Semiparametric multiple kernel estimators and model diagnostics for count regression functions

Lamia Djerroud, Tristan Senga Kiessé, Smail Adjabi

Semiparametric multiple kernel estimators and model diagnostics for count regression functions

Lamia Djerroud [a*], Tristan Senga Kiessé[b] and Smail Adjabi[a]
[a]Research Unit LaMOS, University of Bejaia, Algeria
[b]UMR SAS, INRA, Agrocampus Ouest, F-35000 Rennes, France

Abstract

This study concerns semiparametric approaches to estimate discrete multivariate count regression functions. The semiparametric approaches investigated consist of combining discrete multivariate nonparametric kernel and parametric estimations such that (i) a prior knowledge of the conditional distribution of model response may be incorporated and (ii) the bias of the traditional nonparametric kernel regression estimator of Nadaraya-Watson may be reduced. We are precisely interested in combination of the two estimations approaches with some asymptotic properties of the resulting estimators. Asymptotic normality results were showed for nonparametric correction terms of parametric start function of the estimators. The performance of discrete semiparametric multivariate kernel estimators studied is illustrated using simulations and real count data. In addition, diagnostic checks are performed to test the adequacy of the parametric start model to the true discrete regression model. Finally, using discrete semiparametric multivariate kernel estimators provides a bias reduction when the parametric multivariate regression model used as start regression function belongs to a neighbourhood of the true regression model.

Keywords: Discrete multivariate kernel; Semiparametric count regression; Cross-validation; Model diagnostics

1 Introduction

The choice of suitable models in regression problems has been largely investigated in the literature. Among various methods, the nonparametric approaches have been shown to be useful to estimate regression models. There is particularly a vast literature on continuous kernel based estimators for regression models, including the popular nonparametric estimator proposed by **?** and **?**. Furthermore, combinations of parametric and nonparametric regression approaches through a parametrically guided nonparametric estimation procedure were investigated, to improve nonparametric Nadaraya-Watson (N-W) estimator

(**?**). The semiparametric kernel estimators resulting from the previous combinations enable (i) to incorporate prior knowledge of the conditional distribution of model response; (ii) to reduce the bias of traditional N-W estimator using continuous kernel while keeping the same variance; see, for example, **?** and **?**, in the case of continuous regression functions. In furtherance of the previous works, an extension of the multiplicative combination of parametric and nonparametric regression approaches was proposed by **?** with univariate discrete kernel estimators, focused on the case of count regression function (crf). The parametrically guided nonparametric regression estimators originated from works on the semiparametric estimation of probability density functions (**?**; **?**) and probability mass functions (pmf) (**?**).

Let us consider independent and identically distributed (iid) pairs of random variables $(Y_i, X_i)_{i=1,2,\ldots,n}$ on $\mathbb{R} \times \mathbb{T}^d$, with $\mathbb{R}$ being the set of real numbers and $\mathbb{T}^d$ denoted the product set $\prod_{j=1}^{d} \mathbb{T}$. Following **?** and **?** works, the two-step semiparametric estimation procedure assumes that the conditional mean $\mathbb{E}(Y_i|X_i) = m(X_i)$ can be expressed either as an additive regression function

$$m^{add}(X_i) = r(X_i, \beta) + \delta(X_i), \tag{1}$$

where $r(\cdot, \beta)$ is a parametric function that depends on $\beta = (\beta_1, \beta_2, \ldots, \beta_d)^\top$ and $\delta(\cdot)$ is a nonparametric correction function. For the additive regression function in (1), the previous two step estimation approach results in an estimator $\widehat{m}_n^{add}(X_i) = r(X_i, \widehat{\beta}) + \widehat{\delta}_n(X_i)$ of $m^{add}(X_i)$ with $\widehat{\delta}_n(X_i)$ being a nonparametric estimation of $\mathbb{E}\{Y_i - r(X_i, \widehat{\beta})\} = \delta(X_i)$ and $\widehat{\beta}$ an estimator of β. This procedure supposes the parametric start model r to be sufficiently close from the regression function m. Thus, the asymptotic bias of $\widehat{m}_n^{add}$ should be improved by splitting the estimation problem of the conditional mean function $\mathbb{E}(Y_i|X_i)$ up into a parametric procedure for r and a kernel-based technique for δ. In the literature, similar approaches were developed to achieve bias reduction without affecting the variance of estimators. For instance, **?** proposed a two-step nonparametric estimation procedure for estimating conditional quantiles based on local linear quantile regression "where, in both steps, nonparametric modeling and estimation are done".

The main contribution of this work is twofold: to present discrete multivariate versions of two-step semiparametric estimators of crf and the corresponding models diagnostics, something that was not done until now to our knowledge. In particular, model diagnostics have been presented for the semiparametric density estimation (**?**) but not yet for the semiparametric regression. Section 2 first briefly present basic notions about discrete multivariate associated kernels. Section 3 presents multivariate additive and multiplicative semiparametric regression estimators. Asymptotic bias and variance of the two semiparametric estimators studied are given. Then, information on model diagnostics are derived by studying the estimated additive and multiplicative correction functions. Model diagnostics are useful to reveal if the parametric start regression model r coincides with the true regression model m at each point $\mathbf{x} \in \mathbb{T}^d$. Section 4 and 5 illustrate the performance of the parametric, non-parametric and semi-parametric approaches presented on simulated and real data respectively. Section 6 contains concluding remarks and research prospects.

2 Discrete multiple associated kernel

Let us consider the target vector $\mathbf{x} = (x_1, x_2, \ldots, x_d)^\top \in \mathbb{T}^d \subseteq \mathbb{N}^d$ and the bandwidth matrix $\mathbf{H} = \mathbf{Diag}(h_{11}, \ldots, h_{dd})$ with $h_{jj} > 0$ such that $\mathbf{H} \equiv \mathbf{H}_n$ goes to the null matrix $\mathbf{0_d}$ when the sample size n goes to ∞. The univariate associated kernel $K^{[j]}_{\mathbf{x}_j,h_{jj}}$ is a pmf associated with a random variable (rv) $\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}$ on support $\mathbb{S}_{\mathbf{x}_j,h_{jj}}$ that contains x_j, meaning that we have:

$$0 \leq K^{[j]}_{\mathbf{x}_j,h_{jj}}(y) = \Pr(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}} = y) \leq 1 \text{ and } \sum_{y \in \mathbb{S}_{\mathbf{x}_j,h_{jj}}} K^{[j]}_{\mathbf{x}_j,h_{jj}}(y) = 1.$$

The underlying idea behind the development of discrete associated kernel is that it must attribute the probability mass closest to one at target x_j, while have a smoothing parameter $h_{jj} > 0$ to take into account the probability mass at points $y \neq x_j$ in the neighbourhood $\mathcal{V}_{x_j}$ of x_j. This idea is traduced through the following behaviour of the $K^{[j]}_{\mathbf{x}_j,h_{jj}}$'s modal probability:

$$K^{[j]}_{\mathbf{x}_j,h_{jj}}(x_j) \to D_{x_j}(x_j) = 1, \text{ with } \sum_{y \in \mathbb{S}_{\mathbf{x}_j,h_{jj}} \setminus \{x_j\}} K^{[j]}_{\mathbf{x}_j,h_{jj}}(y) = 1 - K^{[j]}_{\mathbf{x}_j,h_{jj}}(x_j) \to 0, \text{ as } h_{jj} \text{ goes to } \to 0,$$

where D_{x_j} is the pmf of the univariate Dirac type kernel on support $\mathbb{S}_{\mathbf{x}_j} = \{x_j\}$. Without loss of generality, we assume throughout this work that

$$\forall y \in \mathbb{S}_{\mathbf{x}_j,h_{jj}} \subseteq \mathbb{N}, \exists y_0 \in \mathbb{N} : \forall |y| > y_0, K^{[j]}_{\mathbf{x}_j,h_{jj}}(y) = 0.$$

Then, the univariate $\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}$'s expectation and variance are such that:

$$x_j \in \mathbb{S}_{x_j,h_{jj}} \ (A1), \ \lim_{h_{jj} \to 0} \mathbb{E}(\mathcal{K}^{[j]}_{x_j,h_{jj}}) = x_j \ (A2) \text{ and } \lim_{h_{jj} \to 0} \mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}) = 0 \ (A3).$$

It ensues that the multiple associated kernel $K_{\mathbf{x},\mathbf{H}}(\cdot) = \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(\cdot)$ of rv $\mathcal{K}_{\mathbf{x},\mathbf{H}}$ on support $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \times_{j=1}^d \mathbb{S}_{x_j,h_{jj}}$ is a pmf such that

$$\mathbf{x} \in \mathbb{S}_{\mathbf{x},\mathbf{H}}, \ \mathbb{E}(\mathcal{K}_{\mathbf{x},\mathbf{H}}) = \mathbf{x} + \mathbf{U}(\mathbf{x},\mathbf{H}), \ \mathbb{C}\mathrm{ov}(\mathcal{K}_{\mathbf{x},\mathbf{H}}) = \mathbf{B}(\mathbf{x},\mathbf{H}),$$

where $\mathbf{U}(\mathbf{x},\mathbf{H}) = (u_1(\mathbf{x},\mathbf{H}), \ldots, u_d(\mathbf{x},\mathbf{H}))^\top$ and $\mathbf{B}(\mathbf{x},\mathbf{H}) = (b_{ij}(\mathbf{x},\mathbf{H}))_{i,j=1,\ldots,d}$ tend, respectively, to null vector $\mathbf{0}$ and null matrix $\mathbf{0_d}$ as $\mathbf{H} \to \mathbf{0_d}$.
The previous expression of $K_{\mathbf{x},\mathbf{H}}$'s expectation results from the following development:

$$\begin{aligned}
\mathbb{E}(\mathcal{K}_{\mathbf{x},\mathbf{H}}) &= (x_1, x_2, \ldots, x_d) \prod_{j=1}^d K^{[j]}_{x_j,h_{jj}}(x_{ij}) + \sum_{\mathbf{y} \in \mathbb{S}_{\mathbf{x}_j,h_{jj}} \setminus \{x_j\}} (y_1, y_2, \ldots, y_d) \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(y_j) \\
&= (x_1, x_2, \ldots, x_d) + (x_1, x_2, \ldots, x_d) \left\{ \prod_{j=1}^d K^{[j]}_{x_j,h_{jj}}(x_{ij}) - \mathbf{1} \right\} + J_{\mathbf{H}} \\
&= \mathbf{x} + \mathbf{U}(\mathbf{x},\mathbf{H}),
\end{aligned}$$

with

$$J_{\mathbf{H}} = \sum_{\mathbf{y}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}\setminus\{\mathbf{x}\}} (y_1, y_2, \ldots, y_d) \prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(y_j) \to 0 \text{ then } \mathbf{U}(\mathbf{x},\mathbf{H}) \to 0 \text{ as } \mathbf{H} \to \mathbf{0_d},$$

where $\mathbf{1}$ denotes the unit vector. Then, we successively express the $K_{\mathbf{x},\mathbf{H}}$'s covariance as follows:

$$\begin{aligned}
\mathbb{C}\text{ov}(\mathcal{K}_{\mathbf{x},\mathbf{H}}) &= \mathbb{C}\text{ov}\Big(\prod_{j=1}^{d} \mathcal{K}^{[j]}_{x_j,h_{jj}}\Big) \\
&= \text{Diag}_d\Big\{\text{Var}(\mathcal{K}^{[j]}_{x_j,h_{jj}})\Big\} \\
&= \text{Diag}_d\Big\{\sum_{\mathbf{y_j}\in\mathbb{S}_{\mathbf{x}_j,h_{jj}}} y_j^2 K^{[j]}_{\mathbf{x}_j,h_{jj}}(y_j) - \Big\{\sum_{\mathbf{y_j}\in\mathbb{S}_{\mathbf{x}_j,h_{jj}}\setminus\{x_j\}} y_j K^{[j]}_{\mathbf{x}_j,h_{jj}}(y_j)\Big\}^2\Big\} \\
&= \text{Diag}_d\Big\{x_j^2 K^{[j]}_{\mathbf{x}_j,h_{jj}}(x_j)\{1 - K^{[j]}_{\mathbf{x}_j,h_{jj}}(x_j)\} + G_{x_j,h_{jj}}\Big\} \to \mathbf{0_d} \text{ as } \mathbf{H} \to \mathbf{0_d},
\end{aligned}$$

with

$$G_{x_j,h_{jj}} = \sum_{\mathbf{y_j}\in\mathbb{S}_{\mathbf{x}_j,h_{jj}}\setminus\{x_j\}} y_j^2 K^{[j]}_{\mathbf{x}_j,h_{jj}}(y_j) + x_j^2 K^{[j]}_{\mathbf{x}_j,h_{jj}}(x_j) - \Big\{\sum_{\mathbf{y_j}\in\mathbb{S}_{\mathbf{x}_j,h_{jj}}\setminus\{x_j\}} y_j K^{[j]}_{\mathbf{x}_j,h_{jj}}(y_j)\Big\}^2 \to 0 \text{ as } h_{jj} \to 0.$$

Let us now present two examples of discrete associated kernels.

