^{(1)}}{f}\right)(x) \right\} \frac{Var(\mathcal{K}_{x,h})}{2} + O\left(\frac{1}{n}\right) + o(h), \tag{13}$$

$$Var\left\{\widehat{m}_n(x)\right\} = \frac{Var(Y|X=x)}{nf(x)} \{\Pr(\mathcal{K}_{x,h} = x)\}^2 + o\left(\frac{1}{n}\right), \tag{14}$$

where $f^{(1)}$, $\omega^{(1)}$ and $\omega^{(2)}$ are finite differences as defined in (6) and (7).

Proof. See Section 5.

A comparison of the asymptotic expansions of the semiparametric estimator $\widehat{m}_n$ in (12) and the nonparametric discrete Nadaraya-Watson type estimator $\widetilde{m}_n$ in (3) shows that the leading terms in the variance expressions are identical, while the bias expressions are different. More precisely, since under the assumption $m(x) = r_0(x)\omega(x)$ one has

$$m^{(1)} = (r_0\omega)^{(1)} = r_0\omega^{(1)} + r_0^{(1)}\omega$$

$$m^{(2)} = (r_0\omega)^{(2)} = r_0\omega^{(2)} + 2r_0^{(1)}\omega^{(1)} + r_0^{(2)}\omega,$$

it follows that the difference between the leading terms in Bias$\{\widehat{m}_n(x)\}$ and Bias$\{\widetilde{m}_n(x)\}$ is given by:

$$\left\{r_0^{(2)}(x)\omega(x) + 2r_0^{(1)}(x)\omega^{(1)}(x) + 2r_0^{(1)}(x)\omega(x)\left(\frac{f^{(1)}}{f}\right)(x)\right\}\frac{Var(\mathcal{K}_{x,h})}{2}.$$

Unfortunately, the sign of this quantity can be either positive or negative.

3.2 Unknown parametric start

Next, let us consider the semiparametric count regression estimator $\widehat{m}_n$ in (11) of m in (2) with an estimate $\widehat{\beta}$ of β (still constructed by generalized least squared method for example). In the situation where the parametric function $r(x;\beta)$ is misspecified, the estimator $\widehat{\beta}$ of β converges in probability to a certain value β_0 such that $r(x;\beta_0)$ is the best approximant to $m(x)$ with respect to the Kullback-Leibler distance

$$\sum_{x\in\mathbb{N}} m(x)\log\frac{m(x)}{r(x;\beta)} =: d\left\{m(\cdot), r(\cdot;\beta)\right\}$$

of $r(x;\beta)$ from the true function $m(x)$; see, for example, White (1982).

A Taylor expansion of the ratio $r(x;\widehat{\beta})/r(X_i;\widehat{\beta})$ around β_0 provides

$$\begin{aligned}\frac{r(x;\widehat{\beta})}{r(X_i;\widehat{\beta})} &= \exp\{\log r(x;\widehat{\beta}) - \log r(X_i;\widehat{\beta})\}\\ &\doteq \frac{r_0(x)}{r_0(X_i)}\left[1 - \{u_0(X_i) - u_0(x)\}^T(\widehat{\beta}-\beta_0) + \frac{1}{2}(\widehat{\beta}-\beta_0)^T M(x,X_i)(\widehat{\beta}-\beta_0)\right],\end{aligned}$$

where $\doteq$ denotes an asymptotic equivalent, and the matrix $M(x, X_i)$ is

$$M(x,X_i) = v_0(x) - v_0(X_i) + \{u_0(x) - u_0(X_i)\}\{u_0(X_i) - u_0(x)\}^T$$

such that $u_0(x)$ and $v_0(x)$ are the gradient and the Hessian matrix with respect to β, respectively, evaluated in β_0. Using the above expansion, the estimator $\widehat{m}_n$ in (11) can be approximated as

$$\widehat{m}_n(x) \doteq \widehat{m}_n^0(x) + A_n(x) + \frac{1}{2}B_n(x), \qquad (15)$$

where $\widehat{m}_n^0$ is the estimator in (12) with fixed start and the two others terms are

$$A_n(x) = \sum_{i=1}^{n} \frac{[\{u_0(X_i) - u_0(x)\}^T(\widehat{\beta} - \beta_0)]Y_i K_{x,h}(X_i)}{\sum_{j=1}^{n} K_{x,h}(X_j)} \times \frac{r(x;\beta_0)}{r(X_i;\beta_0)},$$

$$B_n(x) = \sum_{i=1}^{n} \frac{\{(\widehat{\beta} - \beta_0)^T M(x, X_i)(\widehat{\beta} - \beta_0)\}Y_i K_{x,h}(X_i)}{\sum_{j=1}^{n} K_{x,h}(X_j)} \times \frac{r(x;\beta_0)}{r(X_i;\beta_0)}.$$

We then formulate the following result.

Theorem 3 *Let x be a given point in $\mathbb{N}$ satisfying $f(x) = \Pr(X = x) > 0$. Assume that the parametric function $r_0(x) = r(x;\beta_0)$ is the best approximant of the regression function $m(x)$ under the Kullback-Leibler criterion, with $\omega(x) = m(x)/r_0(x)$ being a multiplicative correction factor. Then, provided that $n \to \infty$ and $h = h(n) \to 0$, the estimator $\widehat{m}_n(x)$ in (11) satisfies*

$$\begin{aligned} Bias\{\widehat{m}_n(x)\} = & \left\{ r_0(x)\omega^{(2)}(x) + 2r_0(x)\omega^{(1)}(x) \left(\frac{f^{(1)}}{f} \right)(x) \right\} \frac{Var(\mathcal{K}_{x,h})}{2} \\ & + O\left(\frac{1}{n}\right) + o(h), \\ Var\Big\{\widehat{m}_n(x)\Big\} = & \frac{Var(Y|X = x)}{nf(x)} \{\Pr(\mathcal{K}_{x,h} = x)\}^2 + o\left(\frac{1}{n}\right), \end{aligned}$$

where $f^{(1)}$, $\omega^{(1)}$ and $\omega^{(2)}$ are finite differences as defined in (6) and (7).

Proof. See Section 5.

