

## Article

# Bayesian Bandwidths in Semiparametric Modelling for Nonnegative Orthant Data with Diagnostics

Célestin C. Kokonendji <sup>1</sup>  and Sobom M. Somé <sup>2,\*</sup> 

<sup>1</sup> Laboratoire de Mathématiques de Besançon UMR 6623 CNRS-UBFC, Université Bourgogne Franche-Comté, 16 Route de Gray, 25030 Besançon CEDEX, France; celestin.kokonendji@univ-fcomte.fr

<sup>2</sup> Laboratoire d'Analyse Numérique Informatique et de BIOMathématique, Université Joseph KI-ZERBO, Ouagadougou 03 BP 7021, Burkina Faso

\* Correspondence: sobom.some@gmail.com or sobom.some@univ-ouaga.bf

† Current address: Laboratoire Sciences et Techniques, Université Thomas SANKARA, Ouagadougou 12 BP 417, Burkina Faso.

**Abstract:** Multivariate nonnegative orthant data are real vectors bounded to the left by the null vector, and they can be continuous, discrete or mixed. We first review the recent relative variability indexes for multivariate nonnegative continuous and count distributions. As a prelude, the classification of two comparable distributions having the same mean vector is done through under-, equi- and over-variability with respect to the reference distribution. Multivariate associated kernel estimators are then reviewed with new proposals that can accommodate any nonnegative orthant dataset. We focus on bandwidth matrix selections by adaptive and local Bayesian methods for semicontinuous and counting supports, respectively. We finally introduce a flexible semiparametric approach for estimating all these distributions on nonnegative supports. The corresponding estimator is directed by a given parametric part, and a nonparametric part which is a weight function to be estimated through multivariate associated kernels. A diagnostic model is also discussed to make an appropriate choice between the parametric, semiparametric and nonparametric approaches. The retention of pure nonparametric means the inconvenience of parametric part used in the modelization. Multivariate real data examples in semicontinuous setup as reliability are gradually considered to illustrate the proposed approach. Concluding remarks are made for extension to other multiple functions.

**Keywords:** associated kernel; Bayesian selector; dispersion index; model diagnostics; multivariate distribution; variation index; weighted distribution

**MSC:** 62G07; 62G20; 62G99; 62H10; 62H12



**Citation:** Kokonendji, C.C.; Sobom, S.M. Bayesian Bandwidths in Semiparametric Modelling for Nonnegative Orthant Data with Diagnostics. *Stats* **2021**, *4*, 162–183. <https://doi.org/10.3390/stats4010013>

Academic Editor: Eddy Kwessi

Received: 28 January 2021

Accepted: 1 March 2021

Published: 4 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The  $d$ -variate nonnegative orthant data on  $\mathbb{T}_d^+ \subseteq [0, \infty)^d$  are real  $d$ -vectors bounded to the left by the null vector  $\mathbf{0}_d$ , and they can be continuous, discrete (e.g., count, categorical) or mixed. For simplicity, we here assume either  $\mathbb{T}_d^+ = [0, \infty)^d$  for semicontinuous or  $\mathbb{T}_d^+ = \mathbb{N}^d := \{0, 1, 2, \dots\}^d$  for counting; and, we then omit both setups of categorical and mixed which can be a mix of discrete and continuous data (e.g., [1]) or other time scales (see, e.g., [2]). Modeling such datasets of  $\mathbb{T}_d^+$  requires nonnegative orthant distributions which are generally not easy to handle in practical data analysis. The baseline parametric distribution (e.g., [3–5]) for the analysis of nonnegative continuous data is the exponential distribution (e.g., in Reliability) and that of count data is the Poisson one. However, there intrinsic assumptions of the two first moments are often not realistic for many applications. The nonparametric topic of associated kernels, which is adaptable to any support  $\mathbb{T}_d^+$  of probability density or mass function (pdmf), is widely studied in very recent years. We can refer to [6–15] for general results and more specific developments on associated kernel orthant distributions using classical cross-validation and Bayesian methods to select

bandwidth matrices. Thus, a natural question of flexible semiparametric modeling now arises for all these multivariate orthant datasets.

Indeed, we first need a review of the recent relative variability indexes for multivariate semicontinuous ([16]) and count ([17]) distributions. The infinite number and complexity of multivariate parametric distributions require the study of different indexes for comparisons and discriminations between them. Simple classifications of two comparable distributions are done through under-, equi- and over-variability with respect to the reference distribution. We refer to [18] and references therein for univariate categorical data which does not yet have its multivariate version. We then survey multivariate associated kernels that can accommodate any nonnegative orthant dataset. Most useful families shall be pointed out, mainly as a product of univariate associated kernels and including properties and constructions. We shall focus on bandwidth matrix selections by Bayesian methods. Finally, we have to introduce a flexible semiparametric approach for estimating multivariate nonnegative orthant distributions. Following Hjort and Glad [19] for classical kernels, the corresponding estimator shall be directed by a given parametric part, and a nonparametric part which is a weight function to be estimated through multivariate associated kernels. What does it mean for a diagnostic model to make an appropriate choice between the parametric, semiparametric and nonparametric approaches in this multivariate framework? Such a discussion is to highlight practical improvements on standard nonparametric methods for multivariate semicontinuous datasets, through the use of a reasonable parametric-start description. See, for instance, [20–22] for univariate count datasets.

In this paper, the main goal is to introduce a family of semiparametric estimators with multivariate associated kernels for both semicontinuous and count data. They are meant to be flexible compromises between grueling parametric and fuzzy nonparametric approaches. The rest of the paper is organized as follows. Section 2 presents a brief review of the relative variability indexes for multivariate nonnegative orthant distributions, by distinguishing the dispersion for counting and the variation for semicontinuous. Section 3 displays a short panoply of multivariate associated kernels which are useful for semicontinuous and for counting datasets. Properties are reviewed with new proposals, including both appropriated Bayesian methods of bandwidths selections. In Section 4, we introduce the semiparametric kernel estimators with a  $d$ -variate parametric start. We also investigate the corresponding diagnostic model. Section 5 is devoted to numerical illustrations, especially for uni- and multivariate semicontinuous datasets. In Section 6, we make some final remarks in order to extend to other multiple functions, as regression. Eventually, appendixes are exhibited for technical proofs and illustrations.

## 2. Relative Variability Indexes for Orthant Distributions

Let  $\mathbf{X} = (X_1, \dots, X_d)^\top$  be a nonnegative orthant  $d$ -variate random vector on  $\mathbb{T}_d^+ \subseteq [0, \infty)^d$ ,  $d \geq 1$ . We use the following notations:  $\sqrt{\text{var}\mathbf{X}} = (\sqrt{\text{var}X_1}, \dots, \sqrt{\text{var}X_d})^\top$  is the elementwise square root of the variance vector of  $\mathbf{X}$ ;  $\text{diag}\sqrt{\text{var}\mathbf{X}} = \text{diag}_d(\sqrt{\text{var}X_j})$  is the  $d \times d$  diagonal matrix with diagonal entries  $\sqrt{\text{var}X_j}$  and 0 elsewhere; and,  $\text{cov}\mathbf{X} = (\text{cov}(X_i, X_j))_{i,j \in \{1, \dots, d\}}$  denotes the covariance matrix of  $\mathbf{X}$  which is a  $d \times d$  symmetric matrix with entries  $\text{cov}(X_i, X_j)$  such that  $\text{cov}(X_i, X_i) = \text{var}X_i$  is the variance of  $X_i$ . Then, one has

$$\text{cov}\mathbf{X} = (\text{diag}\sqrt{\text{var}\mathbf{X}})(\rho_{\mathbf{X}})(\text{diag}\sqrt{\text{var}\mathbf{X}}), \quad (1)$$

where  $\rho_{\mathbf{X}} = \rho(\mathbf{X})$  is the correlation matrix of  $\mathbf{X}$ ; see, e.g., Equations (2)–(36) [23]. It is noteworthy that there are many multivariate distributions with exponential (resp., Poisson) margins. Therefore, we denote a generic  $d$ -variate exponential distribution by  $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\rho})$ , given specific positive mean vector  $\boldsymbol{\mu}^{-1} := (\mu_1^{-1}, \dots, \mu_d^{-1})^\top$  and correlation matrix  $\boldsymbol{\rho} = (\rho_{ij})_{i,j \in \{1, \dots, d\}}$ . Similarly, a generic  $d$ -variate Poisson distribution is given by  $\mathcal{P}_d(\boldsymbol{\mu}, \boldsymbol{\rho})$ , with positive mean vector  $\boldsymbol{\mu} := (\mu_1, \dots, \mu_d)^\top$  and correlation matrix  $\boldsymbol{\rho}$ . See, e.g., Appendix A for more extensive exponential and Poisson models with possible behaviours in the negative correlation setup. The uncorrelated or independent  $d$ -variate exponential and Poisson

will be written as  $\mathcal{E}_d(\boldsymbol{\mu})$  and  $\mathcal{P}_d(\boldsymbol{\mu})$ , respectively, for  $\boldsymbol{\rho} = \mathbf{I}_d$  the  $d \times d$  unit matrix. Their respective  $d$ -variate probability density function (pdf) and probability mass function (pmf) are the product of  $d$  univariate ones.

According to [16] and following the recent univariate unification of the well-known (Fisher) dispersion and the (Jørgensen) variation indexes by Touré et al. [24], the relative variability index of  $d$ -variate nonnegative orthant distributions can be written as follows. Let  $\mathbf{X}$  and  $\mathbf{Y}$  be two random vectors on the same support  $\mathbb{T}_d^+ \subseteq [0, \infty)^d$  and assume  $\mathbf{m} := \mathbb{E}\mathbf{X} = \mathbb{E}\mathbf{Y}$ ,  $\boldsymbol{\Sigma}_X := \text{cov}\mathbf{X}$  and  $\mathbf{V}_{F_Y}(\mathbf{m}) := \text{cov}(\mathbf{Y})$  fixed, then the relative variability index of  $\mathbf{X}$  with respect to  $\mathbf{Y}$  is defined as the positive quantity

$$\text{RWI}_Y(\mathbf{X}) := \text{tr}[\boldsymbol{\Sigma}_X \mathbf{W}_{F_Y}^+(\mathbf{m})] \geq 1, \tag{2}$$

where “ $\text{tr}(\cdot)$ ” stands for the trace operator and  $\mathbf{W}_{F_Y}^+(\mathbf{m})$  is the unique Moore-Penrose inverse of the associated matrix  $\mathbf{W}_{F_Y}(\mathbf{m}) := [\mathbf{V}_{F_Y}(\mathbf{m})]^{1/2} [\mathbf{V}_{F_Y}(\mathbf{m})]^\top$  to  $\mathbf{V}_{F_Y}(\mathbf{m})$ . From (2),  $\text{RWI}_Y(\mathbf{X}) \geq 1$  means the over- (equi- and under-variability) of  $\mathbf{X}$  compared to  $\mathbf{Y}$  is realized if  $\text{RWI}_Y(\mathbf{X}) > 1$  ( $\text{RWI}_Y(\mathbf{X}) = 1$  and  $\text{RWI}_Y(\mathbf{X}) < 1$ , respectively).

The expression (2) of RWI does not appear to be very easy to handle in this general formulation on  $\mathbb{T}_d^+ \subseteq [0, \infty)^d$ , even the empirical version and interpretations. We now detail both multivariate cases of counting ([17]) and of semicontinuous ([16]). An R package is recently provided in [25].