Example 1
The discrete multiple kernel was proposed for categorical or finite discrete distribution such that

$$K_{\mathbf{x},\mathbf{H}}(\mathbf{y}) = \prod_{j=1}^{d} (1-h_j)^{\mathbb{I}_{y_j=x_j}} \Big(\frac{h_j}{c_j-1}\Big)^{1-\mathbb{I}_{y_j\neq x_j}}, \quad \forall \mathbf{y} \in \mathbb{S}_{\mathbf{x},\mathbf{c}}, \tag{2}$$

with the support $\mathbb{S}_{\mathbf{x},\mathbf{c}}$ being the product $\times_{j=1}^{d}\{0,1,\ldots,c_j-1\}, c_j \in \{2,3,\ldots\}, \forall j = 1,2,\ldots,d$, the bandwidth matrix $\mathbf{H} = \mathbf{Diag}(h_{11},\ldots,h_{dd})$ and $\mathbb{I}_A$ being the indicator function of an event A (**?**). The expectation and covariance of the associated random variable $\mathcal{K}_{\mathbf{x},\mathbf{H}}$ are such that

$$\mathbb{E}(\mathcal{K}_{\mathbf{x},\mathbf{H}}) = \mathbf{x} + \mathbf{H}\Big(1 - \frac{x_1}{c_1-1} + \frac{h_1 c_1}{2}, \ldots, 1 - \frac{x_d}{c_d-1} + \frac{h_d c_d}{2}\Big)^\top = \mathbf{x} + \mathbf{U}(\mathbf{x},\mathbf{H})$$

and

$$\mathbb{C}\text{ov}(\mathcal{K}_{\mathbf{x},\mathbf{H}}) = \mathbf{H}\text{Diag}_d\Big(x_j^2 \frac{c_j^2(1-h_j)-c_j}{(c_j-1)^2} - x_j \frac{c_j^2(1-h_j)-c_j}{c_j-1} + \frac{c_j}{2}\Big(\frac{2c_j-1}{3} - \frac{h_j c_j}{2}\Big)\Big)_j = \mathbf{B}(\mathbf{x},\mathbf{H}),$$

where $\mathbf{U}(\mathbf{x},\mathbf{H})$ and $\mathbf{B}(\mathbf{x},\mathbf{H})$ tend, respectively, to null vector $\mathbf{0}$ and null matrix $\mathbf{0_d}$ as $\mathbf{H} \to \mathbf{0_d}$.

Example 2
We also present the multiple discrete symmetric associated triangular kernels $T_{\mathbf{a};\mathbf{x},\mathbf{H}}(\cdot) = \prod_{j=1}^{d} T^{[j]}_{a_j;x_j,h_{jj}}(\cdot)$ on $\mathbb{S}_{\mathbf{x},\mathbf{a}} = \times_{j=1}^{d}\{x_j, x_j \pm 1, ..., x_j \pm a_j\} = \times_{j=1}^{d}\mathbb{S}_{x_j,a_j}$ such that

$$T^{[j]}_{a_j;x_j,h_{jj}}(y_j) = \frac{(a_j+1)^{h_{jj}} - |y_j - x_j|^{h_{jj}}}{P(a_j,h_{jj})},\ y_j \in \mathbb{S}_{x_j,a_j},$$

with $P(a_j,h_{jj}) = (2a_j+1)(a_j+1)^{h_{jj}} - 2\sum_{k=0}^{a_j} k^{h_{jj}}$, for $(a_j,x_j) \in \mathbb{N}\times\mathbb{T}$ and $h_{jj} > 0$ (**?**). The expectation and covariance of the random variable $\mathcal{T}_{\mathbf{a};\mathbf{x},\mathbf{H}}$ associtated to kernel $T_{\mathbf{a};\mathbf{x},\mathbf{H}}$ are such that

$$\mathbb{E}(\mathcal{T}_{\mathbf{a};\mathbf{x},\mathbf{H}}) = \mathbf{x} \text{ and } \mathbb{C}\text{ov}(\mathcal{T}_{\mathbf{a};\mathbf{x},\mathbf{H}}) \simeq \mathbf{H}\text{Diag}_d\left(\frac{a_j(2a_j^2+3a_j+1)}{3}\log(a_j+1) - 2\sum_{k=1}^{a_j} k^2\log(k)\right)_j + O\Big(\sum_{j=1}^{d} h_{jj}^2\Big).$$

3 Discrete semiparametric multiple kernel regression

3.1 Additive estimator

Let us first recall the definition of discrete nonparametric multivariate kernel estimator of crf on support $\mathbb{T}^d$ included in $\mathbb{N}^d$, the product of set of non-negative integers $\mathbb{N}$ (**?**). Given $(y_i,\mathbf{x}_i)_{i=1,2,...,n} \in \mathbb{R}\times\mathbb{T}^d$ a sequence of iid pairs, the discrete multivariate non-parametric estimator $\widehat{m}_n$ of m is defined as follows:

$$\widehat{m}_n(\mathbf{x};K,\mathbf{H}) = \sum_{i=1}^{n} \frac{y_i \prod_{j=1}^{d} K^{[j]}_{x_j,h_{jj}}(x_{ij})}{\sum_{l=1}^{n}\prod_{j=1}^{d} K^{[j]}_{x_j,h_{jj}}(x_{lj})} = \widehat{m}_n(\mathbf{x}), \tag{3}$$

where $K_{\mathbf{x},\mathbf{H}}$ is the multiple discrete associated kernel with a target vector $\mathbf{x} \in \mathbb{T}^d$ and a bandwidth matrix, with $h_{jj} > 0$, such that $\mathbf{H} \equiv \mathbf{H}_n \to \mathbf{0_d}$ when $n \to \infty$.

Then, the semiparametric multivariate kernel additive estimator of crf in (1) is such that (**?**)

$$\begin{aligned}\widehat{m}_n^{add}(\mathbf{x};K,\mathbf{H}) &= r(\mathbf{x},\widehat{\beta}) + \widehat{\delta}_n(\mathbf{x};K,\mathbf{H}) \\ &= r(\mathbf{x},\widehat{\beta}) + \sum_{i=1}^{n} \frac{\{y_i - r(X_i,\widehat{\beta})\}\prod_{j=1}^{d} K^{[j]}_{x_j,h_{jj}}(x_{ij})}{\sum_{l=1}^{n}\prod_{j=1}^{d} K^{[j]}_{x_j,h_{jj}}(x_{lj})} = \widehat{m}_n^{add}(\mathbf{x}).\end{aligned} \tag{4}$$

Discrete semiparametric and nonparametric regression estimators of N-W type can be viewed as the minimizer of a general loss function (**?**). Thus, the semiparametric additive estimator in equation (4) can be defined by

$$\widehat{m}_n^{add} = \arg\min_{M} \frac{1}{n}\sum_{i=1}^{n}\left(\begin{bmatrix} y_i - r(\mathbf{x}_i,\widehat{\beta}) & M - r(\mathbf{x},\widehat{\beta})\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix}\right)^2 \prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i) \tag{5}$$

and the nonparametric estimator of N-W type in equation (3) is such that $\widehat{m}_n = \arg\min_M \frac{1}{n}\sum_{i=1}^n \left(\begin{bmatrix} y_i & M \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)^2 \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i)$ (similar to the approach of **?**, for semi-parametric density estimation).

Hereafter, we fix the parametric start and put $r_0(\cdot) = r(\cdot,\beta_0)$ on the basis of a goodness-of-fit test or any a priori knowledge about m. The $\widehat{m}_n^{add}$'s bias and variance can be presented as follows.

Theorem 1
Consider the target vector $\mathbf{x}=(x_1,x_2,\cdots,x_d)^\top \in \mathbb{T}^d \subseteq \mathbb{Z}^d$ and the bandwidth $\mathbf{H}=\mathbf{Diag}\,(h_{11},h_{22},\cdots,h_{dd}) \to 0_d$ as $n\to\infty$ with $h_{jj}>0$. Furthermore, consider f_j the univariate pmf of rv X_j such that $f_j(x_j)=\Pr(X_j=x_j)>0$ for $j=1,2,\ldots,d$. Then, the semiparametric estimator $\widehat{m}_n^{add}(\mathbf{x})$ of $m^{add}(\mathbf{x}) = r_0(\mathbf{x})+\delta(\mathbf{x})$ with a discrete multiple associated kernel has bias and variance given by

$$\mathrm{Bias}\{\widehat{m}_n^{add}(\mathbf{x})\} \approx \frac{1}{2}\sum_{j=1}^d \left\{\delta_{jj}^{(2)}(\mathbf{x}) + 2\delta_j^{(1)}(\mathbf{x})\left(\frac{f_j^{(1)}}{f}\right)(\mathbf{x})\right\}\mathrm{Var}(\mathcal{K}^{[j]}_{x_j,h_{jj}}),$$

and

$$\mathrm{Var}\{\widehat{m}_n^{add}(\mathbf{x})\} \approx \frac{\mathbb{E}(Y_1^2|\mathbf{x}_1=\mathbf{x}) - f(\mathbf{x})\mathbb{E}^2(Y_1|\mathbf{x}_1=\mathbf{x})}{nf(\mathbf{x})}\left\{\prod_{j=1}^d \Pr(\mathcal{K}^{[j]}_{x_j,h_{jj}}=x_j)\right\}^2$$

where $f_j^{(1)}$, $\delta_j^{(1)}$ and $\delta_{jj}^{(2)}$ are j-partial finite differences for $j=1,2,\ldots,d$, defined in the sense of any univariate count component $g:\mathbb{N}\mapsto\mathbb{R}$ by

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

and

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

Proof. The proof is postponed to Appendix.

From $\widehat{m}_n^{add}(\mathbf{x}) = r_0(\mathbf{x}) + \widehat{\delta_n}(\mathbf{x})$, $\widehat{\delta_n}$'s bias and variance are given by

$$\mathbb{E}\left\{\widehat{\delta_n}(\mathbf{x})\right\} = \delta(\mathbf{x}) + \mathrm{Bias}\{\widehat{m}_n^{add}(\mathbf{x})\} \text{ and } \mathrm{Var}\left\{\widehat{\delta_n}(\mathbf{x})\right\} = \mathrm{Var}\{\widehat{m}_n^{add}(\mathbf{x})\},\ \mathbf{x}\in\mathbb{T}^d. \tag{8}$$

Remark 1
Consider $(y_i,\mathbf{x}_i)_{i=1,2,\ldots,n}$ a sequence of iid pairs on $\mathbb{R}\times\mathbb{T}^d$. Applying the parametrically guided nonparametric estimation procedure for the crf $m^{mult}(X_i) = r(X_i,\beta)\times\omega(X_i)$ results in the multiplicative estimator such that

$$\widehat{m}_n^{mult}(\mathbf{x};K,\mathbf{H}) = r(\mathbf{x},\widehat{\beta})\times\widehat{\omega}_n(\mathbf{x};K,\mathbf{H}) = r(\mathbf{x},\widehat{\beta})\times\sum_{i=1}^n \frac{y_i\prod_{j=1}^d K^{[j]}_{x_j,h_{jj}}(\mathbf{x}_i)}{r(\mathbf{x}_i,\widehat{\beta})\sum_{l=1}^n\prod_{j=1}^d K^{[j]}_{x_j,h_{jj}}(\mathbf{x}_j)} =: \widehat{m}_n^{mult}(\mathbf{x}), \tag{9}$$

where $K_{\mathbf{x},\mathbf{H}}$ is the multiple discrete associated kernel with $\mathbf{x} \in \mathbb{T}^d$ being a target vector and $\mathbf{H}$ a bandwidth matrix. The parametric function $r(\cdot, \beta)$ depends on $\beta = (\beta_1, \beta_2, \ldots, \beta_d)^\top$ and the nonparametric function $\omega(\cdot)$ is the multiplicative correction factor.