Looking at the previous theorem, the asymptotic bias and variance are the same as in Theorem 2 for the case of the fixed start r_0. Hence, the proposed estimator $\widehat{m}_n$ in (11) of m can once again be shown to be better or not than the nonparametric one in (3). See, for example, Kokonendji et al. (2009a) for some illustrations in the case of count data distributions.

3.3 Multivariate case

The multidimensional estimators (3) and (11) are easily obtained by using a multiplicative d-univariate kernel

$$K_{x,h}(X_i) = \prod_{k=1}^{d} K^{[k]}_{x_k,h_k}(X_{ik}),$$

with $x = (x_1, \ldots, x_d)^T$, $h = (h_1, \ldots, h_d)^T$, $X_i = (X_{i1}, \ldots, X_{id})^T$ and $K^{[k]}_{x_k,h_k}(\cdot)$ is a univariate discrete associated kernel satisfying (4) and (5) for all $k =$

$1, \ldots, d$. We now consider the d-dimensional discrete functions $m(\cdot)$, $r(\cdot;\beta)$ and $\omega(\cdot)$ in (2) with their corresponding partial finite differences in the sense of (6) and (7), for example, $g_k^{(1)}(x) = \partial g(x)/\partial x_k$ and $g_{kl}^{(2)}(x) = \partial^2 g(x)/(\partial x_k \partial x_l)$ and so on.

The asymptotic bias and variance of the d-dimensional discrete Nadaraya-Watson estimator replacing (8) and (9), respectively, are given by

$$Bias\{\widetilde{m}_n(x)\} = \frac{1}{2}\sum_{k=1}^{d}\left\{m_{kk}^{(2)}(x) + 2m_k^{(1)}(x)\left(\frac{f_k^{(1)}}{f}\right)(x)\right\} Var(\mathcal{K}_{x_k,h_k}^{[k]}) + O\left(\frac{1}{n}\right) + o\left(\frac{h_1^2 + h_2^2 + \ldots + h_d^2}{h_1 h_2 \ldots h_d}\right),$$

$$Var\{\widetilde{m}_n(x)\} = \frac{Var(Y|X=x)}{nf(x)}\prod_{k=1}^{d}\{\Pr(\mathcal{K}_{x_k,h_k}^{[k]} = x_k)\}^2 + o\left(\frac{1}{n}\right).$$

The corresponding expression for the bias of the new estimator given in Theorem 3 is

$$Bias\{\widehat{m}_n(x)\} = \frac{1}{2}\sum_{k=1}^{d}\left\{r_0(x)\omega_{kk}^{(2)}(x) + 2r_0(x)\omega_k^{(1)}(x)\left(\frac{f_k^{(1)}}{f}\right)(x)\right\} Var(\mathcal{K}_{x_k,h_k}^{[k]}) + O\left(\frac{1}{n}\right) + o\left(\frac{h_1^2 + h_2^2 + \ldots + h_d^2}{h_1 h_2 \ldots h_d}\right),$$

while the variance is the same as that of the multidimensional discrete Nadaraya-Watson estimator above. Similarly to that for univariate case, the change of bias is clearly point out.

4 Applications

In this section two practical datasets of demography and economy are used for applying the nonparametric and semiparametric discrete regression approaches in comparison with classical parametric regression models. The first example concerns mortality rates (Copas and Haberman, 1983). For each age $x \in \mathbb{N}$, a population of size s_x is exposed to risk, associated to a number of deaths d_x. The crude death rates $m_0(x)$ is calculated as

$$m_0(x) = \frac{d_x}{s_x}, \quad x \in \mathbb{N}.$$

The frequency mortality rate is generally increasing with respect to the age, having a small numbers of deaths at younger ages. The graduated values obtained by applying a logistic model does not fit well to data (see Figure 1). Indeed, the parametric model produces oversmoothing and, in particular, does not succeed to describe the behaviour on the ages $x \in \{42, 44, 45, 46\}$. The second example is a sales dataset with multiple y_i at a given x_i (Kokonendji et al., 2009b). We analyse the amount of daily sales of a new product during the 24 first days. The 151 observations (x_i, y_i), $i = 1, \ldots, 24$, represent the day x_i and the corresponding mean of sales numbers $y_i \in \{y_{Ai}, y_{Bi}, \ldots, y_{Hi}\}$. The number of sale centres for each state (A, B, ..., H) is not available except for the state H where this number is equal to one. We apply generalized linear models (GLM) in comparison with nonparametric and semiparametric models for fitting the sales data (Figure 2). The given examples indicate that the use of continuous semiparametric model may provide fitted values at any point of the real line numbers, while the predictor is an ordinal variable. This motivates to recommend a discrete model that focuses on ordinal covariates and has the same nature. Hence, the nonparametric correction in the two examples is available only for discrete predictors.
For the discrete semiparametric estimator, the logistic model and GLM are used as start functions. The measure of error used is the root mean squared error RMSE defined as

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}(y_j - \widehat{y_j})^2}{n}},$$

where $\widehat{y_j}$ is the adjustement of the j-th observation y_j and n is the number of observations.

4.1 Parametric models

The GLM represents a normal model for the response variable Y_i with a logarithmic link. It has a linear predictor based on a combination of explanatory variables such as

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

The nonlinear model corresponds to a logistic one for the situation of population growth towards a limited value. It is given by

$$y_i = \frac{\theta_1^L}{1 + \exp\left\{-\left(\frac{x_i - \theta_2^L}{\theta_3^L}\right)\right\}} + e_i, \quad x_i \in \mathbb{N}.$$

The fixed effect parameter θ_1^L is the asymptote towards which the population grows. The parameter θ_2^L is the midpoint and corresponds to the time at which $y_i = \theta_1^L/2$. The parameter θ_3^L is the scale and represents the distance on the time axis between the midpoint and the point where the response is $\theta_1^L/(1+e^{-1})$.

4.2 Discrete associated kernel

For $(x, a) \in \mathbb{N} \times \mathbb{N}$ and $h > 0$, the symmetric discrete triangular associated kernel $K_{a;x,h}$ on support $\mathcal{S}_{a;x} = \{x, x \pm 1, \ldots, x \pm a\}$ has discrete probability distribution given by

$$K_{a;x,h}(z) = \frac{(a+1)^h - |z-x|^h}{P(a,h)}, \quad \forall z \in \mathcal{S}_{a;x},$$

where $P(a,h) = (2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^h$. That kernel fulfills assumptions (4) - (5). Note that a general discrete triangular associated kernel has been proposed by Kokonendji and Zocchi (2010); the corresponding R package is also available (see Senga Kiessé et al., 2010).