### 2.1. Relative Dispersion Indexes for Count Distributions

For  $\mathbb{T}_d^+ = \mathbb{N}^d$ , let  $\mathbf{W}_{F_Y}(\mathbf{m}) = \sqrt{\mathbf{m}} \sqrt{\mathbf{m}}^\top$  be the  $d \times d$  matrix of rank 1. Then,  $\boldsymbol{\Sigma}_X \mathbf{W}_{F_Y}^+(\mathbf{m})$  of (2) is also of rank 1 and has only one positive eigenvalue, denoted by

$$\text{GDI}(\mathbf{X}) := \frac{\sqrt{\mathbb{E}\mathbf{X}}^\top (\text{cov}\mathbf{X}) \sqrt{\mathbb{E}\mathbf{X}}}{\mathbb{E}\mathbf{X}^\top \mathbb{E}\mathbf{X}} \geq 1 \tag{3}$$

and called *generalized dispersion index* of  $\mathbf{X}$  compared to  $\mathbf{Y} \sim \mathcal{P}_d(\mathbb{E}\mathbf{X})$  with  $\mathbb{E}\mathbf{Y} = \mathbb{E}\mathbf{X} = \mathbf{m}$  by [17]. For  $d = 1$ ,  $\text{GDI}(X_1) = \text{var}X_1 / \mathbb{E}X_1 = \text{DI}(X_1)$  is the (Fisher) dispersion index with respect to the Poisson distribution. To derive this interpretation of GDI, we successively decompose the denominator of (3) as

$$\mathbb{E}\mathbf{X}^\top \mathbb{E}\mathbf{X} = \sqrt{\mathbb{E}\mathbf{X}}^\top (\text{diag}\mathbb{E}\mathbf{X}) \sqrt{\mathbb{E}\mathbf{X}} = [(\text{diag} \sqrt{\mathbb{E}\mathbf{X}}) \sqrt{\mathbb{E}\mathbf{X}}]^\top (\mathbf{I}_d) [(\text{diag} \sqrt{\mathbb{E}\mathbf{X}}) \sqrt{\mathbb{E}\mathbf{X}}] \tag{4}$$

and the numerator of (3) by using also (1) as

$$\sqrt{\mathbb{E}\mathbf{X}}^\top (\text{cov}\mathbf{X}) \sqrt{\mathbb{E}\mathbf{X}} = [(\text{diag} \sqrt{\text{var}\mathbf{X}}) \sqrt{\mathbb{E}\mathbf{X}}]^\top (\boldsymbol{\rho}_X) [(\text{diag} \sqrt{\text{var}\mathbf{X}}) \sqrt{\mathbb{E}\mathbf{X}}].$$

Thus,  $\text{GDI}(\mathbf{X})$  makes it possible to compare the full variability of  $\mathbf{X}$  (in the numerator) with respect to its expected uncorrelated Poissonian variability (in the denominator) which depends only on  $\mathbb{E}\mathbf{X}$ . In other words, the count random vector  $\mathbf{X}$  is over- (equi- and under-dispersed) with respect to  $\mathcal{P}_d(\mathbb{E}\mathbf{X})$  if  $\text{GDI}(\mathbf{X}) > 1$  ( $\text{GDI}(\mathbf{X}) = 1$  and  $\text{GDI}(\mathbf{X}) < 1$ , respectively). This is a generalization in multivariate framework of the well-known (univariate) dispersion index by [17]. See, e.g., [17,26] for illustrative examples. We can modify  $\text{GDI}(\mathbf{X})$  to  $\text{MDI}(\mathbf{X})$ , as *marginal dispersion index*, by replacing  $\text{cov}\mathbf{X}$  in (3) with  $\text{diag} \sqrt{\text{var}\mathbf{X}}$  to obtain dispersion information only coming from the margins of  $\mathbf{X}$ .

More generally, for two count random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  on the same support  $\mathbb{T}_d^+ \subseteq \mathbb{N}^d$  with  $\mathbb{E}\mathbf{X} = \mathbb{E}\mathbf{Y}$  and  $\text{GDI}(\mathbf{Y}) > 0$ , the *relative dispersion index* is defined by

$$\text{RDI}_Y(\mathbf{X}) := \frac{\text{GDI}(\mathbf{X})}{\text{GDI}(\mathbf{Y})} = \frac{[(\text{diag} \sqrt{\text{var}\mathbf{X}}) \sqrt{\mathbb{E}\mathbf{X}}]^\top (\boldsymbol{\rho}_X) [(\text{diag} \sqrt{\text{var}\mathbf{X}}) \sqrt{\mathbb{E}\mathbf{X}}]}{[(\text{diag} \sqrt{\text{var}\mathbf{Y}}) \sqrt{\mathbb{E}\mathbf{Y}}]^\top (\boldsymbol{\rho}_Y) [(\text{diag} \sqrt{\text{var}\mathbf{Y}}) \sqrt{\mathbb{E}\mathbf{Y}}]} \geq 1; \tag{5}$$

i.e., the **over-** (equi- and under-dispersion) of  $\mathbf{X}$  compared to  $\mathbf{Y}$  is realized if  $\text{GDI}(\mathbf{X}) > \text{GDI}(\mathbf{Y})$  ( $\text{GDI}(\mathbf{X}) = \text{GDI}(\mathbf{Y})$  and  $\text{GDI}(\mathbf{X}) < \text{GDI}(\mathbf{Y})$ , respectively). Obviously, GDI is a

particular case of RDI with any general reference than  $\mathcal{P}_d$ . Consequently, many properties of GDI are easily extended to RDI.

2.2. Relative Variation Indexes for Semicontinuous Distributions

Assuming here  $\mathbb{T}_d^+ = [0, \infty)^d$  and  $\mathbf{W}_{F_Y}(\mathbf{m}) = \mathbf{m}\mathbf{m}^\top$  another  $d \times d$  matrix of rank 1. Then, we also have that  $\Sigma_X \mathbf{W}_{F_Y}^+(\mathbf{m})$  of (2) is of rank 1. Similar to (3), the *generalized variation index* of  $\mathbf{X}$  compared to  $\mathcal{E}_d(\mathbb{E}\mathbf{X})$  is defined by

$$\text{GVI}(\mathbf{X}) := \frac{\mathbb{E}\mathbf{X}^\top (\text{cov}\mathbf{X}) \mathbb{E}\mathbf{X}}{(\mathbb{E}\mathbf{X}^\top \mathbb{E}\mathbf{X})^2} \geq 1; \tag{6}$$

i.e.,  $\mathbf{X}$  is over- (equi- and under-varied) with respect to  $\mathcal{E}_d(\mathbb{E}\mathbf{X})$  if  $\text{GVI}(\mathbf{X}) > 1$  ( $\text{GVI}(\mathbf{X}) = 1$  and  $\text{GVI}(\mathbf{X}) < 1$ , respectively); see [16]. Remark that when  $d = 1$ ,  $\text{GVI}(X_1) = \text{var}X_1 / (\mathbb{E}X_1)^2 = \text{VI}(X_1)$  is the univariate (Jørgensen) variation index which is recently introduced by Abid et al. [27]. From (4) and using again (1) for rewriting the numerator of (6) as

$$\mathbb{E}\mathbf{X}^\top (\text{cov}\mathbf{X}) \mathbb{E}\mathbf{X} = [(\text{diag}\sqrt{\text{var}\mathbf{X}})\mathbb{E}\mathbf{X}]^\top (\rho_X) [(\text{diag}\sqrt{\text{var}\mathbf{X}})\mathbb{E}\mathbf{X}],$$

$\text{GVI}(\mathbf{X})$  of (6) can be interpreted as the ratio of the full variability of  $\mathbf{X}$  with respect to its expected uncorrelated exponential  $\mathcal{E}_d(\mathbb{E}\mathbf{X})$  variability which depends only on  $\mathbb{E}\mathbf{X}$ . Similar to  $\text{MDI}(\mathbf{X})$ , we can define  $\text{MVI}(\mathbf{X})$  from  $\text{GVI}(\mathbf{X})$ . See [16] for properties, numerous examples and numerical illustrations.

The relative variation index is defined, for two semicontinuous random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  on the same support  $\mathbb{T}_d^+ = [0, \infty)^d$  with  $\mathbb{E}\mathbf{X} = \mathbb{E}\mathbf{Y}$  and  $\text{GVI}(\mathbf{Y}) > 0$ , by

$$\text{RVI}_Y(\mathbf{X}) := \frac{\text{GVI}(\mathbf{X})}{\text{GVI}(\mathbf{Y})} = \frac{[(\text{diag}\sqrt{\text{var}\mathbf{X}})\mathbb{E}\mathbf{X}]^\top (\rho_X) [(\text{diag}\sqrt{\text{var}\mathbf{X}})\mathbb{E}\mathbf{X}]}{[(\text{diag}\sqrt{\text{var}\mathbf{Y}})\mathbb{E}\mathbf{Y}]^\top (\rho_Y) [(\text{diag}\sqrt{\text{var}\mathbf{Y}})\mathbb{E}\mathbf{Y}]} \geq 1; \tag{7}$$

i.e., the over- (equi- and under-variation) of  $\mathbf{X}$  compared to  $\mathbf{Y}$  is carried out if  $\text{GVI}(\mathbf{X}) > \text{GVI}(\mathbf{Y})$  ( $\text{GVI}(\mathbf{X}) = \text{GVI}(\mathbf{Y})$  and  $\text{GVI}(\mathbf{X}) < \text{GVI}(\mathbf{Y})$ , respectively). Of course, RVI generalizes GVI for multivariate semicontinuous distributions. For instance, one refers to [16] for more details on its discriminating power in multivariate parametric models from two first moments.

3. Multivariate Orthant Associated Kernels

Nonparametric techniques through associated kernels represent an alternative approach for multivariate orthant data. Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be independent and identically distributed (iid) nonnegative orthant  $d$ -variate random vectors with an unknown joint pdf  $f$  on  $\mathbb{T}_d^+ \subseteq [0, \infty)^d$ , for  $d \geq 1$ . Then the multivariate associated kernel estimator  $\tilde{f}_n$  of  $f$  is expressed as

$$\tilde{f}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathbf{K}_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i), \quad \forall \mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{T}_d^+, \tag{8}$$

where  $\mathbf{H}$  is a given  $d \times d$  bandwidth matrix (i.e., symmetric and positive definite) such that  $\mathbf{H} \equiv \mathbf{H}_n \rightarrow \mathbf{0}_d$  (the  $d \times d$  null matrix) as  $n \rightarrow \infty$ , and  $\mathbf{K}_{\mathbf{x},\mathbf{H}}(\cdot)$  is a multivariate (orthant) associated kernel, parameterized by  $\mathbf{x}$  and  $\mathbf{H}$ ; see, e.g., [10]. More precisely, we have the following refined definition.