Similar to equation (5), it can be shown that the semiparametric multiplicative estimator in equation (9) is defined by

$$\widehat{m}_n^{mult} = \arg\min_M \frac{1}{n} \sum_{i=1}^{n} \left(\left[y_i - r(\mathbf{x}_i, \widehat{\beta}) \quad M - r(\mathbf{x}, \widehat{\beta}) \right] \begin{bmatrix} \frac{r(\mathbf{x}, \widehat{\beta})}{r(\mathbf{x}_i, \widehat{\beta})} \\ -1 \end{bmatrix} \right)^2 \prod_{j=1}^{d} K_{x_j, h_{jj}}^{[j]}(\mathbf{x}_i).$$

Furthermore, the $\widehat{m}_n^{mult}$'s bias and variance can be presented as follows with a fixed parametric start $r_0(\cdot) = r(\cdot, \beta_0)$.

$$\text{Bias}\{\widehat{m}_n^{mult}(\mathbf{x})\} \approx \frac{1}{2} \sum_{j=1}^{d} \left\{ r_0(\mathbf{x}) \omega_{jj}^{(2)}(\mathbf{x}) + 2 r_0(\mathbf{x}) \omega_j^{(1)}(\mathbf{x}) \left(\frac{f_j^{(1)}}{f} \right)(\mathbf{x}) \right\} \text{Var}(\mathcal{K}_{x_j, h_{jj}}^{[j]}), \tag{10}$$

and

$$\text{Var}\left\{ \widehat{m}_n^{mult}(\mathbf{x}) \right\} \approx \frac{\mathbb{E}(Y_1^2 | \mathbf{x}_1 = \mathbf{x}) - f(\mathbf{x}) \mathbb{E}^2(Y_1 | \mathbf{x}_1 = \mathbf{x})}{n f(\mathbf{x})} \left\{ \prod_{j=1}^{d} \Pr(\mathcal{K}_{x_j, h_{jj}}^{[j]} = x_j) \right\}^2, \tag{11}$$

where $f^{(1)}$, $\omega^{(1)}$ and $\omega^{(2)}$ are finite differences as defined in (6) and (7) (see Appendix for the proof). From $\widehat{m}_n^{mult}(\mathbf{x}) = r_0(\mathbf{x}) \times \widehat{\omega}_n(\mathbf{x})$, one can directly derive $\widehat{\omega}_n$'s expectation and variance by using $\widehat{m}_n^{mult}$'s bias and variance as follows:

$$\mathbb{E}\{\widehat{\omega}_n(\mathbf{x})\} = \omega(\mathbf{x}) + \frac{1}{r_0(\mathbf{x})} \text{Bias}\{\widehat{m}_n^{mult}(\mathbf{x})\} \text{ and } Var\{\widehat{\omega}_n(\mathbf{x})\} = \frac{1}{r_0^2(\mathbf{x})} Var\{\widehat{m}_n^{mult}(\mathbf{x})\}, \ \mathbf{x} \in \mathbb{T}^d. \tag{12}$$

Finally, note that the discrete nonparametric multivariate kernel estimator of m in equation (3) has the same variance that the two discrete semiparametric multivariate kernel estimators studied and a bias expression such that (**?**):

$$\text{Bias}\{\widehat{m}_n(\mathbf{x})\} \approx \frac{1}{2} \sum_{j=1}^{d} \left\{ m^{(2)}(\mathbf{x}) + 2 m^{(1)}(\mathbf{x}) \left(\frac{f_j^{(1)}}{f} \right)(\mathbf{x}) \right\} \text{Var}(\mathcal{K}_{x_j, h_{jj}}^{[j]}).$$

Thus, the difference between the performance of the three discrete multivariate kernel estimators comes by comparing their respective bias. For instance, by using the discrete semiparametric multivariate multiplicative kernel estimator provides a smaller bias than by using the discrete nonparametric multivariate kernel estimator if the parametric start function r_0 is such that

$$\left| r_0(\mathbf{x}) \omega_{jj}^{(2)}(\mathbf{x}) + 2 r_0(\mathbf{x}) \omega_j^{(1)}(\mathbf{x}) \left(\frac{f_j^{(1)}}{f} \right)(\mathbf{x}) \right| < \left| m^{(2)}(\mathbf{x}) + 2 m^{(1)}(\mathbf{x}) \left(\frac{f_j^{(1)}}{f} \right)(\mathbf{x}) \right|.$$

Choices of start functions r_0 that influence the performance of the three discrete multivariate kernel estimators are illustrated through simulations (Section 3).

3.2 Model diagnostics

Studying the estimated additive (respectively, multiplicative) correction function is useful to provide information for model diagnostics. The additive (resp., multiplicative) correction function should equal zero (resp., one), if the parametric start regression model coincides with the true regression model. The model adequacy can be checked by looking at a plot of the correction function to see if $\delta(\mathbf{x}) = 0$ (resp., $\omega(\mathbf{x}) = 1$) is reasonable or not, with a confidence interval at each point $\mathbf{x}$. This plot should allow to spot easily where misspecification is locally the largest.

For $\mathbf{x} \in \mathbb{T}^d$, a graphical goodness-of-fit emerges from the results on expectation and variance of $\widehat{\delta_n}$ (resp. $\widehat{\omega_n}$) by plotting the following function $Z(\mathbf{x})$ against $\mathbf{x}$:

$$Z^{add}(\mathbf{x}) = \frac{\widehat{\delta_n}(\mathbf{x}) - \mathbb{E}\{\widehat{\delta_n}(\mathbf{x})\}}{[\mathrm{Var}\{\widehat{\delta_n}(\mathbf{x})\}]^{1/2}}.$$

The model $Z(\mathbf{x})$ follows a standardized normal distribution $\mathcal{N}(0;1)$, when the parametric start coincides with the true regression model, i.e. $m(\cdot) \equiv r_0(\cdot)$. In this situation, we get $Z(\mathbf{x}) \in [-1.96; 1.96]$ about 95% of the time. Assuming that the parametric regression model is such that $\delta(\cdot) = 0$, $\widehat{\delta_n}$'s bias and variance in equations (8) result in

$$Z^{add}(\mathbf{x}) = \frac{\{nf(\mathbf{x})\}^{1/2}\widehat{\delta_n}(\mathbf{x})}{\Big\{\mathbb{E}(Y_1^2|\mathbf{x}_1 = \mathbf{x}) - f(\mathbf{x})\mathbb{E}^2(Y_1|\mathbf{x}_1 = \mathbf{x})\Big\}^{1/2} \prod_{j=1}^d \Pr(\mathcal{K}^{[j]}_{x_j,h_{jj}} = x_j)}.$$

In addition, we formulate the following result on asymptotical normality of $\widehat{\delta_n}$. Without loss of generality, the asymptotic normality is established for the discrete associated kernels with a modal probability satisfying, as $h_{jj} \to 0$:

$$\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}} = \mathbf{x}) \simeq 1 - \Pr(\mathcal{K}_{\mathbf{x}} = \mathbf{x}) \prod_{j=1}^d h_{jj} + o\Big(\sum_{j=1}^d h_{jj}^2\Big) \to 1. \quad (A4)$$

For instance, as $h_{jj} \to 0$, that concerns the multiple discrete associated kernel in Example 1 such that

$$\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}} = \mathbf{x}) = \prod_{j=1}^d (1 - h_j) \simeq 1 - \prod_{j=1}^d (-1)^{d+1} h_{jj} + o\Big(\sum_{j=1}^d h_{jj}^2\Big)$$

and the multiple discrete symmetric associated triangular kernels in Example 2 such that

$$\Pr(\mathcal{T}_{\mathbf{a};\mathbf{x},\mathbf{H}} = \mathbf{x}) \;=\; \prod_{j=1}^d \frac{(a_j+1)^{h_{jj}}}{P(a_j, h_{jj})} \simeq \prod_{j=1}^d \Big[1 - h_{jj}A(a_j) + O(h_{jj}^2)\Big] \simeq 1 - \prod_{j=1}^d (-1)^{d+1} h_{jj} A(a_j) + o\Big(\sum_{j=1}^d h_{jj}^2\Big),$$

with $A(a_j) = 2a_j \log(a_j + 1) - 2\sum_{k=1}^{a_j} \log(k)$.

Theorem 2
For any fixed $x \in \mathbb{N}$, under assumptions (A1)-(A4) on discrete associated kernel, the estimator $\widehat{\delta_n}(x)$ converges in distribution to the normal law as follows, as $n \to \infty$:

$$\sqrt{n \prod_{j=1}^{d} h_{jj}^{-2}} \left\{ \widehat{\delta_n}(x) - \delta(x) \right\} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2 \{\Pr(\mathcal{K}_{\mathbf{x}} = \mathbf{x})\}^2}{f(\mathbf{x})}\right).$$

Proof. See Appendix

For estimator $\widehat{\omega_n}$ of multiplicative correction factor, assuming that the parametric regression model $r_0(\cdot)$ is true i.e. $\omega(\cdot) = 1$, $\widehat{\omega_n}$'s bias and variance in equations (12) result in the following expression of Z:

$$Z^{mult}(\mathbf{x}) = \frac{r_0(\mathbf{x})\{nf(\mathbf{x})\}^{1/2}[\widehat{\omega_n}(\mathbf{x}) - 1]}{\left\{\mathbb{E}(Y_1^2|\mathbf{x}_1 = \mathbf{x}) - f(\mathbf{x})\mathbb{E}^2(Y_1|\mathbf{x}_1 = \mathbf{x})\right\}^{1/2} \prod_{j=1}^{d} \Pr(\mathcal{K}^{[j]}_{x_j,h_{jj}} = x_j)}.$$

Similar to the semiparametric additive estimator, we formulate the following theorem on the normality of $\widehat{\omega_n}$.

Theorem 3
For any fixed $x \in \mathbb{N}$, under assumptions (A1)-(A4) on discrete associated kernel, the estimator $\widehat{\omega_n}(x)$ converges in distribution to the normal law as follows, as $n \to \infty$:

$$\sqrt{n \prod_{j=1}^{d} h_{jj}^{-2}} \left\{ \widehat{\omega_n}(x) - \omega(x) \right\} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2 \{\Pr(\mathcal{K}_{\mathbf{x}} = \mathbf{x})\}^2}{r_0^2(\mathbf{x}) f(\mathbf{x})}\right).$$

Proof. See Appendix

4 Simulations

This section illustrates the performances of discrete nonparametric and semiparametric multivariate kernel estimators through simulations. Two main issues of the discrete kernel method are the choices of kernel and bandwidth. For the kernel choice, we consider the multiple discrete symmetric associated triangular kernels given in Example 2. We assume a constant value for the parameter $a_j = a = 1, j = 1, 2, \ldots, d$ since it was shown to generally give the better estimation.

For $\mathbf{x} \in \mathbb{T}^d$, when considering this multiple associated kernel the functions Z for model diagnostics were given by

$$Z^{mult}(\mathbf{x}) = \frac{r_0(\mathbf{x})\{nf(\mathbf{x})\}^{1/2}[\widehat{\omega_n}(\mathbf{x}) - 1]}{\prod_{i=1}^{d} \left\{\frac{(a+1)^{h_{jj}}}{P(a,h_{jj})}\right\} \left[\mathbb{E}(Y_1^2|\mathbf{x}_1 = \mathbf{x}) - f(\mathbf{x})\mathbb{E}^2(Y_1|\mathbf{x}_1 = \mathbf{x})\right]^{1/2}}$$

and

$$Z^{add}(\mathbf{x}) = \frac{\{nf(\mathbf{x})\}^{1/2}\widehat{\delta}_n(\mathbf{x})}{\prod_{i=1}^{d}\left\{\frac{(a+1)^{h_{jj}}}{P(a,h_{jj})}\right\}\left[\mathbb{E}(Y_1^2|\mathbf{x}_1=\mathbf{x}) - f(\mathbf{x})\mathbb{E}^2(Y_1|\mathbf{x}_1=\mathbf{x})\right]^{1/2}}.$$

For the bandwidth matrix choice, the least squared cross-validation criterion (LSCV) was applied such that $\mathbf{H}_{cv} = \arg\min_{\mathbf{H}>0} \mathrm{LSCV}(\mathbf{H})$ with

$$\mathrm{LSCV}(\mathbf{H}) = \frac{1}{n}\sum_{k=1}^{n}\{y^k - \widehat{m}_{n,-k}(\mathbf{x}^k, \mathbf{H})\}^2,$$

where $\widehat{m}_{n,-k}$ is the multivariate estimator of the regression function calculated from all observations except the observation $\mathbf{x}^k$. The LSCV is an extension of univariate cross-validation criterion to multivariate case.

The following target regression function is proposed with dimension $d = 2$:

$$m(x_1, x_2) = 0.31 - 0.04(x_1 + x_2) + 0.32\log(x_1 x_2) \tag{13}$$

Simulations were carried out for $N = 400$ replications of sizes n=20, 50, 100, 200 and 500. For the semiparametric approach, three parametric models were used as the start function and the nonparametric approach consisted of the estimator in equation (3). The parametric start models are given by

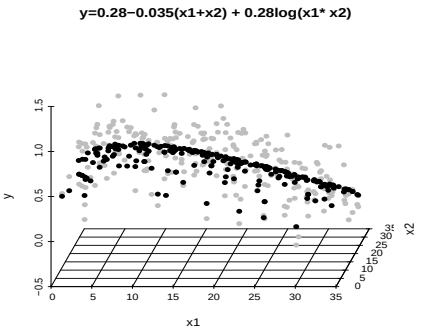
1. $r_0^{(1)}(x_1, x_2) = 0.28 - 0.035(x_1 + x_2) + 0.28\log(x_1 \times x_2)$,
2. $r_0^{(2)}(x_1, x_2) = 0.76 - 0.01(x_1 + x_2)$,
3. $r_0^{(3)}(x_1, x_2) = 0.76 - 0.1\log(x_1 \times x_2)$.

The first start model $r_0^{(1)}$ is expressly chosen to be closest to the target regression function m in equation (13), while the third start model $r_0^{(3)}$ is expressly chosen to be furthest from m. Figure 1 graphically illustrates the models $r_0^{(k)}, k = 1, 2, 3$, for an exemplary run with sample size $n = 250$ simulated from the target regression function.

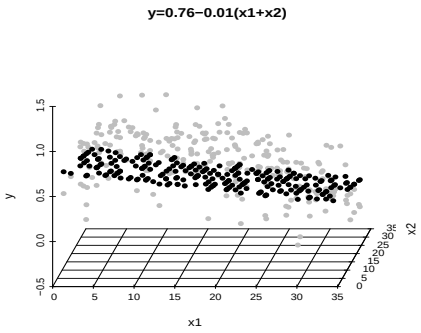
For each replication $j = 1, \ldots, N$, the performance of the three discrete multivariate kernel estimators studied was evaluated using the average squared error (ASE) given by

$$\mathrm{ASE}_j(\mathbf{H}) = \frac{1}{n}\sum_{i=1}^{n}[\widehat{m}^{(j)}(\mathbf{x}_i) - m(\mathbf{x}_i)]^2,$$

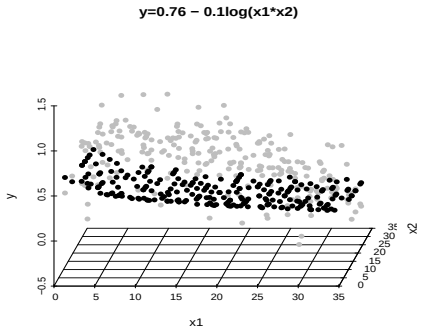
where $\widehat{m}^{(j)}$ is the j-th estimation of the simulated count regression model m in (13).