Concerning the regression estimators with discrete triangular associated kernels, the optimal bandwidth parameter is selected by the cross-validation method; refer to Kokonendji et al. (2009) for discrete case and, Hardle and Marron (1985) for continuous one. For a given associated discrete kernel, the optimal bandwidth is $h_{cv} = \arg\min_{h>0} CV(h)$ with

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

where $\widetilde{m}_{-i}(X_i; h) = \sum_{j\neq i}^{n} Y_j K_{X_i,h}(X_j)/\sum_{j\neq i}^{n} K_{X_i,h}(X_j)$ is the leave-one-out kernel estimator of $\widetilde{m}_n(X_i; h)$. However, this method does not always converge and, thus, other values of the bandwidth parameter $h \in \{0.5, 3.5\}$ are proposed to point out its influence on goodness-of-fit or discrete smoothness. Here, this procedure is applied only for discrete nonparametric estimator but some works are in progress on its development for discrete semiparametric estimator. For the arm $a \in \mathbb{N}$, a value equals to 1 or 2 is sufficient because the error connected to the adjustment is increasing with respect to the arm a (Kokonendji et al., 2007); hence, we consider the discrete triangular associated kernel $a = 2$.

4.3 Mortality rates

The Figure 1 and Table 1 present the results connected to death rates data. For this dataset, the cross-validation procedure does not converge, consequently we use the two values of h proposed previously. First, using semiparametric estimator with discrete triangular kernel $a = 2$, $h = 0.5$ provides the best adjustement which corresponds to the smallest RMSE; the performance of nonparametric and semiparametric approaches are closed and outperform the parametric model. Then, by changing the bandwidth value in $h = 3.5$ improves the amount of smoothing provided by the two discrete associated kernel estimators. Finally, the semiparametric estimator stays the best one in comparison to the nonparametric and the difference between the performance of the two estimators is more clear.

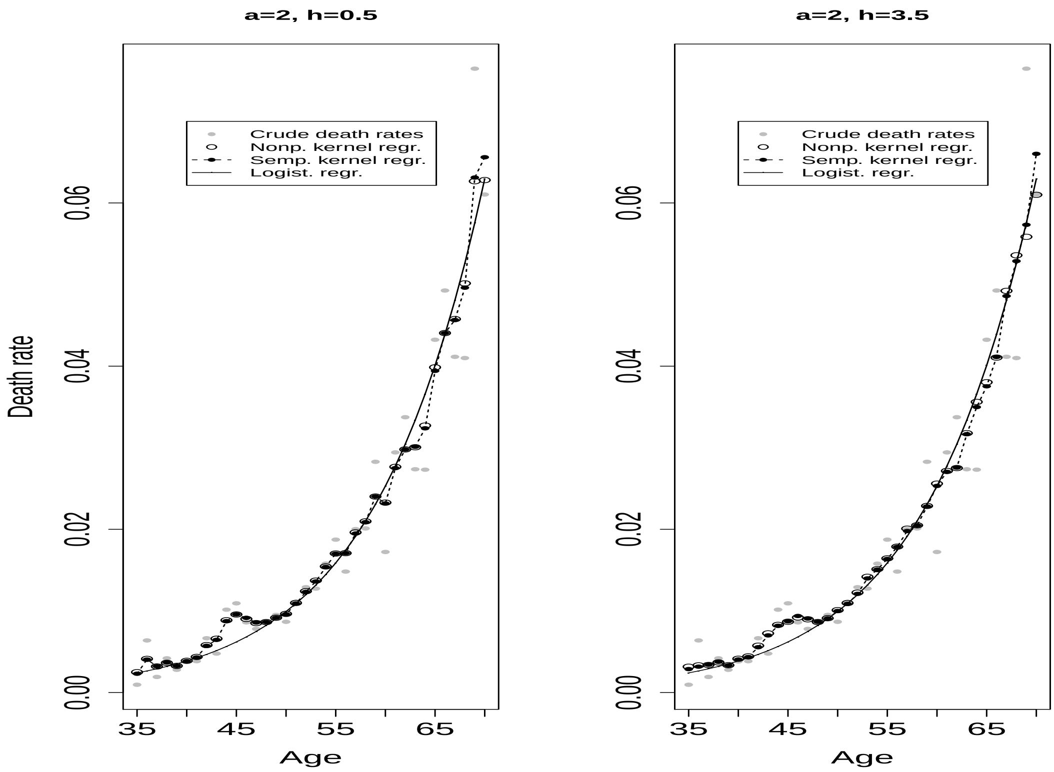

Fig. 1. Logistic regression, nonparametric and semiparametric regressions using symmetric discrete triangular associated kernels on death data

4.4 Sales data

The Figure 2 and Table 2 present the results corresponding to sales data. For this example, the cross-validation method gives an optimal bandwidth value $h_{cv} = 0.132$ for discrete nonparametric estimator; in comparison, we

Table 1
RMSE calculated from logistic model, nonparametric and semiparametric regression estimators using discrete symmetric triangular associated kernels on mortality data

RMSE		
Logistic	Nonp. estimator	Semip. estimator
	with discrete triangular associated kernel $a = 2$	
	$h = 0.5$	
0.500	0.364	0.362
	$h = 3.5$	
	0.535	0.521

applied this value for the discrete semiparametric estimator completed by $h = 3.5$. Similarly to the previous example, the discrete semiparametric triangular model with $a = 2$ and $h = h_{cv} = 0.132$ provides the better adjustment in term of RMSE in comparison to nonparametric and parametric model. However, both satisfying degree of smoothing and fitting are still obtained with $h = 3.5$. Furthermore, the GLM and nonparametric regression model underestimate and overestimate the y-values, respectively, contrary to the good adjustment provided by the semiparametric kernel model.