**Definition 1.** Let  $\mathbb{T}_d^+$  be the support of the pdf to be estimated,  $\mathbf{x} \in \mathbb{T}_d^+$  a target vector and  $\mathbf{H}$  a bandwidth matrix. A parameterized pdf  $\mathbf{K}_{\mathbf{x},\mathbf{H}}(\cdot)$  on support  $\mathbb{S}_{\mathbf{x},\mathbf{H}} \subseteq \mathbb{T}_d^+$  is called “multivariate orthant associated kernel” if the following conditions are satisfied:

$$\mathbf{x} \in \mathbb{S}_{\mathbf{x},\mathbf{H}}, \quad \mathbb{E}\mathbf{Z}_{\mathbf{x},\mathbf{H}} = \mathbf{x} + \mathbf{A}(\mathbf{x}, \mathbf{H}) \rightarrow \mathbf{x} \text{ and } \text{cov}\mathbf{Z}_{\mathbf{x},\mathbf{H}} = \mathbf{B}(\mathbf{x}, \mathbf{H}) \rightarrow \mathbf{0}_d^+,$$

where  $\mathcal{Z}_{\mathbf{x},\mathbf{H}}$  denotes the corresponding orthant random vector with pmf  $\mathbf{K}_{\mathbf{x},\mathbf{H}}$  such that vector  $\mathbf{A}(\mathbf{x}, \mathbf{H}) \rightarrow \mathbf{0}$  (the  $d$ -dimensional null vector) and positive definite matrix  $\mathbf{B}(\mathbf{x}, \mathbf{H}) \rightarrow \mathbf{0}_d^+$  as  $\mathbf{H} \rightarrow \mathbf{0}_d$  (the  $d \times d$  null matrix), and  $\mathbf{0}_d^+$  stands for a symmetric matrix with entries  $u_{ij}$  for  $i, j = 1, \dots, d$  such that  $u_{ij} \in [0, 1)$ .

This definition exists in the univariate count case of [21,28] and encompasses the multivariate one by [10]. The choice of the orthant associated kernel satisfying  $\lim_{\mathbf{H} \rightarrow \mathbf{0}_d} \text{Cov}(\mathcal{Z}_{\mathbf{x},\mathbf{H}}) = \mathbf{0}_d$  assures the convergence of its corresponding estimator named of the second order. Otherwise, the convergence of its corresponding estimator is not guarantee for  $u_{ij} \in (0, 1)$ , a right neighborhood of 0, in Definition 1 and it is said a consistent first-order smoother; see, e.g., [28] for discrete kernels. In general,  $d$ -under-dispersed count associated kernels are appropriated for both small and moderate sample sizes; see, e.g., [28] for univariate cases. As for the selection of the bandwidth  $\mathbf{H}$ , it is very crucial because it controls the degree of smoothing and the form of the orientation of the kernel. As a matter of fact, a simplification can be obtained by considering a diagonal matrix  $\mathbf{H} = \text{diag}_d(h_j)$ . Since it is challenging to obtain a full multivariate orthant distribution  $\mathbf{K}_{\mathbf{x},\mathbf{H}}(\cdot)$  for building a smoother, several authors suggest the product of univariate orthant associated kernels,

$$\mathbf{K}_{\mathbf{x},\mathbf{H}}(\cdot) = \prod_{j=1}^d K_{x_j, h_j}(\cdot), \tag{9}$$

where  $K_{x_j, h_j}$ ,  $j = 1, \dots, d$ , belong either to the same family or to different families of univariate orthant associated kernels. The following two subsections are devoted to the summary of discrete and semicontinuous univariate associated kernels.

Before showing some main properties of the associated kernel estimator (8), let us recall that the family of  $d$ -variate classical (symmetric) kernels  $\mathbf{K}$  on  $\mathbb{S}_d \subseteq \mathbb{R}^d$  (e.g., [29–31]) can be also presented as (classical) associated kernels. Indeed, from (8) and writting for instance

$$\mathbf{K}_{\mathbf{x},\mathbf{H}}(\cdot) = (\det \mathbf{H})^{-1/2} \mathbf{K}[\mathbf{H}^{-1/2}(\mathbf{x} - \cdot)]$$

where “det” is the determinant operator, one has  $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbf{x} - \mathbf{H}^{-1/2} \mathbb{S}_d$ ,  $\mathbf{A}(\mathbf{x}, \mathbf{H}) = \mathbf{0}$  and  $\mathbf{B}(\mathbf{x}, \mathbf{H}) = \mathbf{H}^{1/2} \mathbf{I}_d \mathbf{H}^{1/2} = \mathbf{H}$ . In general, one uses the classical (associated) kernels for smoothing continuous data or pdf having support  $\mathbb{T}_d = \mathbb{R}^d$ .

The purely nonparametric estimator (8) with multivariate associated kernel,  $\tilde{f}_n$  of  $f$ , is generally defined up to the normalizing constant  $C_n$ . Several simulation studies (e.g., Table 3.1 in [10]) are shown that  $C_n = C_n(\mathbf{K}, \mathbf{H})$  (depending on samples, associated kernels and bandwidths) is approximatively 1. Without loss of generality, one here assumes  $C_n = 1$  as for all classical (associated) kernel estimators of pdf. The following proposition finally proves its mean behavior and variability through the integrated bias and integrated variance of  $\tilde{f}_n$ , respectively. In what follows, let us denote by  $\nu$  the reference measure (Lebesgue or counting) on the nonnegative orthant set  $\mathbb{T}_d^+$  and also on any set  $\mathbb{T}_d \subseteq \mathbb{R}^d$ .

**Proposition 1.** Let  $C_n := \int_{\mathbb{T}_d} \tilde{f}_n(\mathbf{x}) \nu(d\mathbf{x}) = C_n(\mathbf{K}, \mathbf{H})$ . Then, for all  $n \geq 1$ :

$$\mathbb{E}(C_n) = 1 + \int_{\mathbb{T}_d} \text{Bias}\{\tilde{f}_n(\mathbf{x})\} \nu(d\mathbf{x}) \quad \text{and} \quad \text{var}(C_n) = \int_{\mathbb{T}_d} \text{var}\{\tilde{f}_n(\mathbf{x})\} \nu(d\mathbf{x}).$$

**Proof.** Let  $n \geq 1$ . One successively has

$$\mathbb{E}(C_n) = \int_{\mathbb{T}_d} [f(\mathbf{x}) + \mathbb{E}\{\tilde{f}_n(\mathbf{x})\} - f(\mathbf{x})] \nu(d\mathbf{x}) = \int_{\mathbb{T}_d} f(\mathbf{x}) \nu(d\mathbf{x}) + \int_{\mathbb{T}_d} [\mathbb{E}\{\tilde{f}_n(\mathbf{x})\} - f(\mathbf{x})] \nu(d\mathbf{x}),$$

which leads to the first result because  $f$  is a pmf on  $\mathbb{T}_d$ . The second result on  $\text{var}(C_n)$  is trivial.  $\square$

The following general result is easily deduced from Proposition 1. To the best of our knowledge, it appears to be new and interesting in the framework of the pdmf (associated) kernel estimators.

**Corollary 1.** *If  $C_n = 1$ , for all  $n \geq 1$ , then:  $\int_{\mathbb{T}_d} \text{Bias}\{\tilde{f}_n(\mathbf{x})\} \nu(d\mathbf{x}) = 0$  and  $\int_{\mathbb{T}_d} \text{var}\{\tilde{f}_n(\mathbf{x})\} \nu(d\mathbf{x}) = 0$ .*

In particular, Corollary 1 holds for all classical (associated) kernel estimators. The two following properties on the corresponding orthant multivariate associated kernels shall be needed subsequently.

**(K1)** There exists the second moment of  $\mathbf{K}_{\mathbf{x},\mathbf{H}}$ :

$$\mu_j^2(\mathbf{K}_{\mathbf{x},\mathbf{H}}) := \int_{\mathbb{S}_{\mathbf{x},\mathbf{H}} \cap \mathbb{T}_d^+} u_j^2 \mathbf{K}_{\mathbf{x},\mathbf{H}}(\mathbf{u}) \nu(d\mathbf{u}) < \infty, \quad \forall j = 1, \dots, d.$$

**(K2)** There exists a real largest number  $r = r(\mathbf{K}_{\mathbf{x},\mathbf{H}}) > 0$  and  $0 < c(\mathbf{x}) < \infty$  such that

$$\|\mathbf{K}_{\mathbf{x},\mathbf{H}}\|_2^2 := \int_{\mathbb{S}_{\mathbf{x},\mathbf{H}} \cap \mathbb{T}_d^+} \{\mathbf{K}_{\mathbf{x},\mathbf{H}}(\mathbf{u})\}^2 \nu(d\mathbf{u}) \leq c(\mathbf{x})(\det \mathbf{H})^{-r}.$$

In fact, **(K1)** is a necessary condition for smoothers to have a finite variance and **(K2)** can be deduced from the continuous univariate cases (e.g., [32]) and also from the discrete ones (e.g., [28]).

We now establish both general asymptotic behaviours of the pointwise bias and variance of the nonparametric estimator (8) on the nonnegative orthant set  $\mathbb{T}_d^+$ ; its proof is given in Appendix B. For that, we need the following assumptions by endowing  $\mathbb{T}_d^+$  with the Euclidean norm  $\|\cdot\|$  and the associated inner product  $\langle \cdot, \cdot \rangle$  such that  $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b}$ .

- (a1)** The unknown pdmf  $f$  is a bounded function and twice differentiable or finite difference in  $\mathbb{T}_d^+$  and  $\nabla f(\mathbf{x})$  and  $\mathcal{H}f(\mathbf{x})$ , which denote, respectively, the gradient vector (in the continuous or discrete sense, respectively) and the corresponding Hessian matrix of the function  $f$  at  $\mathbf{x}$ .
- (a2)** There exists a positive real number  $r > 0$  such that  $\|\mathbf{K}_{\mathbf{x},\mathbf{H}_n}\|_2^2 (\det \mathbf{H}_n)^r \rightarrow c_1(\mathbf{x}) > 0$  as  $n \rightarrow \infty$ .

Note that **(a2)** is obviously a consequence of **(K2)**.

**Proposition 2.** *Under the assumption **(a1)** on  $f$ , then the estimator  $\tilde{f}_n$  in (8) of  $f$  verifies*

$$\text{Bias}\{\tilde{f}_n(\mathbf{x})\} = \langle \nabla f(\mathbf{x}), \mathbf{A}(\mathbf{x}, \mathbf{H}_n) \rangle + \frac{1}{2} \text{tr}\{\mathcal{H}f(\mathbf{x})[\mathbf{B}(\mathbf{x}, \mathbf{H}_n) + \mathbf{A}(\mathbf{x}, \mathbf{H}_n)^\top \mathbf{A}(\mathbf{x}, \mathbf{H}_n)]\} + o\{\text{tr}\{\mathbf{B}(\mathbf{x}, \mathbf{H}_n)\}\}, \quad (10)$$

for any  $\mathbf{x} \in \mathbb{T}_d^+$ . Moreover, if **(a2)** holds then

$$\text{var}\{\tilde{f}_n(\mathbf{x})\} = \frac{1}{n} f(\mathbf{x}) \|\mathbf{K}_{\mathbf{x},\mathbf{H}_n}\|_2^2 + o\left[\frac{1}{n(\det \mathbf{H}_n)^r}\right]. \quad (11)$$

For  $d = 1$  and according to the proof of Proposition 2, one can easily write  $\mathbb{E}\widehat{f}_n(x)$  as follows:

$$\mathbb{E}\widehat{f}_n(x) = \mathbb{E}f(\mathcal{Z}_{x,h}) = \sum_{k \geq 0} \frac{1}{k!} \mathbb{E}(\mathcal{Z}_{x,h} - \mathbb{E}\mathcal{Z}_{x,h})^k f^{(k)}(\mathbb{E}\mathcal{Z}_{x,h}),$$

where  $f^{(k)}$  is the  $k$ th derivative or finite difference of the pdmf  $f$  under the existence of the centered moment of order  $k \geq 2$  of  $\mathcal{Z}_{x,h}$ .