(a) First parametric model



(b) Second parametric model



(c) Third parametric model

Figure 1: *First, second and third parametric estimates (black points) for data simulated from target regression function m in (13) with $n = 250$ (gray points).*

4.1 First parametric model

Performance of discrete multivariate estimators in term of ASE and bias. Table 1 presents the mean ASE ($\overline{\text{ASE}} = (1/N)\sum_{j=1}^{N} \text{ASE}_j$) of nonparametric, semiparametric and first parametric models, as well as the bias of kernel estimators. The purely parametric estimation using the parametric model r_0^1 outperformed the purely nonparametric kernel estimation and the semiparametric kernel estimation using first parametric start $r_0^{(1)}$ to estimate count data simulated from the target regression model m in equation (13). This is due to the fact that the first parametric model is fairly close to the true model (see Figure 1.a). That also results in close performances of the semiparametric and nonparametric approaches, even if semiparametric estimators had a bias that was slightly smaller than the nonparametric estimator. The parametric start model added a valuable prior information of the true model.

	Sample size n	$\overline{ASE} \times 10^3$	Bias
Param. estimator	20	4.931	
	50	4.997	
	100	5.130	
	250	5.059	
	500	5.001	
Semip.mult. estimator	20	25.894	0.346
	50	37.040	4.084
	100	40.182	7.556
	250	37.325	16.363
	500	32.515	34.230
Semip. add. estimator	20	25.894	0.356
	50	37.018	4.089
	100	40.217	7.296
	250	37.366	16.424
	500	32.515	34.522
Nonp. estimator	20	25.894	0.394
	50	37.016	4.119
	100	40.223	7.408
	250	37.362	17.220
	500	32.515	35.220

Table 1: *$\overline{ASE}$ and bias calculated for model in (13) by using nonparametric and semiparametric regression estimators and the first parametric model.*

Model diagnostics. We present the mean and variance of nonparametric estimated additive and multiplicative correction factors (respectively, $\widehat{\delta}_n$ and $\widehat{\omega}_n$) to check model adequacy (Table 2). We observed that $\widehat{\delta}_n$'s expectation was close to zero and $\widehat{\omega}_n$'s expectation was around one as n increased. However, the multiplicative correction factor estimator $\widehat{\omega}_n$ had a larger variance around the mean, in particular for large sample sizes $n = \{250, 500\}$. Hence, for discrete semiparametric multivariate additive kernel estimator, model diagnostics Z were illustrated in Figure 2 for an exemplary run with simulated sample size $n = 250$. Resulting $Z(\mathbf{x})$-values lied within the interval $[-1.96, 1.96]$ about 95.6% of the time, meaning that $Z(\mathbf{x})$ is approximately distributed as standard normal for each target $\mathbf{x}$. That was confirmed by the Shapiro-Wilk nomality test (W-statistic= 0.995 and p-value= 0.539). This suggests that it would be of interest to consider the first parametric model start for modeling these data.

n	$\mathbb{E}(\widehat{\delta}_n)$	$\text{Var}(\widehat{\delta}_n)$	$\mathbb{E}(\widehat{\omega}_n)$	$\text{Var}(\widehat{\omega}_n)$
20	0.002483	0.001004	0.554477	0.116559
50	0.001240	0.000582	0.548013	0.151298
100	0.001921	0.000891	0.582551	0.894777
250	0.001607	0.000726	0.6306623	0.939676
500	0.002508	0.001281	0.683597	0.989539

Table 2: *Mean and variance of estimated additive and multiplicative correction factors (respectively, $\widehat{\delta}_n$ and $\widehat{\omega}_n$) calculated using the first parametric model with $n = 250$.*

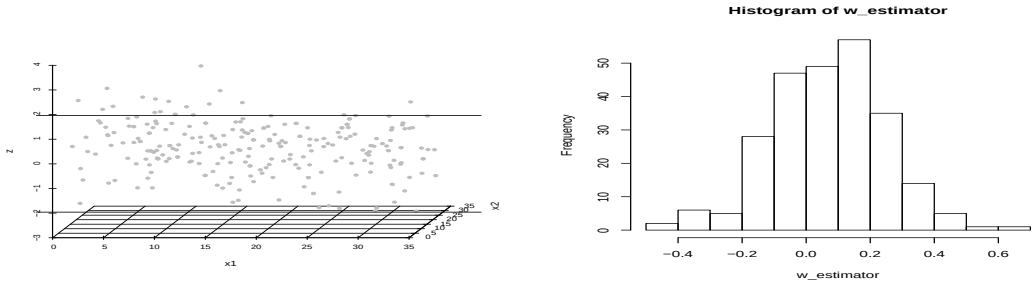


Figure 2: *$Z(\mathbf{x})$-values associated with results of the semiparametric multivariate additive kernel estimator using first parametric start model with $n = 250$. The histogram of $\widehat{\omega}_n$.*

4.2 Second parametric model

Performance of discrete multivariate estimators in term of ASE and bias. When using the second parametric model, the discrete semiparametric multivariate kernel estimators provided the better results to estimate count regression data simulated from model (13) (Table 3). To introduce the additive and multiplicative correction factors in the semiparametric procedure improved the results provided by the nonparametric procedure. The bias of both semiparametric estimators was generally slightly smaller than the bias of the nonparametric estimator, as the sample size increased. This is due to the fact that the second parametric start model is not so far from the target regression model (see Figure 1.b).

Model diagnostics. Table 4 presents mean and variance of estimated additive and multiplicative correction factors (respectively, $\widehat{\delta}_n$ and $\widehat{\omega}_n$) for data simulated from model (13) to check model adequacy. Similar to the results of the previous case, the $\widehat{\delta}_n$'s expectation went to zero and $\widehat{\omega}_n$'s expectation varied around one as the sample size n increased. The additive correction factor estimation δ_n had less variability around its mean than the multiplicative correction factor estimation ω_n. Then, for semiparametric additive kernel estimator, the graphic of $Z(\mathbf{x})$-values for one simulated sample size $n = 250$ shown that $Z(\mathbf{x})$-values lied within the interval $[-1.96, 1.96]$ about 94% of the time (Figure 3). The hypothesis of normality was not rejected by the Shapiro-Wilk test but with a smaller p-value than when using the first parametric model (W-statistic= 0.994 and p-value= 0.491). Finally, the second parametric model could be accepted as an appropriate start function for modeling data simulated from model (13).

	Sample size n	$\overline{ASE} \times 10^3$	Bias
Param. estimator	20	54.677	
	50	56.652	
	100	55.398	
	250	56.003	
	500	55.668	
Semip.mult. estimator	20	40.338	1.996
	50	37.721	7.551
	100	39.840	16.425
	250	33.844	43.352
	500	29.807	85.620
Semip. add. estimator	20	40.339	1.996
	50	37.675	7.550
	100	39.781	16.426
	250	33.863	43.385
	500	29.802	85.620
Nonp. estimator	20	40.338	1.998
	50	37.653	7.557
	100	39.760	16.412
	250	33.901	43.997
	500	30.006	86.281

Table 3: *$\overline{ASE}$ and bias calculated for model in (13) by using nonparametric and semiparametric regression estimators and the second parametric model.*

n	$\mathbb{E}(\widehat{\delta_n})$	$\text{Var}(\widehat{\delta_n})$	$\mathbb{E}(\widehat{\omega_n})$	$\text{Var}(\widehat{\omega_n})$
20	0.004145	0.002226	0.557540	0.479085
50	0.004630	0.001250	0.742093	0.178449
100	0.002855	0.001181	0.752624	0.293568
250	0.002956	0.001347	0.691361	0.244442
500	0.003241	0.001167	0.684382	0.229597

Table 4: *Mean and variance of estimated additive and multiplicative correction factors (respectively, $\widehat{\delta_n}$ and $\widehat{\omega_n}$) calculated using the second parametric model with $n = 250$.*

4.3 Third parametric model

Performance of discrete multivariate estimators in term of ASE and bias. Contrary to the previous case, the nonparametric estimator using discrete triangular kernel outperformed the parametric and semiparametric models to estimate count data simulated from regression model (13) (Table 5). The semiparametric estimates obtained when introducing the additive and multiplicative correction factors did not improve the purely nonparametric estimates. The performance of the nonparametric approaches in comparison with the semiparametric approach could be attributed to the inadequacy of the third parametric start to estimate the model (13) (see Figure 1.c). In this case, the bias of the nonparametric estimator was slightly smaller better than the bias of multiplicative and additive semiparametric estimators, as the sample size n increased. The parametric start model did not add a valuable prior information of the true model.

*Model diagnostics.*Table 6 presents mean and variance of estimated additive and multiplicative correction factors (respectively, $\widehat{\delta_n}$ and $\widehat{\omega_n}$) for data simulated from model (13) to

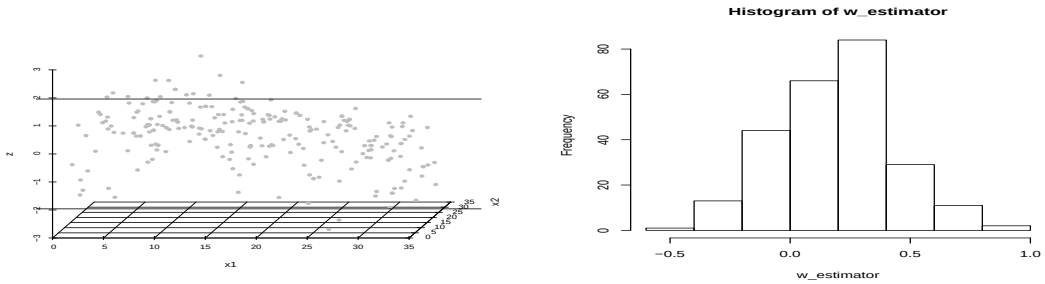


Figure 3: *The $Z(\mathbf{x})$-values associated with results of semiparametric additive kernel estimator using the second parametric model with $n = 250$. The histogram of $\widehat{\omega}_n$.*

	Sample size n	$\overline{ASE} \times 10^3$	Bias
	20	164.100	
	50	164.196	
Param. estimator	100	165.761	
	250	165.397	
	500	158.393	
	20	40.014	5.517
	50	39.091	15.712
Semip.mult. estimator	100	38.036	34.026
	250	33.076	86.504
	500	28.378	119.693
	20	40.015	5.517
	50	39.091	15.723
Semip. add. estimator	100	38.048	34.002
	250	33.070	86.531
	500	28.294	119.517
	20	40.013	5.495
	50	39.089	15.602
Nonp. estimator	100	38.077	34.098
	250	33.023	87.033
	500	28.255	118.127

Table 5: *$\overline{ASE}$ and bias calculated for model in (13) by using nonparametric and semiparametric regression estimators and the third parametric model.*

check model adequacy. Similar to the results of the two previous cases, for semiparametric additive kernel estimator, the graphic of $Z(\mathbf{x})$-values for one simulated sample size $n = 250$ shown that $Z(\mathbf{x})$-values lied within the interval $[-1.96, 1.96]$ about 93% of the time (Figure 4). The normality was not rejected by the Shapiro-Wilk nomality test but the p-value obviously decreased, with regards to the results obtained when using the two first parametric model (W-statistic= 0.993 and p-value= 0.339). In this case, the third parametric start model would not be of interest for modeling these data. Finally, the nonparametric estimator was found to be more appropriate for the data simulated from model (13).

n	$\mathbb{E}(\widehat{\delta_n})$	$\mathrm{Var}(\widehat{\delta_n})$	$\mathbb{E}(\widehat{\omega_n})$	$\mathrm{Var}(\widehat{\omega_n})$
20	-0.001807	0.002897	0.500510	0.309216
50	-0.000611	0.001797	0.856860	0.142295
100	-0.003355	0.001900	0.776986	0.148346
250	-0.003118	0.002138	0.816575	0.196068
500	-0.002805	0.001853	0.841161	0.179138

Table 6: *Mean and variance of estimated additive and multiplicative correction factors (respectively, $\widehat{\delta_n}$ and $\widehat{\omega_n}$) calculated using the third parametric model with $n = 250$.*

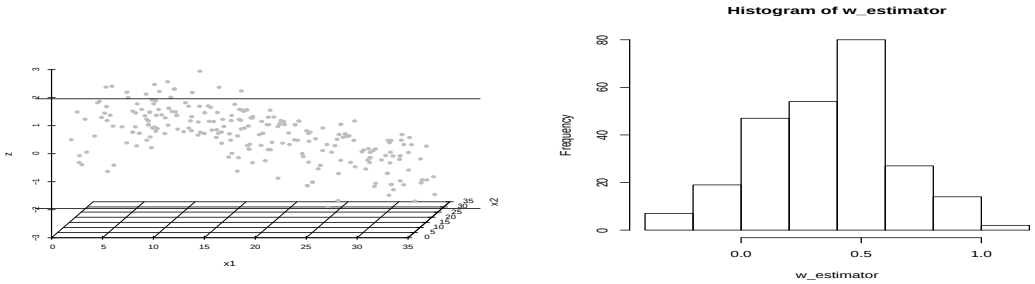


Figure 4: *The $Z(\mathbf{x})$-values associated with results of semiparametric additive kernel estimator of data simulated from model (13) with $n = 250$. The histogram of $\widehat{\omega_n}$.*

5 Real data

The real count data set concerns the study of households' joint choice of the number of leisure trips and number of total nights spent on these trips (**?**). The independent monthly samples, 1990-1996, for Swedish leisure travel are used to estimate models for trips to one of the largest city regions in Sweden: Stockholm. The numbers of the trips to the specific location is denoted by x_1 and the total number of nights on these trips by x_2. The response variable y is the probabilities of each couple. A feature of the data is that an individual making one (two) trips has, at least, one (two) overnight stays in total, and so on. It was considered that trip destinations are observed for at most two trips. Area $I = \{(1,1),(1,2),\ldots,(1,30),(2,2),(2,3),\ldots,(2,30)\}$ contains all possible outcomes that can be observed (Table 7). The area is bounded to the right at 30 nights.
We applied nonparametric and semiparametric regression estimators using discrete multivariate symmetric triangular kernel with $a = 1$ to approximate empirical frequencies for the considered samples, in comparison with the parametric logarithmic model $y_i = \theta_1 + \theta_2 x_i + \theta_2 log(x_i) + \epsilon_i,\ \ x_i \in \mathbb{N} \setminus \{0\}$. In addition, the LSCV was used for bandwidth matrix choice.The pointwise and global evaluations of the performance of the models applied were conducted by using two criteria. We calculated the pointwise absolute difference between observed and estimated values and the root mean squared error (RMSE) defined as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}},$$

where $\hat{y}_i$ is the adjustment of the i the observation y_i with $n = 15$.