Table 2
RMSE calculated from GLM, nonparametric and semiparametric regression estimator using discrete symmetric triangular associated kernels on sales data

RMSE		
GLM	Nonp. estimator	Semip. estimator
	with discrete triangular associated kernel $a = 2$	
	$h_{cv} = 0.132$	$h = 0.132$
2.427	3.002	2.215
	$h = 3.5$	
	3.286	2.479

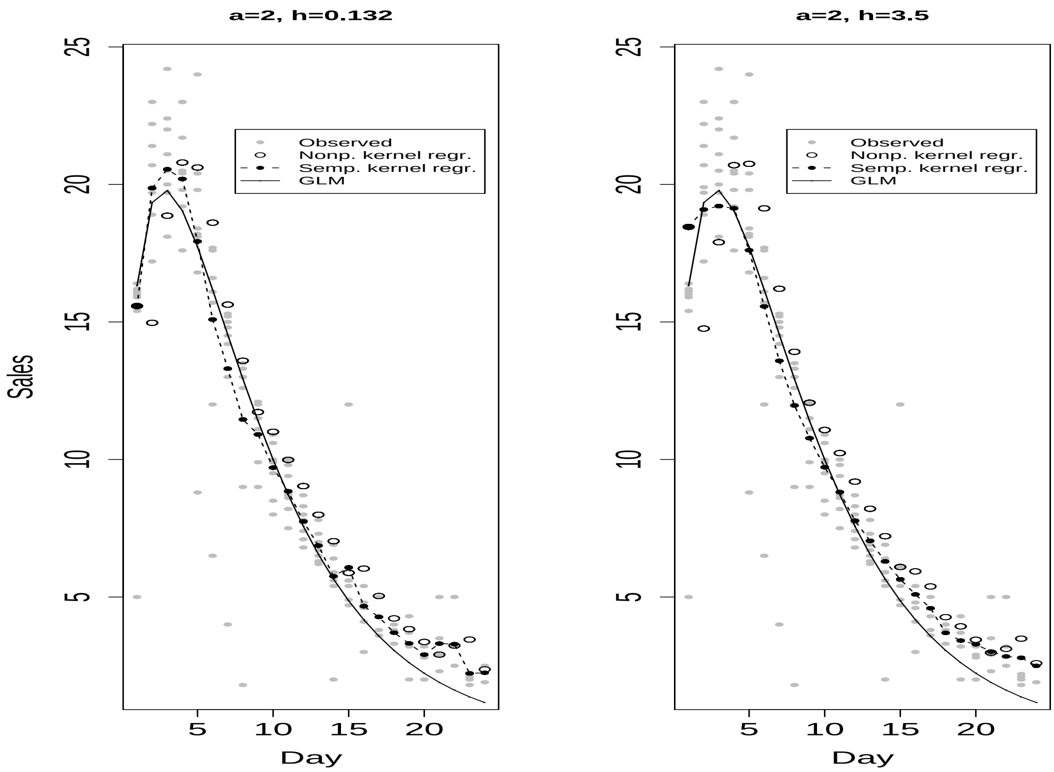

Fig. 2. GLM, nonparametric and semiparametric regressions using symmetric discrete triangular associated kernels on sales data

5 Proofs

5.1 Proof of Theorem 1

Let us rewrite the estimator $\widetilde{m}_n(x)$ in (3), as follows

$$\widetilde{m}_n(x) = \frac{N_n(x;h)}{D_n(x;h)},$$

with $D_n(x;h) = n^{-1}\sum_{j=1}^{n} K_{x,h}(X_j)$ and $N_n(x;h) = n^{-1}\sum_{i=1}^{n} Y_i K_{x,h}(X_i)$. The convergence of $D_n(x;h)$ to $f(x)$ is established by Abdous and Kokonendji (2009; Theorem 2.4) and, similarly, the convergence of $N_n(x;h)$ to $mf(x)$ can de deduced. From here, following Bosq and Lecoutre (1987; p. 119-121), we can write

$$\begin{aligned}\widetilde{m}_n(x) = & \, m(x) + \frac{1}{f(x)}\{N_n(x;h) - (mf)(x)\} - \frac{(mf)(x)}{f^2(x)}\{D_n(x;h) - f(x)\} \\ & - \frac{1}{f^2(x)}\{N_n(x;h) - (mf)(x)\}\{D_n(x;h) - f(x)\} \\ & + \frac{N_n(x;h)}{f^3(x)}\{D_n(x;h) - f(x)\}^2\{1 + o(1)\} \text{ a.s.} \end{aligned} \tag{16}$$

The expectation of $D_n(x;h)$ can be approximated as

$$\begin{aligned}\mathbb{E}\{D_n(x;h)\} = & \, \mathbb{E}\{f(\mathcal{K}_{x,h})\} \\ = & \, f\{\mathbb{E}(\mathcal{K}_{x,h})\} + \frac{1}{2}Var(\mathcal{K}_{x,h})f^{(2)}(x) + o(h),\end{aligned}$$

where $f^{(2)}$ is the finite difference of second order as in (7). Similarly, for the expectation of $N_n(x;h)$, we have

$$\begin{aligned}\mathbb{E}\{N_n(x;h)\} = & \, \mathbb{E}\{Y_1 K_{x,h}(X_1)\} \\ = & \sum_{z \in \mathcal{S}_x} m(z)f(z)\Pr(\mathcal{K}_{x,h} = z) \\ = & \, \mathbb{E}\{(mf)(\mathcal{K}_{x,h})\} \\ = & \, (mf)\{\mathbb{E}(\mathcal{K}_{x,h})\} + \frac{1}{2}Var(\mathcal{K}_{x,h})(mf)^{(2)}(x) + o(h),\end{aligned}$$

Thus, under the condition (4) of discrete associated kernel, we obtain

$$\mathbb{E}\{D_n(x;h)\} - f(x) = \frac{1}{2}Var(\mathcal{K}_{x,h})f^{(2)}(x) + o(h) \tag{17}$$

and

$$\mathbb{E}\{N_n(x;h)\} - (mf)(x) = \frac{1}{2}Var(\mathcal{K}_{x,h})(mf)^{(2)}(x) + o(h). \tag{18}$$