Concerning bandwidth matrix selections in a multivariate associated kernel estimator (8), one generally use the cross-validation technique (e.g., [10,20,28,33,34]). However, it

is tedious and less precise. Many papers have recently proposed Bayesian approaches (e.g., [6,7,13,14,35,36] and references therein). In particular, they have recommended local Bayesian for discrete smoothing of pmf (e.g., [6,7,37]) and adaptive one for continuous smoothing of pdf (e.g., [13,35,36]).

Denote  $\mathcal{M}$  the set of positive definite [diagonal] matrices [from (9), resp.] and let  $\pi$  be a given suitable prior distribution on  $\mathcal{M}$ . Under the squared error loss function, the Bayes estimator of  $\mathbf{H}$  is the mean of the posterior distribution. Then, the local Bayesian bandwidth at the target  $\mathbf{x} \in \mathbb{T}_d^+$  takes the form

$$\tilde{\mathbf{H}}(\mathbf{x}) := \int_{\mathcal{M}} \mathbf{H} \pi(\mathbf{H}) \tilde{f}_n(\mathbf{x}) d\mathbf{H} \left[ \int_{\mathcal{M}} \tilde{f}_n(\mathbf{x}) \pi(\mathbf{H}) d\mathbf{H} \right]^{-1}, \quad \mathbf{x} \in \mathbb{T}_d^+, \quad (12)$$

and the adaptive Bayesian bandwidth for each observation  $\mathbf{X}_i \in \mathbb{T}_d^+$  of  $\mathbf{X}$  is given by

$$\tilde{\mathbf{H}}_i := \int_{\mathcal{M}} \mathbf{H}_i \pi(\mathbf{H}_i) \tilde{f}_{n, \mathbf{H}_i, -i}(\mathbf{X}_i) d\mathbf{H}_i \left[ \int_{\mathcal{M}} \tilde{f}_{n, \mathbf{H}_i, -i}(\mathbf{X}_i) \pi(\mathbf{H}_i) d\mathbf{H}_i \right]^{-1}, \quad i = 1, \dots, n, \quad (13)$$

where  $\tilde{f}_{n, \mathbf{H}_i, -i}(\mathbf{X}_i)$  is the leave-one-out associated kernel estimator of  $f(\mathbf{X}_i)$  deduced from (8) as

$$\tilde{f}_{n, \mathbf{H}_i, -i}(\mathbf{X}_i) := \frac{1}{n-1} \sum_{\ell=1, \ell \neq i}^n \mathbf{K}_{\mathbf{X}_i, \mathbf{H}_i}(\mathbf{X}_\ell). \quad (14)$$

Note that the well-known and classical (global) cross-validation bandwidth matrix  $\tilde{\mathbf{H}}_{CV}$  and the global Bayesian one  $\tilde{\mathbf{H}}_B$  are obtained, respectively, from (14) as

$$\tilde{\mathbf{H}}_{CV} := \arg \min_{\mathbf{H} \in \mathcal{M}} \left[ \int_{\mathbb{T}_d^+} \{\tilde{f}_n(\mathbf{x})\}^2 \nu(d\mathbf{x}) - \frac{2}{n} \sum_{i=1}^n \tilde{f}_{n, \mathbf{H}, -i}(\mathbf{X}_i) \right]$$

and

$$\tilde{\mathbf{H}}_B := \int_{\mathcal{M}} \mathbf{H} \pi(\mathbf{H}) d\mathbf{H} \prod_{i=1}^n \tilde{f}_{n, \mathbf{H}, -i}(\mathbf{X}_i) d\mathbf{H} \left[ \int_{\mathcal{M}} \pi(\mathbf{H}) d\mathbf{H} \prod_{i=1}^n \tilde{f}_{n, \mathbf{H}, -i}(\mathbf{X}_i) \right]^{-1}.$$

### 3.1. Discrete Associated Kernels

We only present three main and useful families of univariate discrete associated kernels for (9) and satisfying (K1) and (K2).

**Example 1** (categorical). For fixed  $c \in \{2, 3, \dots\}$ , the number of categories and  $\mathbb{T}_1^+ = \{0, 1, \dots, c-1\}$ , one defines the Dirac discrete uniform (DirDU) kernel by

$$K_{x,h}^{DirDU}(u) = (1-h)^{\mathbb{1}_{u=x}} \left( \frac{h}{c-1} \right)^{1-\mathbb{1}_{u=x}},$$

for  $x \in \{0, 1, \dots, c-1\}$ ,  $h \in (0, 1]$ , with  $\mathbb{S}_x := \{0, 1, \dots, c-1\} = \mathbb{T}_1^+$ ,  $A(x, h) = h\{c/2 - x - x/(c-1)\}$  and  $B(x, h) = h\{c(2c-1)/6 + x^2 - xc + x^2/(c-1)\} - h^2\{c/2 - x - x/(c-1)\}^2$ .

It was introduced in the multivariate setup by Aitchison and Aitken [38] and investigated as a discrete associated kernel which is symmetric to the target  $x$  by [28] in univariate case; see [7] for a Bayesian approach in multivariate setup. Note here that its normalized constant is always  $1 = C_n$ .

**Example 2** (symmetric count). For fixed  $m \in \mathbb{N}$  and  $\mathbb{T}_1^+ \subseteq \mathbb{Z}$ , the symmetric count triangular kernel is expressed as

$$K_{x,h}^{SCTriang}(u) = \frac{(m+1)^h - |u-x|^h}{P(m,h)} \mathbb{1}_{\{x, x \pm 1, \dots, x \pm m\}}(u),$$

for  $x \in \mathbb{T}_1^+, h > 0$ , with  $\mathbb{S}_x := \{x, x \pm 1, \dots, x \pm m\}, P(m, h) = (2m+1)(m+1)^h - 2 \sum_{\ell=0}^m \ell^h, A(x, h) = 0$  and

$$\begin{aligned} B(x, h) &= \frac{1}{P(m, h)} \left\{ \frac{m(2m+1)(m+1)^{h+1}}{3} - 2 \sum_{\ell=0}^m \ell^{h+2} \right\} \\ &\approx h \left\{ \frac{m(2m^2+3m+1)}{3} \log(m+1) - 2 \sum_{\ell=1}^m \ell^2 \log \ell \right\} + O(h^2), \end{aligned}$$

where  $\approx$  holds for  $h$  sufficiently small.

It was first proposed by Kokonendji et al. [33] and then completed in [39] with an asymmetric version for solving the problem of boundary bias in count kernel estimation.

**Example 3** (standard count). Let  $\mathbb{T}_1^+ \subseteq \mathbb{N}$ , the standard binomial kernel is defined by

$$K_{x,h}^{Binomial}(u) = \frac{(x+1)!}{u!(x+1-u)!} \left(\frac{x+h}{x+1}\right)^u \left(\frac{1-h}{x+1}\right)^{x+1-u} \mathbb{1}_{\{0,1,\dots,x+1\}}(u),$$

for  $x \in \mathbb{T}_1^+, h \in (0, 1]$ , with  $\mathbb{S}_x := \{0, 1, \dots, x+1\}, A(x, h) = h$  and  $B(x, h) = (x+h)(1-h)/(x+1) \rightarrow x/(x+1) \in [0, 1]$  as  $h \rightarrow 0$ .

Here,  $B(x, h)$  tends to  $x/(x+1) \in [0, 1)$  when  $h \rightarrow 0$  and the new Definition 1 holds. This first-order and under-dispersed binomial kernel is introduced in [28] which becomes very useful for smoothing count distribution through small or moderate sample size; see, e.g., [6,7,37] for Bayesian approaches and some references therein. In addition, we have the standard Poisson kernel where  $K_{x,h}^{Poisson}$  follows the equi-dispersed Poisson distribution with mean  $x+h, \mathbb{S}_x := \mathbb{N} =: \mathbb{T}_1^+, A(x, h) = h$  and  $B(x, h) = x+h \rightarrow x \in \mathbb{N}$  as  $h \rightarrow 0$ . Recently, Huang et al. [40] have introduced the Conway-Maxwell-Poisson kernel by exploiting its under-dispersed part and its second-order consistency which can be improved via the mode-dispersion approach of [41]; see also Section 2.4 in [42].

### 3.2. Semicontinuous Associated Kernels

Now, we point out eight main and useful families of univariate semicontinuous associated kernels for (9) and satisfying (K1) and (K2). Which are gamma (G) of [43] (see also [44]), inverse gamma (Ig) (see also [45]) and log-normal 2 (LN2) by [41], inverse Gaussian (IG) and reciprocal inverse Gaussian by [46] (see also [47]), log-normal 1 (LN1) and Birnbaum–Saunders by [48] (see also [49,50]), and Weibull (W) of [51] (see also [50]). It is noteworthy that the link between LN2 of [41] and LN1 of [48] is through changing  $(x, h)$  to  $(x \exp(h^2), 2 \sqrt{\log(1+h)})$ . Several other semicontinuous could be constructed by using the mode-dispersion technique of [41] from any semicontinuous distribution which is unimodal and having a dispersion parameter. Recently, one has the scaled inverse chi-squared kernel of [52].

Table 1 summarizes these eight semicontinuous univariate associated kernels with their ingredients of Definition 1 and order of preference (O.) obtained graphically. In fact, the heuristic classification (O.) is done through the behavior of the shape and scale of the associated kernel around the target  $x$  at the edge as well as inside; see Figure 1 for edge and Figure 2 for inside. Among these eight kernels, we thus have to recommend the five first

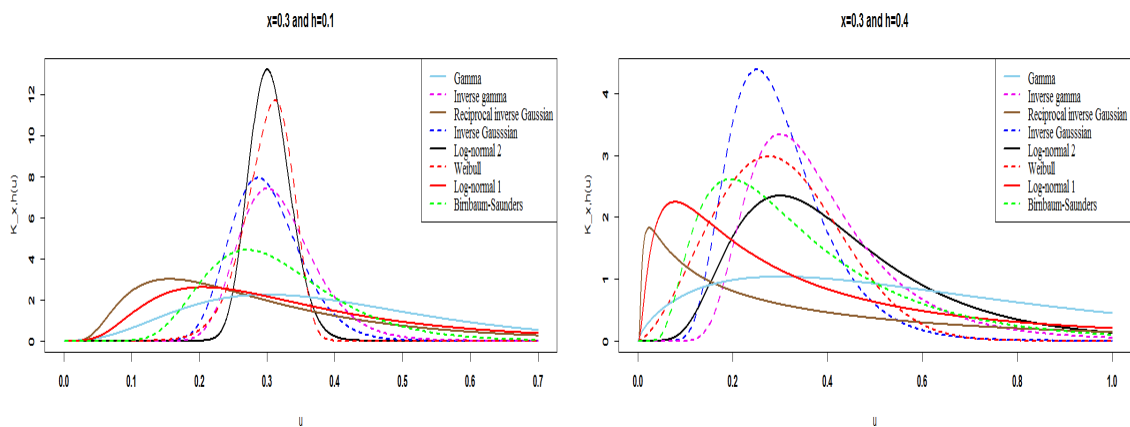


univariate associated kernels of Table 1 for smoothing semicontinuous data. This approach could be improved by a given dataset; see, e.g., [53] for cumulative functions.