Performance of estimators in terms of pointwise absolute difference and RMSE. About the absolute pointwise differences in Table 7, the large probability mass at zero was well estimated by kernel regression estimators while parametric model failed to estimate the excess amount of zeros. In particular, the semiparametric additive kernel estimator provided the best estimation at points $(1, 1)$ and $(1, 2)$ having highest observed frequencies.
Concerning RMSE-values, nonparametric (RMSE= 0.179842) and semiparametric estimators (RMSE^{mult}=0.1797684 and RMSE^{add}= 0.179767) outperformed the parametric logarithmic model (RMSE=0.194891) to estimate count real data for Stockholm city region. However, the three kernel regression estimators had similar performances. $\mathbf{H}_{cv}$ values are also presented in Table 7. Results of parametric and nonparametric models are illustrated in Figure 5.

		|Difference|		
Counts (x_1, x_2)	Observed	|Obs. - Nonp.|	|Obs. - Semip. mult.|	|Obs. - Semip. add.|	|Obs. - Param.|
$(1,1)$	0.255	1.770	0.311	**0.099**	241.200
$(1,2)$	0.299	29.046	27.235	**26.690**	289.604
$(1,3)$	0.177	**1.085**	1.151	1.694	109.980
$(1,4)$	0.066	0.943	0.942	**0.507**	60.525
$(1,5)$	0.043	1.514	1.514	**1.161**	38.615
$(1,6)$	0.030	0.565	0.565	**0.275**	26.417
$(1,7)$	0.032	0.539	0.539	**0.297**	29.021
$(1, 8-30)$	0.056	3.946	3.946	**3.740**	53.479
$(2,2)$	0.021	17.408	16.482	**15.611**	24.174
$(2,3)$	0.018	7.419	7.490	**6.924**	20.371
$(2,4)$	0.028	0.883	0.883	**0.430**	29.849
$(2,5)$	0.009	3.167	3.167	**2.803**	10.481
$(2,6)$	0.008	1.215	1.215	**0.917**	9.210
$(2,7)$	0.003	2.842	2.842	**2.594**	4.006
$(2, 8-30)$	0.015	1.820	1.820	**1.610**	15.851
		RMSE			
		Nonp. estimator	Semip. mult. estimator	Semip. add. estimator	Param. estimator
		179.842	179.768	**179.767**	194.891
$\mathbf{H}_{cv}$		$\begin{pmatrix} 0.09 & 0 \\ 0 & 0.10 \end{pmatrix}$	$\begin{pmatrix} 0.09 & 0 \\ 0 & 0.03 \end{pmatrix}$	$\begin{pmatrix} 0.104 & 0 \\ 0 & 0.105 \end{pmatrix}$	

Table 7: *Absolute differences between empirical and estimated frequencies calculated for Stockholm sample and RMSE obtained by using nonparametric and semiparametric regression estimators and the parametric logarithmic model. Results are multiplied by* 10^3 *and smallest error criterion values are in bold face.*

Model diagnostics. Mean and variance of estimated additive (respectively, multiplicative) correction factor $\widehat{\delta}_n$ (resp. $\widehat{\omega}_n$) were $\mathbb{E}(\widehat{\delta}_n) = 0.042246$ and $\mathrm{Var}(\widehat{\delta}_n) = 0.000955$ (resp. $\mathbb{E}(\widehat{\omega}_n) = 3.289427$ and $\mathrm{Var}(\widehat{\omega}_n) = 6.230134$) for Stockholm sample. To check model adequacy, the graphic of $Z(\mathbf{x})$-values shown that $Z(\mathbf{x})$ lied within the interval $[-1.96, 1.96]$ about 93% of the time (Figure 6). Despite of the point $(1, 1)$ with the largest probability mass which was out of interval $[-1.96, 1.96]$, the logarithmic model could be used as start function for semiparametric modeling of Stockholm city region data. More appropriate parametric models could be investigated to take into account special features of the counting phenomena such as zero-inflation (e.g. zero-inflated Poisson models).

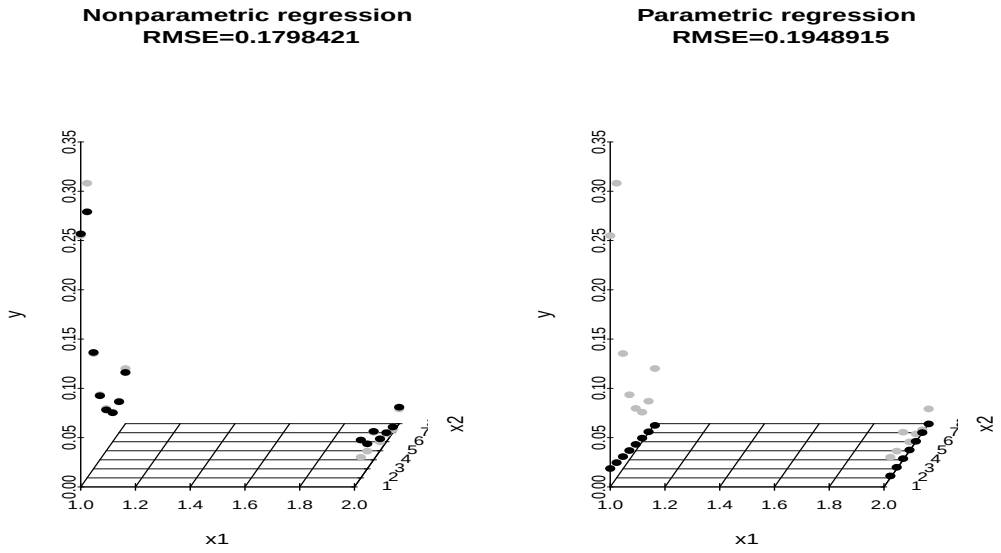


Figure 5: *Parametric logarithmic and nonparametric regressions of empirical frequencies for the Stockholm sample with n=15.*

6 Concluding remarks

This study aimed to present discrete semiparametric multivariate kernel estimators for count regression functions. The semiparametric estimation approach was investigated in two ways: additive and multiplicative combinations of discrete nonparametric kernel and parametric estimations. For each discrete semiparametric multivariate kernel estimator, model diagnostics were used to check the adequacy of parametric start. Simulations shown that the discrete semiparametric multivariate kernel studied outperform the nonparametric kernel estimator when the parametric model used as start function belongs to a neighbourhood of the true regression model. For instance, the application case shown that discrete semiparametric multivariate kernel estimators improved the parametric logarithmic model considered to estimate the excess amount of zeros. Future works would consist of conducting deeper studies to test several start functions and distinguish the performance of additive and multiplicative kernel estimators, something that was not done in this study. In addition, another procedure for bandwidth matrix choice would be also investigated such as Bayesian approach.

Acknowledgments

The authors wish to thank the two anonymous reviewers whose the expertise and comments significantly improved the quality of this paper.

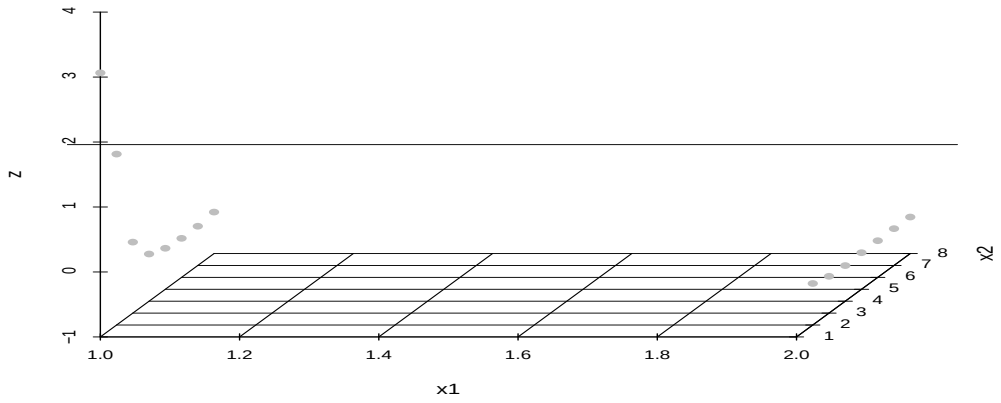


Figure 6: *The $Z(\mathbf{x})$-values associated with the results of semiparametric additive kernel estimator of count data of Stockholm city region with $n = 15$.*

Appendix

Proof of Theorem 1

Proof 1
We present the additive corrected estimator $\widehat{m}^{add}(\mathbf{x})$ as the following ratio

$$\widehat{m}^{add}(\mathbf{x}) = \frac{N_n^{add}(\mathbf{x})}{D_n^{add}(\mathbf{x})}, \ \ \mathbf{x} \in \mathbb{N}^d, \tag{14}$$

with $D_n^{add}(\mathbf{x}) = (1/n)\sum_{i=1}^n \prod_{j=1}^d K_{x_j,h_{jj}}^{[j]}(\mathbf{x}_i) = \widetilde{f}_n(\mathbf{x})$ and
$N_n^{add}(\mathbf{x}) = (1/n)\sum_{i=1}^n [r_0(\mathbf{x}) + \{y_i - r_0(\mathbf{x}_i)\}] \prod_{j=1}^d K_{x_j,h_{jj}}^{[j]}(\mathbf{x}_i)$. Let us consider the pmf f of rv $\mathbf{X}$. To establish the proof of Theorem 1 we assumed the continuity of the pmf f in the sense that

$$\forall\ \epsilon, \exists\ \eta > 0 : \forall\ \mathbf{z} \in (\mathbf{x} - \eta; \mathbf{x} + \eta) \cap \mathbb{T}^d \Longrightarrow \|f(\mathbf{z}) - f(\mathbf{x})\| < \epsilon.$$

In addition, the convergences of D_n^{add} to f and N_n^{add} to mf were required.

Consistency of D_n^{add}. We shown that the pointwise squared error of D_n^{add} was such that

$$\text{MSE}(\mathbf{x}) = \mathbb{E}[\{D_n^{add}(\mathbf{x}) - f(\mathbf{x})\}^2] = \text{Bias}^2\{D_n^{add}(\mathbf{x})\} + \text{Var}\{D_n^{add}(\mathbf{x})\} \to 0 \text{ as } n \to \infty.$$

The bias term was calculated using discrete Taylor expansion around target $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_d)$ such that we obtained

$$\begin{aligned}
\mathbb{E}\{D_n^{add}(\mathbf{x}_1, \ldots, \mathbf{x}_d)\} - f(\mathbf{x}_1, \ldots, \mathbf{x}_d) &= \sum_{\mathbf{z} \in \mathbb{S}_{\mathbf{x},\mathbf{H}}} f(\mathbf{z}) \Pr(\mathcal{K}_{\mathbf{x};\mathbf{H}} = \mathbf{z}) - f(\mathbf{x}_1, \ldots, \mathbf{x}_d) \\
&= \mathbb{E}\Big\{ f(\mathcal{K}^{[1]}_{\mathbf{x}_1,h_{11}}, \ldots, \mathcal{K}^{[d]}_{\mathbf{x}_d,h_{dd}}) \Big\} - f(\mathbf{x}_1, \ldots, \mathbf{x}_d) \\
&= \frac{1}{2} \sum_{j=1}^{d} \mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}) f_j^{(2)}(\mathbf{x}_1, \ldots, \mathbf{x}_d) + o\Big(\sum_{j=1}^{d} h_{jj}^2 \Big).
\end{aligned}$$