Next, we have

$$\begin{aligned}&\mathbb{E}[N_n(x;h)\{D_n(x;h) - f(x)\}^2] \\ &\quad = \mathbb{E}\left([N_n(x;h) - \mathbb{E}\{N_n(x;h)\}][D_n(x;h) - \mathbb{E}\{D_n(x;h)\}]^2\right) \\ &\qquad + 2[\mathbb{E}\{D_n(x;h)\} - f(x)]Cov\{D_n(x;h), N_n(x;h)\} \\ &\qquad + \mathbb{E}[\{D_n(x;h) - f(x)\}^2]\mathbb{E}\{N_n(x;h)\} \\ &\quad = O(1/n)^2 + O(1/n) + \mathbb{E}\{D_n(x;h) - f(x)\}^2\mathbb{E}\{N_n(x;h)\}.\end{aligned} \tag{19}$$

To get the last inequality, we used the fact that

$$
\begin{aligned}
Cov\{D_n(x;h), N_n(x;h)\} &= \frac{1}{n}[\mathbb{E}\{Y_1 K_{x,h}^2(X_1)\} - \mathbb{E}\{Y_1 K_{x,h}(X_1)\}\mathbb{E}\{K_{x,h}(X_2)\}] \\
&= O(1/n).
\end{aligned}
$$

and

$$
\mathbb{E}\Big([N_n(x;h) - \mathbb{E}\{N_n(x;h)\}][D_n(x;h) - \mathbb{E}\{D_n(x;h)\}]^2\Big) = O(1/n)^2.
$$

Similar arguments enable to see that

$$
\begin{aligned}
&\mathbb{E}\{D_n(x;h) - f(x)\}^2\mathbb{E}\{N_n(x;h)\} - f(x)\mathbb{E}[\{N_n(x;h) - (mf)(x)\}\{D_n(x;h) - f(x)\}] \\
&\quad = O(1/n) + o(h). \qquad (20)
\end{aligned}
$$

Upon plugging the expansions (17)-(20) into (16), we end up with

$$
\mathbb{E}\{\widetilde{m}_n(x)\} - m(x) = \left\{\frac{(mf)^{(2)}(x)}{f(x)} - \frac{mf^{(2)}(x)}{f(x)}\right\}\frac{Var(\mathcal{K}_{x,h})}{2} + O(1/n) + o(h).
$$

Finally, to obtain (8), simply note that $(mf)^{(2)} = m^{(2)}f + 2m^{(1)}f^{(1)} + mf^{(2)}$.

As for the variance expansion (9), it follows from (16) that

$$
\begin{aligned}
Var\{\widetilde{m}_n(x)\} &= \frac{Var\{N_n(x;h)\}}{f^2(x)} + \frac{(mf)^2(x)}{f^4(x)}Var\{D_n(x;h)\} \\
&\quad -2\frac{(mf)(x)}{f^3(x)}Cov\{N_n(x;h), D_n(x;h)\} + o\left(\frac{1}{n}\right). \qquad (21)
\end{aligned}
$$

The variance of $D_n(x;h)$ might be written as follows

$$
\begin{aligned}
Var\left\{D_n(x;h)\right\} &= \frac{1}{n}\sum_{y\in\mathcal{S}_x} f(y)\left\{\Pr(\mathcal{K}_{x,h} = y)\right\}^2 - \frac{1}{n}\left\{\sum_{y\in\mathcal{S}_x} f(y)\Pr(\mathcal{K}_{x,h} = y)\right\}^2 \\
&= \frac{1}{n}f(x)\{\Pr(\mathcal{K}_{x,h} = x)\}^2 - \frac{1}{n}f^2(x) + R_n(x;h),
\end{aligned}
$$

with

$$
\begin{aligned}
R_n(x;h) &= \frac{1}{n}\sum_{y\in\mathcal{S}_x\setminus\{x\}} f(y)\left\{\Pr(\mathcal{K}_{x,h} = y)\right\}^2 + \frac{1}{n}f^2(x) \\
&\quad -\frac{1}{n}\left[f(x) + \sum_{y\in\mathcal{S}_x}\{f(y) - f(x)\}\Pr(\mathcal{K}_{x,h} = y)\right]^2. \qquad (22)
\end{aligned}
$$

This quantity becomes negligible under the hypothesis (4) and (5) of discrete associated kernel; i.e. for any $x \in \mathbb{N}$, $R_n(x;h) \to 0$ when $n \to \infty$ and $h = h(n) \to 0$. Indeed, let $y \in \mathcal{S}_x \setminus \{x\}$ we can find a constant $\eta = \eta(y) > 0$ such that

$$\begin{aligned}\Pr(\mathcal{K}_{x,h} = y) &\le \Pr(|\mathcal{K}_{x,h} - x| > \eta)\\ &\le \frac{1}{\eta^2}\mathbb{E}\{(\mathcal{K}_{x,h} - x)^2\} = \frac{1}{\eta^2}[Var(\mathcal{K}_{x,h}) + \{\mathbb{E}(\mathcal{K}_{x,h}) - x\}^2] \to 0 \text{ as } h \to 0,\end{aligned}$$

and for $y = x$ we deduce the asymptotic modal probability $\Pr(\mathcal{K}_{x,h} = x) \to 1$ when $h \to 0$.

Similarly, the variance of $N_n(x;h)$ can be expressed

$$\begin{aligned}Var\{N_n(x;h)\} &= \frac{1}{n}\sum_{y\in S_x}\mathbb{E}(Y_1^2|X_1 = y)f(y)\{\Pr(\mathcal{K}_{x,h} = y)\}^2\\ &\quad -\frac{1}{n}\left\{\sum_{z\in S_x}\mathbb{E}(Y_1|X_1 = z)f(z)\Pr(\mathcal{K}_{x,h} = z)\right\}^2\\ &= \frac{1}{n}\mathbb{E}(Y_1^2|X_1 = x)f(x)\{\Pr(\mathcal{K}_{x,h} = x)\}^2\\ &\quad +\frac{1}{n}\sum_{y\in S_x\setminus\{x\}}\mathbb{E}(Y_1^2|X_1 = y)f(y)\{\Pr(\mathcal{K}_{x,h} = y)\}^2\\ &\quad -\frac{1}{n}\left[f(x)\mathbb{E}(Y_1|X_1 = x) + \sum_{y\in S_x}\{f(y) - f(x)\}\mathbb{E}(Y_1|X_1 = y)\Pr(\mathcal{K}_{x,h} = y)\right]^2\\ &= \frac{1}{n}f(x)\{\Pr(\mathcal{K}_{x,h} = x)\}^2\{Var(Y|X = x) + m^2(x)\}\\ &\quad -\frac{1}{n}(mf)^2(x) + Q_n(x;h),\end{aligned}$$

where

$$\begin{aligned}Q_n(x;h) &= \frac{1}{n}\sum_{y\in S_x\setminus\{x\}}\mathbb{E}(Y^2|X = x)f(y)\{\Pr(\mathcal{K}_{x,h} = y)\}^2 + \frac{1}{n}(mf)^2(x)\\ &\quad -\frac{1}{n}\left[(mf)(x) + \sum_{y\in S_x}\{(mf)(y) - (mf)(x)\}\Pr(\mathcal{K}_{x,h} = y)\right]^2.\end{aligned}$$