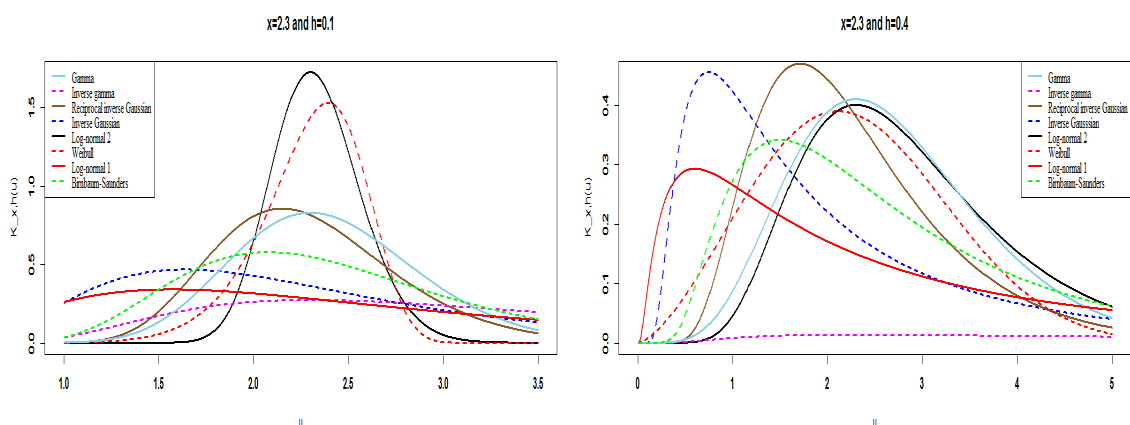
**Table 1.** Eight semicontinuous univariate associated kernels on  $\mathbb{S}_{x,h} \subseteq [0, \infty)$  and classified by “O.”

O.	Name	$K_{x,h}(u)$	$A(x, h)$	$B(x, h)$
1	LN2 [41]	$(uh\sqrt{2\pi})^{-1} \exp\left(\frac{[\log\{x \exp(h^2)\} - \log u]}{2h^2}\right)$	$x[\exp(3h^2/2) - 1]$	$x^2 \exp(3h^2)[\exp(h^2) - 1]$
2	W [51]	$[\Gamma(h)/x][u\Gamma(1+h)/x]^{1/h-1} \exp\{-[u\Gamma(1+h)/x]^{1/h}\}$	0	$x^2[\Gamma(1+2h)/\Gamma^2(1+2h) - 1]$
3	G [43]	$h^{-1-x/h}u^{x/h} \exp(-u/h)/\Gamma(1+x/h)$	$h$	$(x+h)h$
4	BS [48]	$(uh\sqrt{2\pi})^{-1}[(xu)^{-1/2} + (x/u^3)^{-1/2}] \exp[(2-u/x-x/u)/2h]$	$xh/2$	$x^2h(2+5h/2)/2$
5	Ig [41]	$h^{1-1/xh}u^{-1/xh} \exp(-1/uh)/\Gamma(1/xh-1)$	$2x^2h/(1-2xh)$	$x^3h/[(1-3xh)(1-2xh)^2]$
6	RIG [46]	$(\sqrt{2\pi uh})^{-1} \exp\{[x-h][2-(x-h)/u-u/(x-h)]/2h\}$	0	$(x-h)h$
7	IG [46]	$(\sqrt{2\pi hu^3})^{-1} \exp\{[2-x/u-u/x]/2hx\}$	0	$x^3h$
8	LN1 [48]	$(u\sqrt{8\pi \log(1+h)})^{-1} \exp(-[\log u - \log x]^2/[8 \log(1+h)])$	$xh(h+2)$	$x^2(1+h)^4[(1+h)^4 - 1]$

$\Gamma(v) := \int_0^\infty s^{v-1} \exp(-s) ds$  is the classical gamma function with  $v > 0$ .



**Figure 1.** Comparative graphics of the eight univariate semicontinuous associated kernels of Table 1 on the edge ( $x = 0.3$ ) with  $h = 0.1$  and  $h = 0.4$ .



**Figure 2.** Comparative graphics of the eight univariate semicontinuous associated kernels of Table 1 inside ( $x = 2.3$ ) with  $h = 0.1$  and  $h = 0.4$ .

#### 4. Semiparametric Kernel Estimation with $d$ -Variate Parametric Start

We investigate the semiparametric orthant kernel approach which is a compromise between the pure parametric and the nonparametric methods. This concept was proposed by Hjort and Glad [19] for continuous data, treated by Kokonendji et al. [20] for discrete univariate data and, recently, studied by Kokonendji et al. [21] with an application to radiation biodosimetry.

Without loss of generality, we here assume that any  $d$ -variate pdf  $f$  can be formulated (e.g., [54] for  $d = 1$ ) as

$$f(\mathbf{x}) = w(\mathbf{x}; \boldsymbol{\theta}) p_d(\mathbf{x}; \boldsymbol{\theta}), \quad \forall \mathbf{x} \in \mathbb{T}_d^+, \quad (15)$$

where  $p_d(\cdot; \boldsymbol{\theta})$  is the non-singular parametric part according to a reference  $d$ -variate distribution with corresponding unknown parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^\top$  and  $w(\cdot; \boldsymbol{\theta}) := f(\cdot)/p_d(\cdot; \boldsymbol{\theta})$  is the unknown orthant weight function part, to be estimated with a multivariate orthant associated kernel. The weight function at each point can be considered as the local multiplicative correction factor aimed to accommodate any pointwise departure from the reference  $d$ -variate distribution. However, one cannot consider the best fit of parametric models as the start distribution in this semiparametric approach. Because the corresponding weight function is close to zero and becomes a noise which is unappropriated to smooth by an associated kernel, especially for the continuous cases.

Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be iid nonnegative orthant  $d$ -variate random vectors with unknown pdf  $f$  on  $\mathbb{T}_d^+ \subseteq [0, \infty)^d$ . The semiparametric estimator of (15) with (9) is expressed as follows:

$$\begin{aligned} \widehat{f}_n(\mathbf{x}) &= p_d(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n) \frac{1}{n} \sum_{i=1}^n \frac{1}{p_d(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_n)} \mathbf{K}_{\mathbf{x}, \mathbf{H}}(\mathbf{X}_i) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{p_d(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n)}{p_d(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_n)} \mathbf{K}_{\mathbf{x}, \mathbf{H}}(\mathbf{X}_i), \quad \mathbf{x} \in \mathbb{T}_d^+, \end{aligned} \quad (16)$$

where  $\widehat{\boldsymbol{\theta}}_n$  is the estimated parameter of  $\boldsymbol{\theta}$ . From (16), we then deduce the nonparametric orthant associated kernel estimate

$$\widetilde{w}_n(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n) = \frac{1}{n} \sum_{i=1}^n \frac{1}{p_d(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_n)} \mathbf{K}_{\mathbf{x}, \mathbf{H}}(\mathbf{X}_i) \quad (17)$$

of the weight function  $x \mapsto w(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n)$  which depends on  $\widehat{\boldsymbol{\theta}}_n$ . One can observe that Proposition 1 also holds for  $\widehat{f}_n(\cdot) = p_d(\cdot; \widehat{\boldsymbol{\theta}}_n) \widetilde{w}_n(\cdot; \widehat{\boldsymbol{\theta}}_n)$ . However, we have to prove below the analogous fact to Proposition 2.

##### 4.1. Known $d$ -Variate Parametric Model

Let  $p_d(\cdot; \boldsymbol{\theta}_0)$  be a fixed orthant distribution in (15) with  $\boldsymbol{\theta}_0$  known. Writing  $f(\mathbf{x}) = p_d(\mathbf{x}; \boldsymbol{\theta}_0) w(\mathbf{x})$ , we estimate the nonparametric weight function  $w$  by  $\widetilde{w}_n(\mathbf{x}) = n^{-1} \sum_{i=1}^n \mathbf{K}_{\mathbf{x}, \mathbf{H}}(\mathbf{X}_i) / p_d(\mathbf{X}_i; \boldsymbol{\theta}_0)$  with an orthant associated kernel method, resulting in the estimator

$$\widehat{f}_n(\mathbf{x}) = p_d(\mathbf{x}; \boldsymbol{\theta}_0) \widetilde{w}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{p_d(\mathbf{x}; \boldsymbol{\theta}_0)}{p_d(\mathbf{X}_i; \boldsymbol{\theta}_0)} \mathbf{K}_{\mathbf{x}, \mathbf{H}}(\mathbf{X}_i), \quad \mathbf{x} \in \mathbb{T}_d^+. \quad (18)$$

The following proposition is proven in Appendix B.

**Proposition 3.** Under the assumption (a1) on  $f(\cdot) = p_d(\cdot; \theta_0)w(\cdot)$ , then the estimator  $\widehat{f}_n(\cdot) = p_d(\cdot; \theta_0)\widetilde{w}_n(\cdot)$  in (18) of  $f$  satisfies

$$\begin{aligned} \text{Bias}\{\widehat{f}_n(\mathbf{x})\} &= p_d(\mathbf{x}; \theta_0)[w(\mathbf{x}) - f(\mathbf{x})\{p_d(\mathbf{x}; \theta_0)\}^{-1} + \langle \nabla w(\mathbf{x}), \mathbf{A}(\mathbf{x}, \mathbf{H}_n) \rangle] \\ &\quad + \frac{1}{2} p_d(\mathbf{x}; \theta_0) \left( \text{tr}\{\mathcal{H}w(\mathbf{x})[\mathbf{B}(\mathbf{x}, \mathbf{H}_n) + \mathbf{A}(\mathbf{x}, \mathbf{H}_n)^\top \mathbf{A}(\mathbf{x}, \mathbf{H}_n)]\} \right) \\ &\quad + (1 + o\{\text{tr}[\mathbf{B}(\mathbf{x}, \mathbf{H}_n)]\}), \end{aligned}$$

for any  $\mathbf{x} \in \mathbb{T}_d^+$ . Furthermore, if (a2) holds then one has  $\text{var}\{\widehat{f}_n(\mathbf{x})\} = \text{var}\{\widetilde{f}_n(\mathbf{x})\}$  of (11).

It is expected that the bias here is quite different from that of (10).

#### 4.2. Unknown $d$ -Variate Parametric Model

Let us now consider the more realistic and practical semiparametric estimator  $\widehat{f}_n(\cdot) = p_d(\cdot; \widehat{\theta}_n)\widetilde{w}_n(\cdot; \widehat{\theta}_n)$  presented in (16) of  $f(\cdot) = p_d(\cdot; \theta)w(\cdot; \theta)$  in (15) such that the parametric estimator  $\widehat{\theta}_n$  of  $\theta$  can be obtained by the maximum likelihood method; see [19] for quite a general estimator of  $\theta$ . In fact, if the  $d$ -variate parametric model  $p_d(\cdot; \theta)$  is misspecified then this  $\widehat{\theta}_n$  converges in probability to the pseudotrue value  $\theta_0$  satisfying

$$\theta_0 := \arg \min_{\theta} \int_{\mathbf{x} \in \mathbb{T}_d^+} f(\mathbf{x}) \log[f(\mathbf{x})/p_d(\mathbf{x}; \theta)] \nu(d\mathbf{x})$$

from the Kullback–Leibler divergence (see, e.g., [55]).