Hence, the D_n^{add}'s bias was consistent since $\mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}) \to 0$ as $n \to \infty$ and $h_{jj} \to 0$. The The variance term was decomposed as the sum of a main term depending on target $\mathbf{x} \in \mathbb{S}^d_{\mathbf{x},\mathbf{H}}$ and a remaining term depending on points $\mathbf{z} \in \mathbb{S}^d_{\mathbf{x},\mathbf{H}} \setminus \{\mathbf{x}\}$. Then, we expressed

$$\begin{aligned}
\mathrm{Var}\{D_n^{add}(\mathbf{x})\} &= \mathrm{Var}\Big\{ \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(x_{ij}) \Big\} \\
&= \frac{1}{n} \Big[f(\mathbf{x}) \Big\{ \prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_j) \Big\}^2 - f^2(\mathbf{x}) \Big\{ \prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_j) \Big\}^2 \Big] + R_n \\
&= \frac{1}{n} f(\mathbf{x}) \{1 - f(\mathbf{x})\} \Big\{ \prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_j) \Big\}^2 + R_n,
\end{aligned}$$

where the term

$$R_n = \frac{1}{n} \Big[\sum_{\mathbf{y} \in \mathbb{S}^d_{\mathbf{x},\mathbf{H}} \setminus \{\mathbf{x}\}} f(\mathbf{y}) \Big\{ \prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{y}_j) \Big\}^2 - \Big\{ \sum_{\mathbf{y} \in \mathbb{S}^d_{\mathbf{x},\mathbf{H}} \setminus \{\mathbf{x}\}} f(\mathbf{y}) \prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{y}_j) \Big\}^2 \Big] \tag{15}$$

went to 0 as $\mathbf{H} \to 0_d$. Indeed, let $\mathbf{y} \in \mathbb{S}^d_{\mathbf{x},\mathbf{H}} \setminus \{\mathbf{x}\}$ we can find a constant $\eta > 0$ such that

$$\begin{aligned}
K^{[1]}_{x_1,h_{11}}(y_1) \ldots K^{[d]}_{x_d,h_{dd}}(y_d) &\leq \Pr(|\mathcal{K}^{[1]}_{x_1,h_{11}} - x_1| > \eta, \ldots, |\mathcal{K}^{[d]}_{x_d,h_{dd}} - x_d| > \eta) \\
&\leq \frac{\mathbb{E}(\mathcal{K}^{[d]}_{x_1,h_{11}} - x_1)^2}{\eta^2} \cdots \frac{\mathbb{E}(\mathcal{K}^{[d]}_{x_d,h_{dd}} - x_d)^2}{\eta^2} \\
&= \frac{1}{\eta^{2d}} \prod_{j=1}^{d} \Big[\mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}) - \{\mathbb{E}(\mathcal{K}^{j}_{\mathbf{x},h_{jj}}) - \mathbf{x}_j\}^2 \Big] \to 0
\end{aligned}$$

under the two following hypotheses:

$$\lim_{h_{jj} \to 0} \mathbb{E}(\mathcal{K}_{x_1,h_{11}}, \ldots, \mathcal{K}_{x_d,h_{dd}}) = \mathbf{x} \text{ and } \lim_{h_{jj} \to 0} \mathrm{Var}(\mathcal{K}_{x_1,h_{11}}, \ldots, \mathcal{K}_{x_d,h_{dd}}) = 0_d.$$

For $\mathbf{y} = \mathbf{x}$ we deduced the asymptotic modal probability $\prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(x_1) \to 1$ when $h_{jj} \to 0$. Thus, the variance term went to 0 as $n \to \infty$ and $h_{jj} \to 0$.
We omitted to present the consistency of N_n^{add} since it was obtained in a similar way. ■

Now we are able to present the expectation and variance of $\widehat{m}^{add}$.

Expectation of $\widehat{m}^{add}$. To express the expectation of $\widehat{m}^{add}(\mathbf{x})$, we calculate the expectation of N_n^{add} in a way similar to that of D_n^{add} using discrete Taylor expansion such that

$$\begin{aligned}
\mathbb{E}\{N_n^{add}(\mathbf{x})\} &= \sum_{\mathbf{z}\in\mathbb{S}_{\mathbf{x}}} \Big[r_0(\mathbf{x}) + \{m(\mathbf{z}) - r_0(\mathbf{z})\}\Big] f(\mathbf{z}) \Pr(\mathcal{K}_{\mathbf{x};\mathbf{H}} = \mathbf{z}) \\
&= r_0(\mathbf{x}_1,\ldots,\mathbf{x}_d)\mathbb{E}\Big\{f(\mathcal{K}^{[1]}_{\mathbf{x}_1,h_{11}},\ldots,\mathcal{K}^{[d]}_{\mathbf{x}_d,h_{dd}})\Big\} + \mathbb{E}\Big\{(\delta f)(\mathcal{K}^{[1]}_{\mathbf{x}_1,h_{11}},\ldots,\mathcal{K}^{[d]}_{\mathbf{x}_d,h_{dd}})\Big\} \\
&= r_0(\mathbf{x}_1,\ldots,\mathbf{x}_d)\Big\{f(\mathbf{x}_1,\ldots,\mathbf{x}_d) + \frac{1}{2}\sum_{j=1}^{d} \mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}) f_j^{(2)}(\mathbf{x}_1,\ldots,\mathbf{x}_d)\Big\} \\
&\quad + (\delta f)(\mathbf{x}_1,\ldots,\mathbf{x}_d)\frac{1}{2}\sum_{j=1}^{d} \mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}})(\delta f)_j^{(2)}(\mathbf{x}_1,\ldots,\mathbf{x}_d) + o\Big(\sum_{j=1}^{d} h_{jj}^2\Big).
\end{aligned}$$

Then, $\widehat{m}_n^{add}$'s expectation is obtained by using the following approximation from **?**, pp.119-121:

$$\widehat{m}(\mathbf{x}) \simeq m(\mathbf{x}) + \frac{N_n^{add}(\mathbf{x}) - (mf)(\mathbf{x})}{f(\mathbf{x})} - \frac{(mf)(\mathbf{x})\{D_n^{add}(\mathbf{x}) - f(\mathbf{x})\}}{f^2(\mathbf{x})}, \tag{16}$$

such that

$$\begin{aligned}
\mathbb{E}\{\widehat{m}_n^{add}(\mathbf{x})\} &\simeq m(\mathbf{x}) + \frac{1}{f(\mathbf{x})}\mathbb{E}\{N_n^{add}(\mathbf{x}) - (mf)(\mathbf{x})\} - \frac{(mf)(\mathbf{x})}{f^2(\mathbf{x})}\mathbb{E}\{D_n^{add}(\mathbf{x}) - f(\mathbf{x})\} \\
&\simeq m(\mathbf{x}) + \frac{1}{2}\sum_{j=1}^{d} \frac{r_0(\mathbf{x}) f_j^{(2)}(\mathbf{x}) + (\delta f)_j^{(2)}(\mathbf{x}) - (r_0 + \delta_j)(\mathbf{x}) f_j^{(2)}(\mathbf{x})}{f_j(\mathbf{x})} \mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}) \\
&\simeq m(\mathbf{x}) + \frac{1}{2}\sum_{j=1}^{d} \left\{\delta_j^{(2)}(\mathbf{x}) + 2\delta_j^{(1)}\frac{f^{(1)}(\mathbf{x})}{f(\mathbf{x})}\right\} \mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}),
\end{aligned}$$

where f_j is the univariate pmf of rv X_j and $\delta_j(\mathbf{x})$ is a nonparametric correction function. The functions $f_j^{(1)}$, $\delta_j^{(1)}$ and $\delta_{jj}^{(2)}$ are j-partial finite differences for $j = 1, 2, \ldots, d$, defined in 6 and 7 respectively. The finite difference of second order $(\delta f)^{(2)}$ is equal to $\delta^{(2)} f + 2\delta^{(1)} f^{(1)} + \delta f^{(2)}$. Finally, we get the $\widehat{m}_n^{add}$'s bias expressed by

$$\mathrm{Bias}\{\widehat{m}_n^{add}(\mathbf{x})\} \approx \frac{1}{2}\sum_{j=1}^{d} \left[\delta_{jj}^{(2)} + 2\delta_j^{(1)}\left\{\frac{f_j^{(1)}}{f}(\mathbf{x})\right\}\right] \mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}}).$$

Variance of $\widehat{m}^{add}$. The expression of $\widehat{m}^{add}$'s variance required to calculate N_n^{add}'s variance in a way similar to that of D_n^{add} such that

$$\begin{aligned}
\text{Var}\{N_n^{add}(\mathbf{x})\} &= \frac{1}{n}\text{Var}\Big[\Big\{r_0(\mathbf{x}) + \delta(X_{1j}) + \epsilon_1\Big\}\prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(X_{1j})\Big] \\
&= \frac{1}{n}\sum_{\mathbf{z}\in\mathbb{S}^d_{\mathbf{x},\mathbf{H}}}\Big\{r_0(\mathbf{x}) + \delta(\mathbf{z}) + \epsilon_1\Big\}^2 f(\mathbf{z})\Big\{\prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{z})\Big\}^2 \\
&\quad -\frac{1}{n}\Big[\sum_{\mathbf{z}\in\mathbb{S}^d_{\mathbf{x},\mathbf{H}}}\Big\{r_0(\mathbf{x}) + \delta(\mathbf{z}) + \epsilon_1\Big\} f(\mathbf{z})\prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{z})\Big]^2 \\
&= \frac{1}{n}\Big\{\mathbb{E}(Y_1^2|\mathbf{X_1} = \mathbf{x}) - \mathbb{E}^2(Y_1|\mathbf{X_1} = \mathbf{x})f(\mathbf{x})\Big\}f(\mathbf{x})\Big\{\prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_j)\Big\}^2 + Q_n
\end{aligned}$$

where Q_n contained the remaining terms depending on $\mathbf{z} \in \mathbb{S}^d_{\mathbf{x},\mathbf{H}} \setminus \{\mathbf{x}\}$. The same arguments similar to those used for R_n in (15) might be used to shown that Q_n went to 0 as $\mathbf{H} \to 0_d$. From (16) the $\widehat{m}_n^{add}$'s variance resulted in

$$\begin{aligned}
\text{Var}\{\widehat{m}_n^{add}(\mathbf{x})\} &\approx \frac{\text{Var}\{N_n^{add}(\mathbf{x})\}}{f^2(\mathbf{x})} + \frac{(mf)^2(\mathbf{x})}{f^4(\mathbf{x})}\text{Var}(D_n^{add}(\mathbf{x})) - 2\frac{(mf)(\mathbf{x})}{f^3(\mathbf{x})}Cov\{D_n^{add}(\mathbf{x}), N_n^{add}(\mathbf{x})\} \\
&\approx \frac{\text{Var}\{N_n^{add}(\mathbf{x})\}}{f^2(\mathbf{x})} + O\Big(\frac{1}{n}\Big) - O\Big(\frac{1}{n}\Big) \\
&\approx \frac{\{\mathbb{E}(Y_1^2|\mathbf{x}_1 = \mathbf{x}) - f(\mathbf{x})\mathbb{E}^2(Y_1|\mathbf{x}_1 = \mathbf{x})\}}{nf(\mathbf{x})}\Big\{\prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_j)\Big\}^2
\end{aligned}$$

with

$$Cov(D_n^{add}(\mathbf{x}), N_n^{add}(\mathbf{x})) = \frac{1}{n}\Big[\mathbf{E}(Y_1\{\prod_{j=1}^{d} K^{[j]}_{x_j,h_j}(x_{1j})\}^2) - \mathbf{E}(Y_1\prod_{j=1}^{d} K^{[j]}_{x_j,h_j}(x_{1j}))\mathbf{E}(\prod_{j=1}^{d} K^{[j]}_{x_j,h_j}(x_{2j}))\Big] = O\Big(\frac{1}{n}\Big),$$

and

$$\text{Var}\{D_n^{add}(\mathbf{x},\mathbf{H})\} = O\Big(\frac{1}{n}\Big).$$

■

Bias (10) and variance (11) of multiplicative setimator

Similar to the additive case, assume that the multiplicative estimator in equation (9) can be given by

$$\widehat{m}^{mult}(\mathbf{x}) = \frac{N_n^{mult}(\mathbf{x})}{D_n^{mult}(\mathbf{x})}, \quad \mathbf{x} \in \mathbb{N}^d,$$

The main difference with the proof of additive case comes from the expressions of the N_n^{add}'s expectation and variance since $D_n^{mult} = D_n^{add}$. Thus, let us consider

$$N_n^{mult}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\Big\{r_0(\mathbf{x}) \times \frac{y_i}{r_0(\mathbf{x}_i)}\Big\}\prod_{j=1}^{d} K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i).$$

We expressed the expectation of N_n^{mult} by

$$\begin{aligned}
\mathbb{E}\{N_n^{mult}(\mathbf{x})\} &= \sum_{\mathbf{z}\in\mathbb{S}_{\mathbf{x}}}\Big\{r_0(\mathbf{x}) \times \frac{m(\mathbf{z})}{r_0(\mathbf{z})}\Big\}f(\mathbf{z})\Pr(\mathcal{K}_{\mathbf{x};\mathbf{H}} = \mathbf{z}) \\
&= r_0(\mathbf{x})\mathbb{E}\Big\{(\omega f)(\mathcal{K}^{[1]}_{\mathbf{x}_1,h_{11}},\ldots,\mathcal{K}^{[d]}_{\mathbf{x}_d,h_{dd}})\Big\} \\
&= r_0(\mathbf{x})\Big\{(\omega f)(\mathbf{x}) + \frac{1}{2}\sum_{j=1}^{d}\mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}})(\omega f)^{(2)}_j(\mathbf{x}) + o\Big(\sum_{j=1}^{d}h_{jj}^2\Big)\Big\} \\
&= (mf)(\mathbf{x}) + \frac{1}{2}\sum_{j=1}^{d}\mathrm{Var}(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}})r_0(\mathbf{x})(\omega f)^{(2)}_j(\mathbf{x}) + o\Big(\sum_{j=1}^{d}h_{jj}^2\Big).
\end{aligned}$$