The same arguments as those used for $R_n(x;h)$ in (22) might be used to show that $Q_n(x;h)$ goes to 0 as $h \to 0$. Finally, the relationship in (21) together with the above asymptotic expansions and the fact that $Cov\{N_n(x;h), D_n(x;h)\} = O(1/n)$ lead to the desired result (9). ■

5.2 Proof of Theorem 2

The proof follows the same lines as those of Theorem 1. In fact, we write $\widehat{m}_n(x) = H_n(x;h)/F_n(x;h)$ to establish an expression similar to (16) with $H_n(x;h) = n^{-1}\sum_{i=1}^{n}\{r_0(x)/r_0(X_i)\}Y_iK_{x,h}(X_i)$ and $F_n(x;h) = n^{-1}\sum_{j=1}^{n}K_{x,h}(X_j) = D_n(x;h)$.

In order to express the expectation of $\widehat{m}_n(x)$, we first calculate

$$\begin{aligned}\mathbb{E}\{H_n(x;h)\} &= r_0(x)\mathbb{E}[\{r_0(X_1)\}^{-1}Y_1K_{x,h}(X_1)]\\ &= r_0(x)\sum_{z\in S_x}\{r_0(z)\}^{-1}f(z)m(z)\Pr(\mathcal{K}_{x,h}=z)\\ &= r_0(x)\mathbb{E}\{(\omega f)(\mathcal{K}_{x,h})\}\\ &= r_0(x)\left[(\omega f)\{\mathbb{E}(\mathcal{K}_{x,h})\} + \frac{1}{2}Var(\mathcal{K}_{x,h})(\omega f)^{(2)}(x) + o(h)\right].\end{aligned}$$

Thus, under condition (4) of discrete associated kernel, we get

$$\mathbb{E}\{H_n(x;h)\} = (mf)(x) + \frac{1}{2}Var(\mathcal{K}_{x,h})r_0(x)(\omega f)^{(2)}(x) + o(h).$$

In addition, we also have

$$\begin{aligned}Cov\{F_n(x;h), H_n(x;h)\} &= \frac{r_0(x)}{n}\mathbb{E}\left[\{r_0(X_1)\}^{-1}Y_1K_{x,h}^2(X_1)\right]\\ &\quad -\frac{r_0(x)}{n}\mathbb{E}\left[\{r_0(X_1)\}^{-1}Y_1K_{x,h}(X_1)\right]\mathbb{E}\{K_{x,h}(X_2)\}\\ &= O(1/n).\end{aligned}$$

Hence, it suffices to use the same arguments as in the proof of Theorem 1 to obtain the bias of $\widehat{m}_n(x)$.

As for the variance of $\widehat{m}_n(x)$, the proof being quite similar to that of $\widetilde{m}_n(x)$ in Theorem 1, details are therefore omitted.■

5.3 Proof of Theorem 3

In order to establish the results of Theorem 3, the difference $\widehat{\beta} - \beta_0$ will be expressed as an average of i.i.d. variables with mean zero plus remainder term as in (23). Indeed, let P be the generating joint distribution of (X, Y) and P_n its corresponding empirical distribution. We consider functional estimators of β of the form $\widehat{\beta} = T(P_n)$ for which $\beta = T(P)$ realizes the best approximant $r(x;\beta_0)$ to $m(x)$ with respect to some distance measure, for instance the Kullback-Leibler distance. We then introduce the influence function of dimension p (of β)

$$I(X,Y) = \lim_{\epsilon\to 0}[T\{(1-\epsilon)P + \epsilon\delta_{(X,Y)}\} - T(P)]/\epsilon,$$

which has zero mean and finite covariance matrix, and where $\delta_{(X,Y)}$ denotes the unit point mass in (X,Y). Hence, one has

$$\widehat{\beta} - \beta_0 = \frac{1}{n}\sum_{i=1}^{n} I(X_i,Y_i) + \frac{b}{n} + \epsilon_n, \qquad (23)$$

where b/n can be considered as the bias of the estimator $\widehat{\beta}$ and the last term ϵ_n is such that $\mathbb{E}(\epsilon_n) = O(n^{-2})$.

In addition, we note

$$U_i = \frac{r_0(x)}{r_0(X_i)} Y_i K_{x,h}(X_i), \qquad V_i = \frac{r_0(x)}{r_0(X_i)} Y_i K_{x,h}(X_i)\{u_0(x) - u_0(X_i)\}^T,$$

$$W_i = \frac{r_0(x)}{r_0(X_i)} Y_i K_{x,h}(X_i) M(x,X_i), \qquad Z_i = K_{x,h}(X_i),$$

and we then write $U^* = (1/n)\sum_{i=1}^{n} U_i$ and the same for V^*, W^* and Z^*. Thus, it comes that $\widehat{m}_n^0(x) = U^*/Z^*$, $A_n(x) = (\widehat{\beta} - \beta_0)V^*/Z^*$ and $B_n(x) = (\widehat{\beta} - \beta_0)^T W^*(\widehat{\beta} - \beta_0)/Z^*$.

Looking first to the expectation of the estimator $\widehat{m}_n$ such that

$$\mathbb{E}\{\widehat{m}_n^0(x)\} \doteq \mathbb{E}\{\widehat{m}_n^0(x)\} + \mathbb{E}\{A_n(x)\} + \frac{1}{2}\mathbb{E}\{B_n(x)\}.$$

Let us recall that the expectation $\mathbb{E}\{\widehat{m}_n^0(x)\}$ can be deduced from (13) of Theorem 2 as $\mathbb{E}\{\widehat{m}_n^0(x)\} = \text{bias}\{\widehat{m}_n(x)\} + r_0(x)\omega(x)$.