By writing  $p_0(\cdot) := p_d(\cdot; \theta_0)$  this best  $d$ -variate parametric approximant, but this  $p_0(\cdot)$  is not explicitly expressible as the one in (18). According to [19] (see also [20]), we can represent the proposed estimator  $\widehat{f}_n(\cdot) = p_d(\cdot; \widehat{\theta}_n)\widetilde{w}_n(\cdot; \widehat{\theta}_n)$  in (16) as

$$\widehat{f}_n(\mathbf{x}) \doteq \frac{1}{n} \sum_{i=1}^n \frac{p_0(\mathbf{x})}{p_0(\mathbf{X}_i)} \mathbf{K}_{\mathbf{x}, \mathbf{H}}(\mathbf{X}_i), \quad \mathbf{x} \in \mathbb{T}_d^+. \tag{19}$$

Thus, the following result provides approximate bias and variance. We omit its proof since it is analogous to the one of Proposition 3.

**Proposition 4.** Let  $p_0(\cdot) := p_d(\cdot; \theta_0)$  be the best  $d$ -variate approximant of the unknown pmf  $f(\cdot) = p_d(\cdot; \theta)w(\cdot; \theta)$  as (15) under the Kullback–Leibler criterion, and let  $w(\cdot) := f(\cdot)/p_0(\cdot)$  be the corresponding  $d$ -variate weight function. As  $n \rightarrow \infty$  and under the assumption (a1) on  $f$ , then the estimator  $\widehat{f}_n(\cdot) = p_d(\cdot; \widehat{\theta}_n)\widetilde{w}_n(\cdot; \widehat{\theta}_n)$  in (16) of  $f$  and reformulated in (19) satisfies

$$\begin{aligned} \text{Bias}\{\widehat{f}_n(\mathbf{x})\} &= p_0(\mathbf{x})[w(\mathbf{x}) - f(\mathbf{x})\{p_0(\mathbf{x})\}^{-1} + \langle \nabla w(\mathbf{x}), \mathbf{A}(\mathbf{x}, \mathbf{H}_n) \rangle] \\ &\quad + \frac{1}{2} p_0(\mathbf{x}) \left( \text{tr}\{\mathcal{H}w(\mathbf{x})[\mathbf{B}(\mathbf{x}, \mathbf{H}_n) + \mathbf{A}(\mathbf{x}, \mathbf{H}_n)^\top \mathbf{A}(\mathbf{x}, \mathbf{H}_n)]\} \right) \\ &\quad + (1 + o\{\text{tr}[\mathbf{B}(\mathbf{x}, \mathbf{H}_n)]\} + n^{-2}), \end{aligned}$$

for any  $\mathbf{x} \in \mathbb{T}_d^+$ . Furthermore, if (a2) holds then we have  $\text{var}\{\widehat{f}_n(\mathbf{x})\} = \text{var}\{\widetilde{f}_n(\mathbf{x})\}$  of (11).

Once again, the bias is different from that of (10). Thus, the proposed semiparametric estimator  $\widehat{f}_n$  in (16) of  $f$  can be shown to be better (or not) than the traditional nonparametric one  $\widetilde{f}_n$  in (8). The following subsection provides a practical solution.

### 4.3. Model Diagnostics

The estimated weight function  $\widehat{w}_n(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n)$  given in (17) provides useful information for model diagnostics. The  $d$ -variate weight function  $w(\cdot)$  is equal one if the  $d$ -variate parametric start model  $p_d(\cdot; \boldsymbol{\theta}_0)$  is indeed the true pmf. Hjort and Glad [19] proposed to check this adequacy by examining a plot of the weight function for various potential models with pointwise confidence bands to see whether or not  $w(\mathbf{x}) = 1$  is reasonable. See also [20,21] for univariate count setups.

In fact, without technical details here we use the model diagnostics for verifying the adequacy of the model by examining a plot of  $\mathbf{x} \mapsto \widehat{w}_n(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n)$  or

$$\widehat{W}_n(\mathbf{x}) := \log \widehat{w}_n(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n) = \log[\widehat{f}_n(\mathbf{x})/p_d(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n)] \tag{20}$$

for all  $\mathbf{x} = \mathbf{X}_i, i = 1, \dots, n$ , with a pointwise confidence band of  $\pm 1.96$  for large  $n$ ; that is to see how far away it is from zero. More precisely, for instance,  $\widehat{W}_n$  is  $< 5\%$  for pure nonparametric, it belongs to  $[5\%, 95\%]$  for semiparametric, and it is  $> 95\%$  for full parametric models. It is noteworthy that the retention of pure nonparametric means the inconvenience of parametric part considered in this approach; hence, the orthant dataset is left free.

## 5. Semicontinuous Examples of Application with Discussions

For a practical implementation of our approach, we propose to use the popular multiple gamma kernels as (9) by selecting the adaptive Bayesian procedure of [13] to smooth  $\widehat{w}_n(\mathbf{x}; \widehat{\boldsymbol{\theta}}_n)$ . Hence, we shall gradually consider  $d$ -variate semicontinuous cases with  $d = 1, 2, 3$  for real datasets. All computations and graphics have been done with the R software [56].

### 5.1. Adaptive Bayesian Bandwidth Selection for Multiple Gamma Kernels

From Table 1, the function  $G_{x,h}(\cdot)$  is the gamma kernel [43] given on the support  $\mathbb{S}_{x,h} = [0, \infty) = \mathbb{T}_1^+$  with  $x \geq 0$  and  $h > 0$ :

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

where  $\mathbb{1}_E$  denotes the indicator function of any given event  $E$ . This gamma kernel  $G_{x,h}(\cdot)$  appears to be the pdf of the gamma distribution, denoted by  $\mathcal{G}(1 + x/h, h)$  with shape parameter  $1 + x/h$  and scale parameter  $h$ . The multiple gamma kernel from (9) is written as  $\mathbf{K}_{\mathbf{x},\mathbf{H}}(\cdot) = \prod_{j=1}^d G_{x_j, h_j}(\cdot)$  with  $\mathbf{H} = \text{diag}_d(h_j)$ .

For applying (13) and (14) in the framework of semiparametric estimator  $\widehat{f}_n$  in (16), we assume that each component  $h_{i\ell} = h_{i\ell}(n), \ell = 1, \dots, d$ , of  $\mathbf{H}_i$  has the univariate inverse gamma prior  $Ig(\alpha, \beta_\ell)$  distribution with same shape parameters  $\alpha > 0$  and, eventually, different scale parameters  $\beta_\ell > 0$  such that  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top$ . We here recall that the pdf of  $Ig(\alpha, \beta_\ell)$  with  $\alpha, \beta_\ell > 0$  is defined by

$$Ig_{\alpha, \beta_\ell}(u) = \frac{\beta_\ell^\alpha}{\Gamma(\alpha)} u^{-\alpha-1} \exp(-\beta_\ell/u) \mathbb{1}_{(0,\infty)}(u), \ell = 1, \dots, d. \tag{21}$$

The mean and the variance of the prior distribution (21) for each component  $h_{i\ell}$  of the vector  $\mathbf{H}_i$  are given by  $\beta_\ell/(\alpha - 1)$  for  $\alpha > 1$ , and  $\beta_\ell^2/(\alpha - 1)^2(\alpha - 2)$  for  $\alpha > 2$ , respectively. Note that for a fixed  $\beta_\ell > 0, \ell = 1, \dots, d$ , and if  $\alpha \rightarrow \infty$ , then the distribution of the bandwidth vector  $\mathbf{H}_i$  is concentrated around the null vector  $\mathbf{0}$ .

From those considerations, the closed form of the posterior density and the Bayesian estimator of  $\mathbf{H}_i$  are given in the following proposition which is proven in Appendix B.

**Proposition 5.** For fixed  $i \in \{1, 2, \dots, n\}$ , consider each observation  $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^\top$  with its corresponding  $\mathbf{H}_i = \text{diag}_d(h_{ij})$  of univariate bandwidths and defining the subset

$\mathbb{I}_i = \{k \in \{1, \dots, d\}; X_{ik} = 0\}$  and its complementary set  $\mathbb{I}_i^c = \{\ell \in \{1, \dots, d\}; X_{i\ell} \in (0, \infty)\}$ . Using the inverse gamma prior  $Ig_{\alpha, \beta_\ell}$  of (21) for each component  $h_{i\ell}$  of  $\mathbf{H}_i$  in the multiple gamma estimator with  $\alpha > 1/2$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top \in (0, \infty)^d$ , then:

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

$$\pi(\mathbf{H}_i | \mathbf{X}_i) = \frac{p_d(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_n)}{D_i(\alpha, \boldsymbol{\beta})} \sum_{j=1, j \neq i}^n \frac{1}{p_d(\mathbf{X}_j; \widehat{\boldsymbol{\theta}}_n)} \left( \prod_{k \in \mathbb{I}_i} C_{jk}(\alpha, \beta_k) Ig_{\alpha+1, X_{jk} + \beta_k}(h_{ik}) \right) \left( \prod_{\ell \in \mathbb{I}_i^c} A_{ij\ell}(\alpha, \beta_\ell) Ig_{\alpha+1/2, B_{ij\ell}(\beta_\ell)}(h_{i\ell}) \right),$$

with  $A_{ij\ell}(\alpha, \beta_\ell) = [\Gamma(\alpha + 1/2)] / (\beta_\ell^{-\alpha} X_{i\ell}^{1/2} \sqrt{2\pi} [B_{ij\ell}(\beta_\ell)]^{\alpha+1/2})$ ,  $B_{ij\ell}(\beta_\ell) = X_{i\ell} \log(X_{i\ell} / X_{j\ell}) + X_{j\ell} - X_{i\ell} + \beta_\ell$ ,  $C_{jk}(\alpha, \beta_k) = [\Gamma(\alpha + 1)] / [\beta_k^{-\alpha} (X_{jk} + \beta_k)^{\alpha+1}]$ , and  $D_i(\alpha, \boldsymbol{\beta}) = p_d(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_n) \sum_{j=1, j \neq i}^n \left( \prod_{k \in \mathbb{I}_i} A_{ijk}(\alpha, \beta_k) \right) \left( \prod_{\ell \in \mathbb{I}_i^c} B_{ij\ell}(\beta_\ell) \right) / p_d(\mathbf{X}_j; \widehat{\boldsymbol{\theta}}_n)$ ;

(ii) under the quadratic loss function, the Bayesian estimator  $\widehat{\mathbf{H}}_i = \text{diag}_d(\widehat{h}_{im})$  of  $\mathbf{H}_i$  in (16) is

$$\widehat{h}_{im} = \frac{p_d(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_n)}{D_i(\alpha, \boldsymbol{\beta})} \sum_{j=1, j \neq i}^n \frac{1}{p_d(\mathbf{X}_j; \widehat{\boldsymbol{\theta}}_n)} \left( \prod_{k \in \mathbb{I}_i} C_{jk}(\alpha, \beta_k) \right) \left( \prod_{\ell \in \mathbb{I}_i^c} A_{ij\ell}(\alpha, \beta_\ell) \right) \left( \frac{X_{jm} + \beta_m}{\alpha} \mathbb{1}_{(0)}(X_{im}) + \frac{B_{ijm}(\beta_m)}{\alpha - 1/2} \mathbb{1}_{(0, \infty)}(X_{im}) \right),$$

for  $m = 1, 2, \dots, d$ , with the previous notations of  $A_{ij\ell}(\alpha, \beta_\ell)$ ,  $B_{ijm}(\beta_m)$ ,  $C_{jk}(\alpha, \beta_k)$  et  $D_i(\alpha, \boldsymbol{\beta})$ .

Following Somé and Kokonendji [13] for nonparametric approach, we have to select the prior parameters  $\alpha$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top$  of the multiple inverse gamma of  $Ig(\alpha, \beta_\ell)$  in (21) as follows:  $\alpha = \alpha_n = n^{2/5} > 2$  and  $\beta_\ell > 0$ ,  $\ell = 1, \dots, d$ , to obtain the convergence of the variable bandwidths to zero with a rate close to that of an optimal bandwidth. For practical use, we here consider each  $\beta_\ell = 1$ .