In addition, for the variance we obtained

$$\begin{aligned}
\mathrm{Var}\{N_n^{mult}(\mathbf{x})\} &= \frac{1}{n}\mathrm{Var}\Big[\Big\{r_0(\mathbf{x})\omega(X_{1j}) + \epsilon_1\Big\}\prod_{j=1}^{d}K^{[j]}_{\mathbf{x}_j,h_{jj}}(X_{1j})\Big] \\
&= \frac{1}{n}\sum_{\mathbf{z}\in\mathbb{S}^d_{\mathbf{x},\mathbf{H}}}\Big\{r_0(\mathbf{x})\omega(\mathbf{z}) + \epsilon_1\Big\}^2 f(\mathbf{z})\Big\{\prod_{j=1}^{d}K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{z})\Big\}^2 \\
&\quad -\frac{1}{n}\Big[\sum_{\mathbf{z}\in\mathbb{S}^d_{\mathbf{x},\mathbf{H}}}\Big\{r_0(\mathbf{x})\omega(\mathbf{z}) + \epsilon_1\Big\}f(\mathbf{z})\prod_{j=1}^{d}K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{z})\Big]^2 \\
&\approx \frac{1}{n}\Big\{\mathbb{E}(Y_1^2|\mathbf{X_1} = \mathbf{x}) - \mathbb{E}^2(Y_1|\mathbf{X_1} = \mathbf{x})f(\mathbf{x})\Big\}f(\mathbf{x})\Big\{\prod_{j=1}^{d}K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_j)\Big\}^2,
\end{aligned}$$

where we omitted to present the remaining term that converges to 0 as $n \to \infty$ and $h_{jj} \to 0$. ■

In what follows, to prove the asymptotic normality for semiparametric multivariate additive and semiparametric estimators, we need to recall the Lyapounov central limit theorem for triangular arrays (**?**).

Theorem (Lyapounov) Assume that $\{X_{n,j}, j = 1,\ldots,k_n\}$ are zero-mean independent random variables, $n = 1, 2, \cdots$. If

$$\lim_{n\to\infty}\sum_{j=1}^{k_n}\mathbb{E}(X_{n,j}^2) = \Sigma^2 > 0,$$

$$\lim_{n\to\infty}\sum_{j=1}^{k_n}\mathbb{E}(|X_{n,j}|^3)=0,$$

then $S_n = X_{n,1}+\cdots+X_{n,k_n}$ converges in distribution to the normal law with the mean zero and the variance Σ^2

$$S_n \xrightarrow{d} \mathcal{N}(0,\Sigma^2)\ as\ n\to\infty.$$

The notation "$\xrightarrow{d}$" stands for convergence in distribution.

Proof of Theorem 2

We consider the expression of $\widehat{m}_n^{add}$ in (14) with $D_n^{add}(\mathbf{x}) = (1/n)\sum_{i=1}^n\prod_{j=1}^d K_{x_j,h_{jj}}^{[j]}(\mathbf{x}_i) = \widetilde{f_n}(\mathbf{x})$.

Proof 2

$$\begin{aligned}
\{\widehat{\delta_n}(\mathbf{x})-\delta(\mathbf{x})\}\times D_n^{add}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n\prod_{j=1}^d K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_i)\{(\mathbf{y}_i)-\widehat{r}(\mathbf{x}_i)\}-\delta(\mathbf{x})\times D_n^{add}(\mathbf{x})\\
&= \frac{1}{n}\sum_{i=1}^n\prod_{j=1}^d K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_i)\{\delta(\mathbf{x}_i)+\epsilon_i\}-\delta(\mathbf{x})\times D_n^{add}(\mathbf{x})\\
&= \frac{1}{n}\sum_{i=1}^n K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\{\delta(\mathbf{x}_i)-\delta(\mathbf{x})\}+\frac{1}{n}\sum_{i=1}^n K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i \qquad (17)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\Big[\Big\{\widehat{\delta_n}(\mathbf{x})-\delta(\mathbf{x})\Big\}\times D_n^{add}(\mathbf{x})\Big] &= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\{\delta(\mathbf{x}_i)-\delta(\mathbf{x})\}\right]+\frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^n K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i\right\} \qquad (18)\\
&= \sum_{\mathbf{y}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}^d}(\delta f)(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_d)\prod_{j=1}^d\Pr(\mathcal{K}_{\mathbf{x}_j,h_{jj}}^{[j]}=\mathbf{y}_j)\\
&\quad -\delta(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_d)\sum_{\mathbf{y}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}^d}f(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_d)\prod_{j=1}^d\Pr(\mathcal{K}_{\mathbf{x}_j,h_{jj}}^{[j]}=\mathbf{y}_j)\\
&= (\delta f)\{\mathbb{E}(\mathcal{K}_{\mathbf{x}_1,h_{11}}^{[j]},\mathcal{K}_{\mathbf{x}_2,h_{22}}^{[2]},\ldots,\mathcal{K}_{\mathbf{x}_d,h_{dd}}^{[d]})\}\\
&\quad -\delta(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_d)f\{\mathbb{E}(\mathcal{K}_{\mathbf{x}_1,h_{11}}^{[j]},\mathcal{K}_{\mathbf{x}_2,h_{22}}^{[2]},\ldots,\mathcal{K}_{\mathbf{x}_d,h_{dd}}^{[d]})\}+o\Big(\sum_{j=1}^d h_{jj}^2\Big)\\
&= \frac{1}{2}\sum_{j=1}^d\mathrm{Var}(\mathcal{K}_{x_j,h_{jj}})\{\delta_{jj}^{(2)}(\mathbf{x})f(\mathbf{x})+2\delta_j^{(1)}(\mathbf{x})f_j^{(1)}(\mathbf{x})\}+o\Big(\sum_{j=1}^d h_{jj}^2\Big).
\end{aligned}$$

Then, for the variance of (17) we have

$$\begin{aligned}
\mathrm{Var}\Big[\big\{\widehat{\delta_n}(\mathbf{x}) - \delta(\mathbf{x})\big\} \times D_n^{add}(\mathbf{x})\Big] &= \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\{\delta(\mathbf{x}_i) - \delta(\mathbf{x})\}\right] + \frac{1}{n^2}\mathrm{Var}\left\{\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i\right\} \\
&\quad -\frac{2}{n^2}\mathrm{Cov}\left(\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\{\delta(\mathbf{x}_i) - \delta(\mathbf{x})\}, \sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i\right) \\
&= \frac{f(\mathbf{x})\sigma^2}{n}\{\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}} = \mathbf{x})\}^2 + o\left(\frac{\sum_{j=1}^{d} h_{jj}^2}{n} + \frac{1}{n}\right) + O\left(\frac{1}{n}\right) \\
&= \frac{f(\mathbf{x})\sigma^2}{n}\Big\{1 - \Pr(\mathcal{K}_{\mathbf{x}} = \mathbf{x})\prod_{j=1}^{d} h_{jj}\Big\}^2 + o\left(\frac{\sum_{j=1}^{d} h_{jj}^2}{n} + \frac{1}{n}\right) + O\left(\frac{1}{n}\right),
\end{aligned}$$

with $\sigma^2 = \mathrm{Var}(\epsilon_i) < \infty$. This result is essentially due to the second term in (18) given by $(1/n)\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i$, which is a sum of i.i.d. random variables. Indeed, under assumptions (A1)-(A3), we have

$$\begin{aligned}
\frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\{\delta(\mathbf{x}_i) - \delta(\mathbf{x})\}\right] &= \frac{1}{n}\Big[\sum_{\mathbf{z}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}^d\setminus\{\mathbf{x}\}}\{\delta(\mathbf{z}) - \delta(\mathbf{x})\}^2\Big\{\prod_{j=1}^{d}\Pr(\mathcal{K}_{\mathbf{x}_j,h_{jj}}^{[j]} = \mathbf{z}_j)\Big\}^2 f(\mathbf{z}) \\
&\quad -\Big\{\sum_{\mathbf{z}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}^d\setminus\{\mathbf{x}\}}\{\delta(\mathbf{z}) - \delta(\mathbf{x})\}\prod_{j=1}^{d}\Pr(\mathcal{K}_{\mathbf{x}_j,h_{jj}}^{[j]} = \mathbf{z}_j)f(\mathbf{z})\Big\}^2\Big] = o\left(\frac{\sum_{j=1}^{d} h_{jj}^2}{n}\right)
\end{aligned}$$

and

$$\frac{1}{n^2}\mathrm{Cov}\left(\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\{\delta(\mathbf{x}_i) - \delta(\mathbf{x})\}, \sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i\right) = O\left(\frac{1}{n}\right).$$

Then, we mainly have

$$\begin{aligned}
\frac{1}{n^2}\mathrm{Var}\left\{\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i\right\} &= \frac{1}{n}\left[\mathbb{E}\left\{K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i\right\}^2 - \mathbb{E}^2\left\{K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i\right\}\right] \\
&= \frac{1}{n}\sum_{\mathbf{z}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}^d}\epsilon_1^2 f(\mathbf{z})\Big\{\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{z})\Big\}^2 - \frac{1}{n}\Big[\sum_{\mathbf{z}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}^d}\epsilon_1 f(\mathbf{z})\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{z})\Big]^2 \\
&= \frac{\mathbb{E}(\epsilon_1^2)f(\mathbf{x})}{n}\Big\{\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_j)\Big\}^2 + S_n
\end{aligned}$$

where

$$S_n = \frac{1}{n}\sum_{\mathbf{z}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}^d}\epsilon_1^2 f(\mathbf{z})\Big\{\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{z})\Big\}^2 - \frac{1}{n}\mathbb{E}(\epsilon_1^2)f(\mathbf{x})\Big\{\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_j)\Big\}^2 - \frac{1}{n}\Big\{\sum_{\mathbf{z}\in\mathbb{S}_{\mathbf{x},\mathbf{H}}^d}\epsilon_1 f(\mathbf{z})\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{z})\Big\}^2$$

tends to 0 as $n \to \infty$ and $\mathbf{H} \to \mathbf{0_d}$, similar to the term R in equation (15). Rather, by applying the Lyapounov's central limit theorem on $\frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\epsilon_i$, we have

$$\sqrt{n}\Big[\Big\{\widehat{\delta_n}(\mathbf{x}) - \delta(\mathbf{x})\Big\} \times D_n^{add}(\mathbf{x})\Big] \xrightarrow{d} \mathcal{N}(0, f(\mathbf{x})\sigma^2\{\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}} = \mathbf{x})\}^2).$$

Finally, by considering the convergence of $\widetilde{f_n}$ to f states by **?**, it results that

$$\sqrt{n}\{\widehat{\delta_n}(\mathbf{x}) - \delta(\mathbf{x})\}D_n^{add}(\mathbf{x}) = \sqrt{n}\{\widehat{\delta_n}(\mathbf{x}) - \delta(\mathbf{x})\}f(\mathbf{x}) + o_p(1)$$

such that

$$\mathbb{E}\{\widehat{\delta_n}(\mathbf{x}) - \delta(\mathbf{x})\} = \frac{1}{2}\sum_{j=1}^{d} \mathrm{Var}(\mathcal{K}_{x_j,h_{jj}})\Big\{\delta_{jj}^{(2)}(\mathbf{x}) + 2\delta_j^{(1)}(\mathbf{x})\Big(\frac{f_j^{(1)}}{f}\Big)(\mathbf{x})\Big\} + o\Big(\sum_{j=1}^{d} h_{jj}^2\Big) \to 0, \text{ as } h_{jj} \to 0,$$

and

$$\mathrm{Var}\{\widehat{\delta_n}(\mathbf{x}) - \delta(\mathbf{x})\} = \frac{1}{nf(\mathbf{x})}\sigma^2\{\Pr(\mathcal{K}_{\mathbf{x}} = \mathbf{x})\}^2 \prod_{j=1}^{d} h_{jj}^2 + o\Big(\frac{1}{n} + \frac{\sum_{j=1}^{d} h_{jj}^2}{n}\Big).$$

■

Proof of Theorem 3

Proof 3

For $\mathbf{x} \in \mathbb{N}^d$ and $\mathbf{H} = \mathbf{Diag}(h_{11}, \ldots, h_{dd})$ with $h_{jj} > 0$, let us consider the semiparametric estimator $\widehat{\omega}_n$ in (9) and the sequence $D_n^{mult}(\mathbf{x}) = (1/n)\sum_{j=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_j)$. By using a similar approximation as in equation (16), we have

$$\begin{aligned}\frac{1}{\widehat{r}(\mathbf{x}_i)} = \frac{\widehat{r}(\mathbf{x})}{\widehat{r}(\mathbf{x})\widehat{r}(\mathbf{x}_i)} &= \frac{r_0(\mathbf{x})}{r_0(\mathbf{x})r_0(\mathbf{x}_i)} + \frac{\widehat{r}(\mathbf{x}) - r_0(\mathbf{x})}{r_0(\mathbf{x})r_0(\mathbf{x}_i)} - \frac{r_0(\mathbf{x})\{\widehat{r}(\mathbf{x}) - r_0(\mathbf{x})r_0(\mathbf{x}_i)\}}{r_0^2(\mathbf{x})r_0^2(\mathbf{x}_i)} \\ &= \frac{1}{r_0(\mathbf{x}_i)} + \frac{\widehat{r}(\mathbf{x}) - r_0(\mathbf{x})}{r_0(\mathbf{x})r_0(\mathbf{x}_i)} - \frac{\widehat{r}(\mathbf{x}) - r_0(\mathbf{x})r_0(\mathbf{x}_i)}{r_0(\mathbf{x})r_0^2(\mathbf{x}_i)}.\end{aligned}$$