For $A_n(x)$, we reformulate it as shown in Eq. (16). Thus, to calculate $\mathbb{E}\{A_n(x)\}$, we need $\mathbb{E}\{(\widehat{\beta} - \beta_0)V^*\}$, $\mathbb{E}(Z^*)$ and $Cov\{(\widehat{\beta} - \beta_0)V^*, Z^*\}$. Note that the expectation of $Z^* = D_n(x;h)$ is already found in the proof of Theorem 1. From (23) and $\mathbb{E}\{I(X_iY_i)\} = 0$ for all i, we get

$$\begin{aligned}\mathbb{E}\{(\widehat{\beta} - \beta_0)V^*\} &= n^{-1}\mathbb{E}\{V_1 I(X_1,Y_1)\} + n^{-1}b\mathbb{E}(V_1) + O(n^{-2})\\ &= o(hn^{-1}) + O(n^{-2}).\end{aligned}$$

Indeed, using both the discrete Taylor expansion and the condition (4) of the discrete associated kernel, we have successively

$$
\begin{aligned}
\mathbb{E}\{V_1 I(X_1 Y_1)\} &= \mathbb{E}\left[\frac{r_0(x)}{r_0(X_1)} Y_1 K_{x,h}(X_1) I(X_1, Y_1)\{u_0(x) - u_0(X_1)\}^T\right] \\
&= \sum_z \frac{r_0(x)}{r_0(z)} K_{x,h}(z)\{u_0(x) - u_0(z)\}^T f(z) \sum_y y I(z,y) g(y|z) \\
&= r_0(x) \sum_z \Pr(\mathcal{K}_{x,h} = z)\{u_0(x) - u_0(z)\}^T \frac{f(z)q(z)}{r_0(z)} \\
&= r_0(x) \left[u_0(x) - u_0\{\mathbb{E}(\mathcal{K}_{x,h})\}\right]^T s\{\mathbb{E}(\mathcal{K}_{x,h})\} \\
&\quad + \frac{1}{2} r_0(x) Var(\mathcal{K}_{x,h}) \left([u_0(x) - u_0\{\mathbb{E}(\mathcal{K}_{x,h})\}]^T s\{\mathbb{E}(\mathcal{K}_{x,h})\}\right)^{(2)} + o(h) \\
&= -\frac{1}{2} r_0(x) Var(\mathcal{K}_{x,h}) \left[\{u_0^{(2)}\}^T(x) s(x) + 2\{u_0^{(1)}\}^T(x) s^{(1)}(x)\right] + o(h) \\
&= a(x) + o(h),
\end{aligned}
$$

where $a(x) = (-1/2) r_0(x) Var(\mathcal{K}_{x,h}) \left[\{u_0^{(2)}\}^T(x) s(x) + 2\{u_0^{(1)}\}^T(x) s^{(1)}(x)\right]$, $q(x) = \mathbb{E}\{I(X,Y)Y|X = x\}$, $s(x) = f(x)q(x)/r_0(x)$. Hence, under condition (5) on the variance of $\mathcal{K}_{x,h}$, we have $\mathbb{E}\{V_1 I(X_1 Y_1)\}$ and $\mathbb{E}(V_1)$ which are of order $o(h)$. Thus, in the reformulation of $A_n(x)$ as seen in (16), $n^{-1}a(x)$ takes the place of mf such that

$$
\mathbb{E}\{(\widehat{\beta} - \beta_0)V^*\} - n^{-1}a(x) = o(hn^{-1}) + O(n^{-2}).
$$

Then, we express

$$
\begin{aligned}
Cov\{(\widehat{\beta} - \beta_0)V^*, Z^*\} &= Cov(V^* I^*, Z^*) + n^{-1} Cov(bV^*, Z^*) + Cov(\epsilon_n V^*, Z^*) \\
&= o(hn^{-1}) + O(n^{-2}),
\end{aligned}
$$

where

$$
\begin{aligned}
Cov(V^* I^*, Z^*) &= n^{-1}\mathbb{E}(Z_1)\mathbb{E}(V_1 I_1) + O(n^{-2}), \\
n^{-1} Cov(bV^*, Z^*) &= n^{-1} b\mathbb{E}(Z_1)\mathbb{E}(V_1) + O(n^{-2})
\end{aligned}
$$

and $\mathbb{E}(V_1) = o(h)$. In addition, from the decomposition of $A_n(x)$ as in (16), we have

$$
\mathbb{E}[(\widehat{\beta} - \beta_0)V^*\{Z^* - f(x)\}^2] = O(n^{-2} + n^{-1}) + o(h).
$$

Finally, we obtain $\mathbb{E}(A_n) = o(hn^{-1} + h) + O(n^{-2} + n^{-1})$.

In the same manner, we calculate the expectation $\mathbb{E}(B_n) = o(hn^{-1} + h) + O(n^{-2} + n^{-1})$. We omit here to present all the calculus and we give only the mains results. Firstly, we obtain

$$
\begin{aligned}
\mathbb{E}\{(\widehat{\beta} - \beta_0)W^*(\widehat{\beta} - \beta_0)^T\} &= n^{-1} Tr\{\mathbb{E}(W_1)\mathbb{E}(I_1 I_1^T)\} + O(n^{-2}) \\
&= o(hn^{-1}) + O(n^{-2})
\end{aligned}
$$

and $\mathbb{E}(W_1) = o(h)$. Then, we get

$$Cov\{(\widehat{\beta} - \beta_0)W^*(\widehat{\beta} - \beta_0)^T, Z^*\} = O(n^{-2}).$$

Lastly, we find

$$\mathbb{E}[\{(\widehat{\beta} - \beta_0)W^*(\widehat{\beta} - \beta_0)^T\}\{Z^* - f(x)\}^2] = O(n^{-2} + n^{-1}) + o(h)$$

which appears in the decomposition of $B_n(x)$ as in (16). The previous term allow to calculate $\mathbb{E}(B_n)$.