### 5.2. Semicontinuous Datasets

The numerical illustrations shall be done through the following dataset of Table 2 which are recently used in [13] for non-parametric approach and only in the trivariate setup as semicontinuous data. It concerns three measurements (with  $n = 42$ ) of drinking water pumps installed in the Sahel. The first variable  $X_1$  represents the failure times (in months) and, also, it is recently used by Touré et al. [24]. The second variable  $X_2$  refers to the distance (in kilometers) between each water pump and the repair center in the Sahel, while the third one  $X_3$  stands for average volume (in  $m^3$ ) of water per day.

**Table 2.** Drinking water pumps trivariate data measured in the Sahel with  $n = 42$ .

$X_1$ :	23	261	87	10	120	14	62	15	47	225	71	20	246	21
$X_2$ :	97	93	94	100	98	84	96	110	121	73	90	93	103	116
$X_3$ :	26	52	22	39	23	26	32	17	10	39	31	42	52	26
$X_1$ :	19	42	20	5	12	120	17	11	3	14	71	11	5	14
$X_2$ :	114	82	96	94	77	91	117	103	99	113	79	109	84	118
$X_3$ :	26	36	43	36	6	27	15	36	9	52	11	20	25	37
$X_1$ :	11	16	90	1	16	52	95	10	1	14	4	7	14	20
$X_2$ :	98	93	94	103	109	110	89	108	101	93	102	138	103	96
$X_3$ :	25	18	43	43	24	38	6	40	21	34	15	23	68	37

Table 3 displays all empirical univariate, bivariate and trivariate variation (6) and dispersion (3) indexes from Table 2. Hence, each  $X_j$ ,  $(X_j, X_k)$  and  $(X_1, X_2, X_3)$  is over-dispersed compared to the corresponding uncorrelated Poisson distribution. However, only  $(X_1, X_3)$  (resp.  $X_1$ ) can be considered as a bivariate equi-variation (resp. univariate over-variation) with respect to the corresponding uncorrelated exponential distribution; and, other  $X_j$ ,  $(X_j, X_k)$  and  $(X_1, X_2, X_3)$  are under-varied. In fact, we just compute dispersion indexes for curiosity since all values in Table 2 are positive integers; and, we here now omit the counting point of view in the remainder of the analysis.

**Table 3.** Empirical univariate (in diagonal), bivariate (off diagonal) and trivariate (at the corner) variation (6) and dispersion (3) indexes.

$\widehat{GVI}_3 = 0.0533$	$X_1$	$X_2$	$X_3$	$\widehat{GDI}_3 = 15.1229$	$X_1$	$X_2$	$X_3$
$X_1$	1.9425	0.0557	1.0549	$X_1$	89.5860	14.3223	70.7096
$X_2$	0.0557	0.0167	0.0157	$X_2$	14.3223	1.6623	2.0884
$X_3$	1.0549	0.0157	0.2122	$X_3$	70.7096	2.0884	6.3192

Thus, we are gradually investing in semiparametric approaches for three univariates, three bivariate and only one trivariate from  $(X_1, X_2, X_3)$  of Table 2.

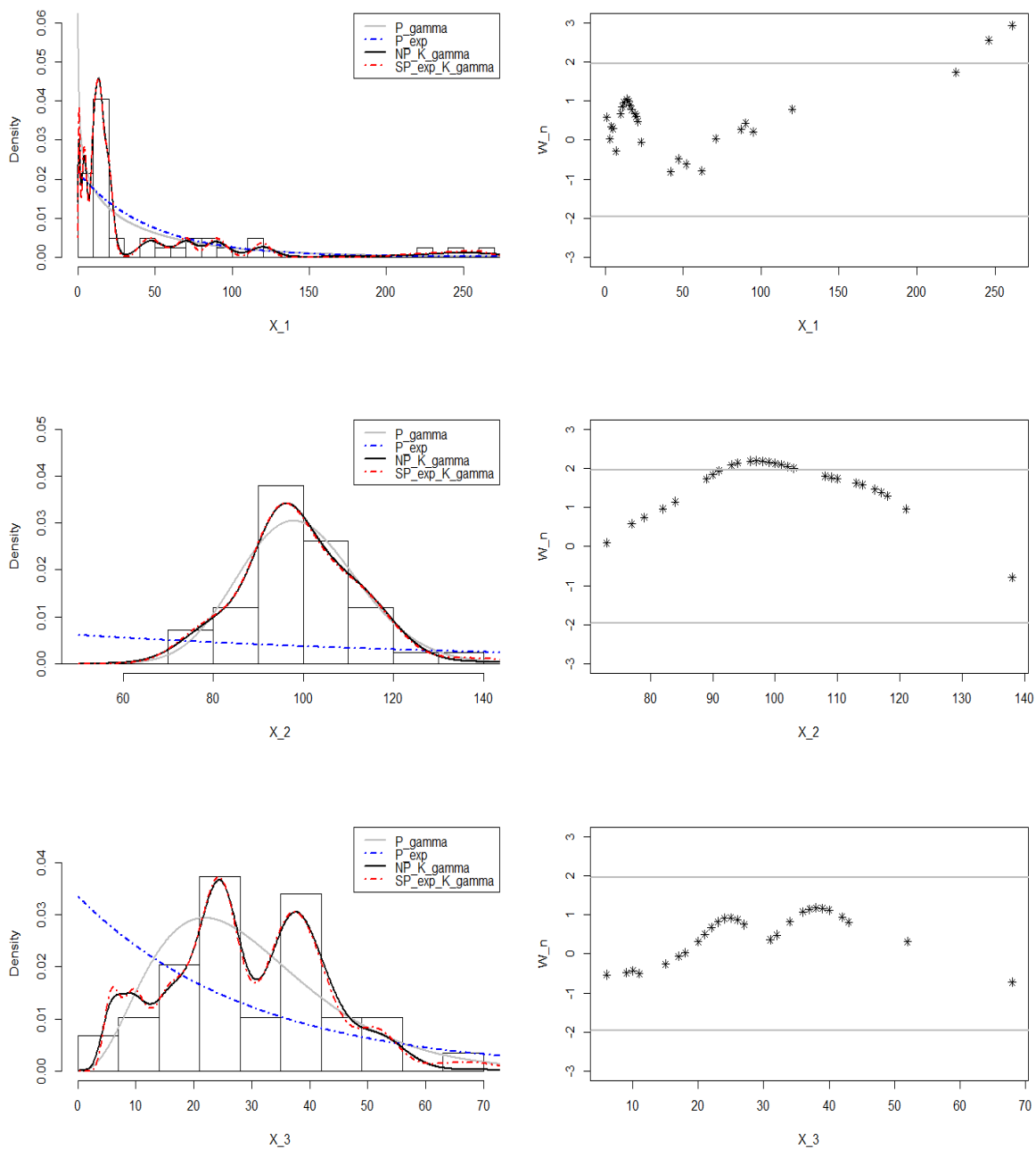
### 5.3. Univariate Examples

For each univariate semicontinuous dataset  $X_j, j = 1, 2, 3$ , we have already computed the GVI in Table 3 which belongs in  $(0.01, 1.95) \ni 1$ . This allows to consider our flexible semiparametric estimation  $\widehat{f}_{n,j}$  with an exponential  $\mathcal{E}_1(\mu_j)$  as start in (16) and using adaptive Bayesian bandwidth in gamma kernel of Proposition 5. Hence, we deduce the corresponding diagnostic percent  $\widetilde{W}_{n,j}$  from (20) for deciding an appropriate approach. In addition, we first present the univariate nonparametric estimation  $\widetilde{f}_{n,j}$  with adaptive Bayesian bandwidth in gamma kernel of [35] and then propose another parametric estimation of  $X_j$  by the standard gamma model with shape  $(a_j)$  and scale  $(b_j)$  parameters.

Table 4 transcribes parameter maximum likelihood estimates of exponential and gamma models with diagnostic percent from Table 2. Figure 3 exhibits histogram,  $\widetilde{f}_{n,j}$ ,  $\widehat{f}_{n,j}$ , exponential, gamma and diagnostic  $\widetilde{W}_{n,j}$  graphs for each univariate data  $X_j$ . One can observe differences with the naked eye between  $\widetilde{f}_{n,j}$  and  $\widehat{f}_{n,j}$  although they are very near and with the same pace. The diagnostic  $\widetilde{W}_{n,j}$  graphics lead to conclude to semiparametric approach for  $X_2$  and to full parametric models for  $X_3$  and slightly also for  $X_1$ . Thus, we have suggested the gamma model with two parameters for improving the starting exponential model; see, e.g., Table 2 in [54], for alternative parametric models.

**Table 4.** Parameter estimates of models and diagnostic percents of univariate datasets.

Estimate	$\widehat{\mu}_j$	$\widetilde{W}_{n,j}$ (%)	$\widehat{a}_j$	$\widehat{b}_j$
$X_1$	0.0217	95.2381	0.7256	63.5618
$X_2$	0.0100	76.1905	56.9817	1.7470
$X_3$	0.0336	100.0000	3.7512	7.9403



**Figure 3.** Comparative graphs of estimates of  $X_1$ ,  $X_2$  and  $X_3$  with their corresponding diagnostics.

5.4. Bivariate and Trivariate Examples

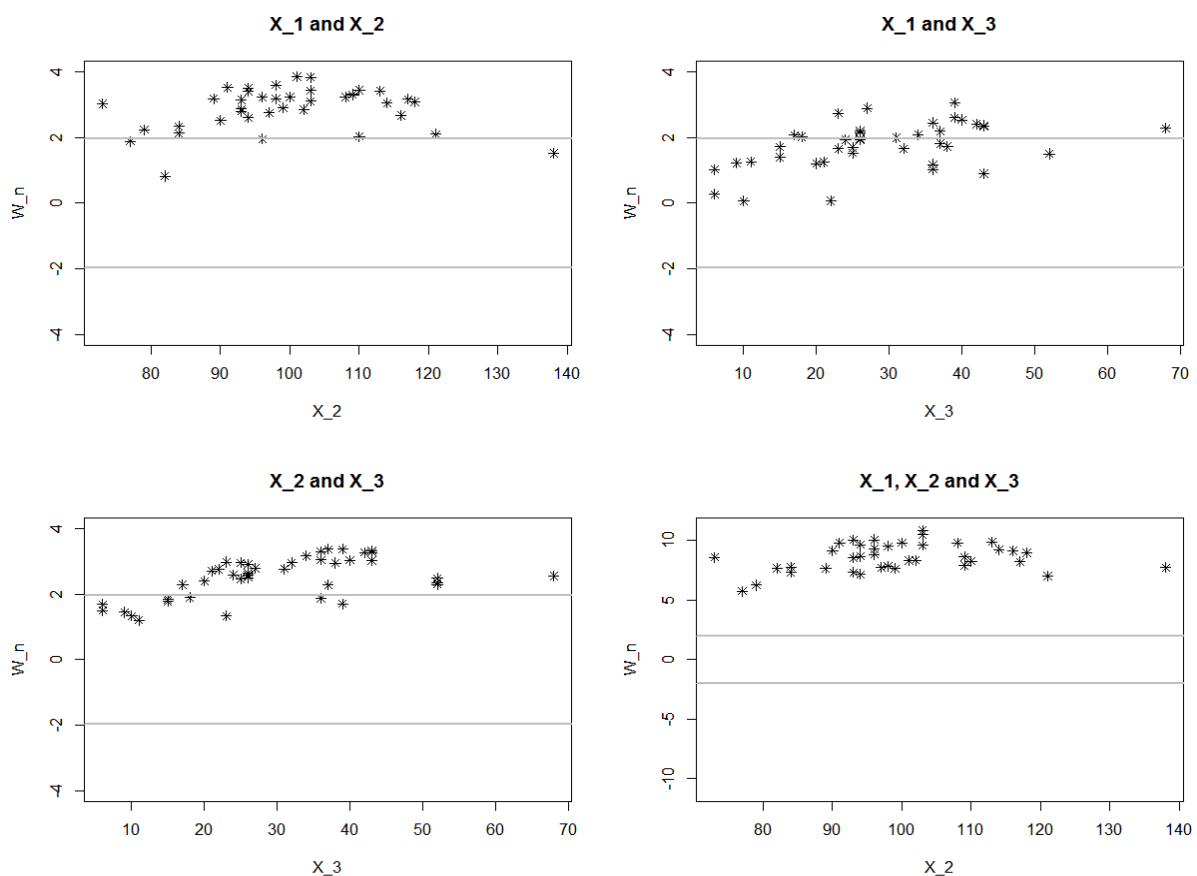
For the sake of flexibility and efficiency, we here analyze our proposed semiparametric estimation  $\widehat{f}_n$  with an uncorrelated exponential as a start in (16) and using adaptive Bayesian bandwidth in gamma kernel of Proposition 5. This concerns all bivariate and trivariate datasets from Table 2 for which their GVI are in  $(0.01, 1.06) \ni 1$  from Table 3. All the computation times are almost instantaneous.