Then, we have

$$\begin{aligned}\widehat{\omega}_n(\mathbf{x}) \times D_n^{mult}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_i)\mathbf{y}_i\frac{1}{r_0(\mathbf{x}_i)} + \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_i)\frac{\mathbf{y}_i}{r_0(\mathbf{x})r_0(\mathbf{x}_i)}\{\widehat{r}(\mathbf{x}_j) - r_0(\mathbf{x}_j)\} \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_i)\mathbf{y}_i\frac{1}{r_0^2(\mathbf{x}_i)r_0(\mathbf{x}_i)}\{\widehat{r}(\mathbf{x}_i) - r_0(\mathbf{x}_j)r_0(\mathbf{x}_i)\}\end{aligned}$$

By using the equation $\mathbf{y}_i = r_0(\mathbf{x}_i)\omega(\mathbf{x}_i) + \epsilon_\mathbf{i}$ where $\mathbf{x}_i$ and ϵ_i are independent variables, the terms

$$\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_i)\frac{\epsilon_i}{r_0(\mathbf{x}_j)r_0(\mathbf{x}_i)}\{\widehat{r}(\mathbf{x}_j) - r_0(\mathbf{x}_j)\} \text{ and } \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} K_{\mathbf{x}_j,h_{jj}}^{[j]}(\mathbf{x}_i)\epsilon_i\frac{1}{r_0^2(\mathbf{x}_i)r_0(\mathbf{x}_j)}\{\widehat{r}(\mathbf{x}_i) - r_0(\mathbf{x}_j)r_0(\mathbf{x}_i)\}$$

are of order $o_p\Big(\sum_{j=1}^d h_{jj}^2\Big)$. It ensues the following equalities:

$$
\begin{aligned}
\{\widehat{\omega}_n(\mathbf{x}) - \omega(\mathbf{x})\} \times D_n^{mult}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i)\{r_0(\mathbf{x}_i)\omega(\mathbf{x}_i) + \epsilon_i\}\frac{1}{r_0(\mathbf{x}_i)} - \omega(\mathbf{x}) \times D_n^{mult}(\mathbf{x}) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i)\frac{\{r_0(\mathbf{x}_i)\omega(\mathbf{x}_i) + \epsilon_i\}}{r_0(\mathbf{x}_j)r_0(\mathbf{x}_i)}\{\widehat{r}(\mathbf{x}_j) - r_0(\mathbf{x}_j)\} \\
&\quad - \frac{1}{n}\sum_{i=1}^n \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i)\{r_0(\mathbf{x}_i)\omega(\mathbf{x}_i) + \epsilon_i\}\frac{1}{r_0^2(\mathbf{x}_i)r_0(\mathbf{x}_j)}\{\widehat{r}(\mathbf{x}_i) - r_0(\mathbf{x}_j)r_0(\mathbf{x}_i)\} \\
&= \frac{1}{n}\sum_{i=1}^n \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i)\Big\{\omega(\mathbf{x}_i) + \frac{\epsilon_i}{r_0(\mathbf{x}_i)}\Big\} - \omega(\mathbf{x}) \times D_n^{mult}(\mathbf{x}) \\
&\quad + \frac{1}{nr_0(\mathbf{x})}\sum_{i=1}^n \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i)\omega(\mathbf{x}_i)\Big\{\widehat{r}(\mathbf{x}_j) - r_0(\mathbf{x}_j)\Big\} \\
&\quad - \frac{1}{n}\sum_{i=1}^n \prod_{j=1}^d K^{[j]}_{\mathbf{x}_j,h_{jj}}(\mathbf{x}_i)\frac{\omega(\mathbf{x}_i)}{r_0(\mathbf{x}_i)r_0(\mathbf{x}_j)}\{\widehat{r}(\mathbf{x}_i) - r_0(\mathbf{x}_j)r_0(\mathbf{x}_i)\} + o_p\Big(\sum_{j=1}^d h_{jj}^2\Big) \\
&= A_n(\mathbf{x},\mathbf{H}) + B_n(\mathbf{x},\mathbf{H}) - C_n(\mathbf{x},\mathbf{H}) + o_p\Big(\sum_{j=1}^d h_{jj}^2\Big). \qquad (19)
\end{aligned}
$$

For calculating the expectation of (19), we begin by the first term A_n. Under assumptions (A1)-(A3) and using the discrete Taylor expansion, we have successively

$$
\begin{aligned}
\mathbb{E}\{A_n(\mathbf{x},\mathbf{H})\} &= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\{\omega(\mathbf{x}_i) - \omega(\mathbf{x})\}\right] + \frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^n K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i)\frac{\epsilon_i}{r_0(\mathbf{x}_i)}\right\} \qquad (20) \\
&= \sum_{\mathbf{y}\in\mathbb{S}^d_{\mathbf{x},\mathbf{H}}} (\omega f)(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_d)\prod_{j=1}^d \Pr(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}} = \mathbf{y}_j) \\
&\quad - \omega(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_d)\sum_{\mathbf{y}\in\mathbb{S}^d_{\mathbf{x},\mathbf{H}}} f(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_d)\prod_{j=1}^d \Pr(\mathcal{K}^{[j]}_{\mathbf{x}_j,h_{jj}} = \mathbf{y}_j) \\
&= (\omega f)\{\mathbb{E}(\mathcal{K}^{[j]}_{\mathbf{x}_1,h_{11}}, \mathcal{K}^{[2]}_{\mathbf{x}_2,h_{22}},\ldots,\mathcal{K}^{[d]}_{\mathbf{x}_d,h_{dd}})\} \\
&\quad - \omega(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_d) f\{\mathbb{E}(\mathcal{K}^{[j]}_{\mathbf{x}_1,h_{11}}, \mathcal{K}^{[2]}_{\mathbf{x}_2,h_{22}},\ldots,\mathcal{K}^{[d]}_{\mathbf{x}_d,h_{dd}})\} + o\Big(\sum_{j=1}^d h_{jj}^2\Big) \\
&= \frac{1}{2}\sum_{j=1}^d \mathrm{Var}(\mathcal{K}_{x_j,h_{jj}})\{\omega^{(2)}_{jj}(\mathbf{x})f(\mathbf{x}) + 2\omega^{(1)}_j(\mathbf{x})f^{(1)}_j(\mathbf{x})\} + o\Big(\sum_{j=1}^d h_{jj}^2\Big).
\end{aligned}
$$

The expectations of the second and third terms B_n and C_n in (19) are given by

$$\begin{aligned}\mathbb{E}\{B_n(\mathbf{x},\mathbf{H})\} &= \{1/r_0(\mathbf{x})\}\mathbb{E}\left\{K_{\mathbf{x};\mathbf{H}}(\mathbf{x}_1)\omega(\mathbf{x}_1)\right\}\mathbb{E}_{\mathbf{x}_i}\{\widehat{r}(\mathbf{x})-r_0(\mathbf{x})\}\\ &= \{1/r_0(\mathbf{x})\}\sum_{y\in\mathbb{S}^d_{\mathbf{x},\mathbf{H}}}\omega(\mathbf{y})f(\mathbf{y})\Pr(\mathcal{K}_{\mathbf{x};\mathbf{H}}=\mathbf{y})\mathbb{E}_{X_i}\{\widehat{r}(\mathbf{x})-r_0(\mathbf{x})\}\\ &= \{1/r_0(\mathbf{x})\}\omega(\mathbf{x})f(\mathbf{x})\mathbb{E}_{X_i}\{\widehat{r}(\mathbf{x})-r_0(\mathbf{x})\}+o\Big(\sum_{j=1}^d h_{jj}^2\Big),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\{C_n(\mathbf{x},\mathbf{H})\} &= \mathbb{E}\left\{K_{\mathbf{x},\mathbf{H}}(X_1)\frac{\omega(\mathbf{x}_1)}{r_0(\mathbf{x}_1)r_0(\mathbf{x})}\right\}\mathbb{E}_{\mathbf{x}_i}\{\widehat{r}(\mathbf{x}_i)-r_0(\mathbf{x})r_0(\mathbf{x}_i)\}\\ &= \{1/r_0^2(\mathbf{x})\}\omega(\mathbf{x})f(\mathbf{x})\mathbb{E}_{\mathbf{x}_i}\{\widehat{r}(\mathbf{x}_i)-r_0(\mathbf{x})r_0(\mathbf{x}_i)\}+o\Big(\sum_{j=1}^d h_{jj}^2\Big).\end{aligned}$$

It results $\mathbb{E}[\{\widehat{\omega}_n(\mathbf{x})-\omega(\mathbf{x})\}\times D_n^{mult}(\mathbf{x})]=\mathbb{E}\{A_n(\mathbf{x},\mathbf{H})\}+o\Big(\sum_{j=1}^d h_{jj}^2\Big)$. Then, for the variance of (19) we have

$$\mathrm{Var}\{A_n(\mathbf{x},\mathbf{H})+B_n(\mathbf{x},\mathbf{H})-C_n(\mathbf{x},\mathbf{H})\}=\frac{f(\mathbf{x})\sigma^2}{nr_0^2(\mathbf{x})}\{\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}}=\mathbf{x})\}^2+o\left(\frac{\sum_{j=1}^d h_{jj}^2}{n}\right)$$

with $\sigma^2=\mathrm{Var}(\epsilon_i)<\infty$. This result is essentially due to the second term in (20) given by

$$A_{1n}(\mathbf{x},\mathbf{H})=n^{-1}\sum_{i=1}^n \epsilon_i r_0^{-1}(\mathbf{x}_i)K_{\mathbf{x},\mathbf{H}}(\mathbf{x}_i),$$

which is a sum of i.i.d. random variables; thus we have $\mathbb{E}\{A_{1n}(\mathbf{x},\mathbf{H})\}=0$ and, under assumptions (A1)-(A3),

$$\mathrm{Var}\{A_{1n}(\mathbf{x},\mathbf{H})\}=\frac{\mathbb{E}^2(\epsilon_1)}{n}\sum_{y\in\mathbb{S}_{\mathbf{x},\mathbf{H}}} f(\mathbf{y})r_0^{-2}(\mathbf{y})\{\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}}=\mathbf{y})\}^2=\frac{f(\mathbf{x})\sigma^2}{nr_0^2(\mathbf{x})}\{\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}}=\mathbf{x})\}^2+W,$$

where

$$W=\frac{\sigma^2}{n}\sum_{y\in\mathbb{S}^d_{\mathbf{x},\mathbf{H}}\setminus\{\mathbf{x}\}} r_0^{-2}(\mathbf{y})f(\mathbf{y})\{\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}}=\mathbf{y})\}^2$$

tends to 0 as $n\to\infty$ and $\mathbf{H}\to\mathbf{0_d}$, similar to the term R in equation (15). The other terms in the variance of (19) provide the order $o\Big(\sum_{j=1}^d h_{jj}^2/n\Big)$; we omit to detail here all these calculations. Rather, by applying the Lyapounov's central limit theorem on A_{1n}, we have

$$\sqrt{n}A_{1n}(\mathbf{x},\mathbf{H})\xrightarrow{d}\mathcal{N}(0,f(\mathbf{x})\sigma^2/r_0^2(\mathbf{x})\{\Pr(\mathcal{K}_{\mathbf{x},\mathbf{H}}=\mathbf{x})\}^2).$$

Finally, by considering the convergence of $\widetilde{f_n}$ to f states by **?**, it results that

$$\sqrt{n}\{\widehat{\omega}_n(\mathbf{x}) - \omega(\mathbf{x})\}D_n^{mult}(\mathbf{x}) = \sqrt{n}\{\widehat{\omega}_n(\mathbf{x}) - \omega(\mathbf{x})\}f(\mathbf{x}) + o_p(1) = \mu f(\mathbf{x}) + \sqrt{n}A_{1n}(\mathbf{x}, \mathbf{H}) + o_p(1),$$

such that

$$\mathbb{E}\{\widehat{\omega}_n(\mathbf{x}) - \omega(\mathbf{x})\} = \frac{1}{2}\sum_{j=1}^{d} \mathrm{Var}(\mathcal{K}_{x_j,h_{jj}})\Big\{\omega_{jj}^{(2)}(\mathbf{x}) + 2\omega_j^{(1)}(\mathbf{x})\Big(\frac{f_j^{(1)}}{f}\Big)(\mathbf{x})\Big\} + o\Big(\sum_{j=1}^{d} h_{jj}^2\Big) \to 0, \text{ as } h_{jj} \to 0,$$

and

$$\mathrm{Var}\{\widehat{\omega}_n(\mathbf{x}) - \omega(\mathbf{x})\} = \frac{1}{nr_0^2(\mathbf{x})f(\mathbf{x})}\sigma^2\{\Pr(\mathcal{K}_{\mathbf{x}} = \mathbf{x})\}^2 \prod_{j=1}^{d} h_{jj}^2 + o\Big(\frac{1}{n} + \frac{\sum_{j=1}^{d} h_{jj}^2}{n}\Big).$$

■