Finally, the expectation of the estimator $\widehat{m}_n$ results in $\mathbb{E}\{\widehat{m}_n(x)\} = \mathbb{E}\{\widehat{m}_n^0(x)\}+ O(n^{-2} + n^{-1}) + o(hn^{-1} + h)$ which provides the desired result asymptotically.

From (15), we express the variance of $\widehat{m}_n(x)$ as

$$\begin{aligned} Var\{\widehat{m}_n(x)\} \doteq & Var\{\widehat{m}_n^0(x)\} + Var\{A_n(x)\} + \frac{1}{4}Var\{B_n(x)\} \\ & + Cov\{\widehat{m}_n^0(x), A_n(x)\} + Cov\{\widehat{m}_n^0(x), B_n(x)\} \\ & + \frac{1}{2}Cov\{A_n(x), B_n(x)\}. \end{aligned}$$

The expression of $Var\{\widehat{m}_n^0(x)\}$ is arleady found in (14) of Theorem 2. For the variance of A_n, by establishing an expression of $Var\{A_n(x)\}$ similarly to (21), it follows that we need to calculate

$$\begin{aligned} Var\{(\widehat{\beta} - \beta_0)V^*\} = & \mathbb{E}\{(\widehat{\beta} - \beta_0)(W^*)^2(\widehat{\beta} - \beta_0)^T\} + O(n^{-2}) \\ = & n^{-1}\mathbb{E}(W_1)\mathbb{E}(I_1 I_1^T)\mathbb{E}(B_1)^T + O(n^{-2}) \\ = & O(n^{-2}) + o(h^2 n^{-1}). \end{aligned}$$

Then, by taking into account the others terms resulting from the presentation of the variance of $A_n(x)$ as in (21), it ensues that $Var\{A_n(x)\} = O(n^{-2} + n^{-1}) + o(h^2 n^{-1} + n^{-1})$. Furthermore, it can be proved

$$\begin{aligned} Cov\{\widehat{m}_n^0(x), A_n(x)\} = & \mathbb{E}\left\{\frac{U^*(\widehat{\beta} - \beta_0)V^*}{(V^*)^2}\right\} - \mathbb{E}\{\widehat{m}_n^0(x)\}\mathbb{E}\{A_n(x)\} \\ = & o(hn^{-1}) + O(n^{-2}) \end{aligned}$$

because in the expression of $\mathbb{E}\{\widehat{m}_n^0 A_n\}$ the more influential terms are

$$\begin{aligned} \mathbb{E}\left\{U^*(\widehat{\beta} - \beta_0)V^*\right\} = & \mathbb{E}(U^* I^* V^*) + n^{-1}b\mathbb{E}(U_1)\mathbb{E}(V_1) + O(n^{-2}) \\ = & o\left(hn^{-1}\right) + O(n^{-2}) \end{aligned}$$

and $Cov\{U^*(\widehat{\beta}-\beta_0)V^*, (Z^*)^2\}$ which can be also shown to be of order $o(hn^{-1})+O(n^{-2})$.

For B_n, without give the details here, it can proved that $Var\{B_n(x)\} = o(h^2n^{-2})+O(n^{-2})$, $Cov\{\widehat{m}_n^0(x), B_n(x)\} = o(hn^{-2})+O(n^{-2})$ and also $Cov\{A_n(x), B_n(x)\} = o(h^2n^{-1}) + O(n^{-2})$.

Finally, we obtain $Var\{\widehat{m}_n(x)\} = Var\{\widehat{m}_n^0(x)\} + O(n^{-2}) + o(hn^{-1} + hn^{-2} + h^2n^{-2} + n^{-1})$ and it ensues the desired asymptotic result. ■

Acknowledgments

We sincerely thank the anonymous referee and Associate Editor for their valuable and constructive comments.

References

Abdous, B. and Kokonendji, C.C. (2009). Consistency and asymptotic normality for discrete associated kernel estimator. *African Diaspora Journal of Mathematics* **8**, 63–70.

Bosq, D. and Lecoutre, J.P. (1987). *Théorie de l'Estimation Fonctionnelle.* Economica, Paris.

Copas, J.B. and Haberman, M.A. (1983). Non-parametric graduation using kernel methods, *Journal of the Institute of Actuaries* **110**, 135–156.

Fan, J., Wu, Y. and Feng, Y. (2009). Local quasi-likelihood with a parametric guide. *The Annals of Statistics* **37**, 4153–4183.

Hardle,W. and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics* **13**, 1465–1481.

Kokonendji, C.C. and Zocchi, S.S. (2010). Extensions of discrete triangular distributions and boundary bias in kernel estimation for discrete functions. *Statistics and Probability Letters* **80**, 1655–1662.

Kokonendji, C.C., Senga Kiessé, T. and Balakrishnan, N. (2009a). Semiparametric estimation for count data through weighted distributions. *Journal of Statistical Planning and Inference* **139**, 3625–3638.

Kokonendji, C.C., Senga Kiessé, T. and Demétrio, C.G.B. (2009b). Appropriate kernel regression on a count explanatory variable and applications. *Advances and Applications in Statistics* **12**, 99-126.

Kokonendji, C.C., Senga Kiessé, T. and Zocchi, S.S. (2007). Discrete triangular distributions and non-parametric estimation for probability mass function. *Journal of Nonparametric Statistics* **19**, 241–254.

Martins-Filho, C., Mishra, S. and Ullah, A. (2008). A class of improved parametrically guided nonparametric regression estimators. *Econometric Reviews* **27**, 542–573.

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and its Applications* **9**, 141–142.

Schumaker, L.L. (1981). *Spline Functions: Basic Theory.* Wiley, New York.

Senga Kiessé, T. (2009). *Nonparametric Approach by Discrete Associated-Kernel for Count Data.* Ph.D. manuscript (in French), University of Pau. URL http://tel.archives-ouvertes.fr/tel-00372180/fr/

Senga Kiessé, T., Libengué, F.G., Zocchi, S.S. and Kokonendji, C.C. (2010). The R package for general discrete triangular distributions. URL http://cran.r-project.org/web/packages/TRIANGG/.

Su, L. and Ullah, A. (2008). Nonparametric prewhitening estimators for conditional quantiles. *Statistica Sinica* **18**, 1131–1152.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26**, 359–372.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–26.