Table 5 reports the main numerical results of the corresponding correlations, MVI, parameter estimates and finally diagnostic percent  $\widetilde{W}_n$  from (20) that we intentionally omit to represent some graphics in three or four dimensions. However, Figure 4 displays some representative projections of  $\widetilde{W}_n$ . From Table 5, the cross empirical correlations are closed to 0 and all MVI are smaller than 1 which allows us to consider uncorrelated

exponential start-models. The maximum likelihood method is also used for estimating the parameters  $\mu_j$  for getting the same results as in Table 4. Thus, the obtained numerical values of  $\widetilde{W}_n$  indicate semiparametric approaches for all bivariate datasets and the purely nonparametric method for the trivariate one; see [13] for more details on this nonparametric analysis. This progressive semiparametric analysis of the trivariate dataset of Table 2 shows the necessity of a suitable choice of the parametric start-models, which may take into account the correlation structure. Hence, the retention of pure nonparametric means the inconvenience of parametric part used in the modelization. Note that we could consider the Marshall–Olkin exponential distributions with nonnegative correlations as start-models; but, they are singular. See Appendix A for a brief review.

**Table 5.** Correlations, MVI, parameter estimates and diagnostic percents of bi- and trivariate cases.

Dataset	$(X_1, X_2)$	$(X_1, X_3)$	$(X_2, X_3)$	$(X_1, X_2, X_3)$
$\widehat{\rho}(X_j, X_k)$	-0.3090	0.2597	0.0245	$\det \widehat{\rho} = 0.8325$
$\widehat{MVI}$	0.0720	0.9857	0.0155	0.0634
$(\widehat{\mu}_j)$	(0.0217, 0.0100)	(0.0217, 0.0336)	(0.0100, 0.0336)	(0.0217, 0.0100, 0.0336)
$W_n$ (%)	9.5238	52.3809	26.1005	0.0000



**Figure 4.** Univariate projections of diagnostic graphs for bivariate and trivariate models.

### 6. Concluding Remarks

In this paper, we have presented a flexible semiparametric approach for multivariate nonnegative orthant distributions. We have first recalled multivariate variability indexes GVI, MVI, RVI, GDI, MDI and RDI from RWI as a prelude to the second-order discrimination for these parametric distributions. We have then reviewed and provided new proposals to



the nonparametric estimators through multivariate associated kernels; e.g., Proposition 1 and Corollary 1. Both effective adaptive and local Bayesian selectors of bandwidth matrices are suggested for semicontinuous and counting data, respectively.

All these previous ingredients were finally used to develop the semiparametric modeling for multivariate nonnegative orthant distributions. Numerical illustrations have been simply done for univariate and multivariate semicontinuous datasets with the uncorrelated exponential start-models after examining GVI and MVI. The adaptive Bayesian bandwidth selection (13) in multiple gamma kernel (Proposition 5) were here required for applications. Finally, the diagnostic models have played a very interesting role in helping to the appropriate approach, even if it means improving it later.

At the meantime, Kokonendji et al. [37] proposed an in-depth practical analysis of multivariate count datasets starting with multivariate (un)correlated Poisson models after reviewing GDI and RDI. They have also established an equivalent of our Proposition 5 for the local Bayesian bandwidth selection (12) by using the multiple binomial kernels from Example 3. As one of the many perspectives, one could consider the categorical setup with local Bayesian version of the multivariate associated kernel of Aitchison and Aitken [38] from Example 1 of the univariate case.

At this stage of analysis, all the main foundations are now available for working in a multivariate setup such as variability indexes, associated kernels, Bayesian selectors and model diagnostics. We just have to adapt them to each situation encountered. For instance, we have the semiparametric regression modeling; see, e.g., Abdous et al. [57] devoted to counting explanatory variables and [22]. An opportunity will be opened for hazard rate functions (e.g., [51]). The near future of other functional groups, such as categorical and mixed, can now be considered with objectivity and feasibility.

**Author Contributions:** Conceptualization (C.C.K., S.M.S); Formal analysis (C.C.K., S.M.S); Investigation (C.C.K., S.M.S); Methodology (C.C.K., S.M.S); Software (C.C.K., S.M.S); Supervision (C.C.K.); Writing—original draft preparation (C.C.K., S.M.S); Writing—review and editing (C.C.K., S.M.S). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** The authors are grateful to the Associate Editor and two anonymous referees for their constructive comments. We also sincerely thank Mohamed Elmi Assoweh for some interesting discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

GDI	Generalized dispersion index
GVI	Generalized variation index
iid	Independent and identically distributed
MDI	Marginal dispersion index
MVI	Marginal variation index
pdf	Probability density function
pdmf	Probability density or mass function
pmf	Probability mass function
RDI	Relative dispersion index
RVI	Relative variation index
RWI	Relative variability index

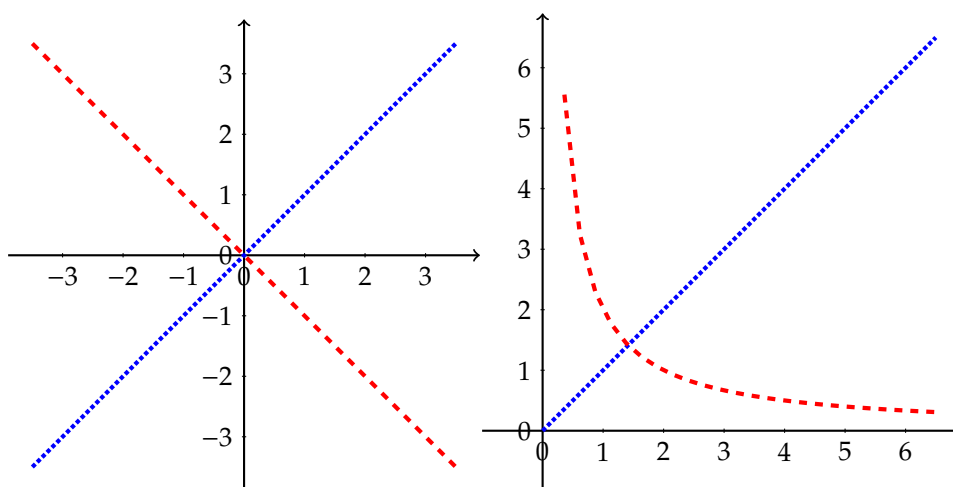
### Appendix A. On a Broader $d$ -Variate Parametric Models and the Marshall-Olkin Exponential

According to Cuenin et al. [58], taking  $p \in \{1, 2\}$  in their multivariate Tweedie models of flexible dependence structure, another way to define the  $d$ -variate Poisson and exponential distributions is given by  $\mathcal{P}_d(\Lambda)$  and  $\mathcal{E}_d(\Lambda)$ , respectively. The  $d \times d$  symmetric variation matrix  $\Lambda = (\lambda_{ij})_{i,j \in \{1, \dots, d\}}$  is such that  $\lambda_{ij} = \lambda_{ji} \geq 0$ , the mean of the corresponding marginal distribution is  $\lambda_{ii} > 0$ , and the non-negative correlation terms satisfy

$$\rho_{ij} = \frac{\lambda_{ij}}{\sqrt{\lambda_{ii}\lambda_{jj}}} \in [0, \min\{R(i, j), R(j, i)\}], \tag{A1}$$

with  $R(i, j) = \sqrt{\lambda_{ii}/\lambda_{jj}}(1 - \lambda_{ii}^{-1} \sum_{\ell \neq i, j} \lambda_{i\ell}) \in (0, 1)$ . Their constructions are perfectly defined having  $d(d + 1)/2$  parameters as in  $\mathcal{P}_d(\mu, \rho)$  and  $\mathcal{E}_d(\mu, \rho)$ . Moreover, we attain the exact bounds of the correlation terms in (A1). Cuenin et al. [58] have pointed out the construction and simulation of the negative correlation structure from the positive one of (A1) by considering the inversion method.

The negativity of a correlation component is crucial for the phenomenon of under-variability in a bivariate/multivariate positive orthant model. Figure A1 (right) plots a limit shape of any bivariate positive orthant distribution with very strong negative correlation (in red), which is not the diagonal line of the upper bound (+1) of positive correlation (in blue); see, e.g., [58] for details on both bivariate orthant (i.e., continuous and count) models. Conversely, Figure A1 (left) represents the classic lower (-1) and upper (+1) bounds of correlations on  $\mathbb{R}^2$  or finite support.



**Figure A1.** Support of bivariate distributions with maximum correlations (positive in blue and negative in red): model on  $\mathbb{R}^2$  (left) and also finite support; model on  $\mathbb{T}_2^+ \subseteq [0, \infty)^2$  (right), without finite support.

The  $d$ -variate exponential  $\mathbf{X} = (X_1, \dots, X_d)^\top \sim \mathcal{E}_d(\mu, \mu_0)$  of Marshall and Olkin [59] is built as follows. Let  $Y_1, \dots, Y_d$  and  $Z$  be univariate exponential random variables with parameters  $\mu_1 > 0, \dots, \mu_d > 0$  and  $\mu_0 \geq 0$ , respectively. Then, by setting  $X_j := Y_j + Z$  for  $j = 1, \dots, d$ , one has  $\mathbb{E}X_j = 1/(\mu_j + \mu_0) = \sqrt{\text{var}X_j}$  and  $\text{cov}(X_j, X_\ell) = \mu_0 / \{(\mu_j + \mu_0)(\mu_\ell + \mu_0)(\mu_j + \mu_\ell + \mu_0)\}$  for all  $j \neq \ell$ . Each correlation  $\rho(X_j, X_\ell) = \mu_0 / (\mu_j + \mu_\ell + \mu_0)$  lies in  $[0, 1]$  if and only if  $\mu_0 \geq 0$ . From its survival (or reliability) function

$$S(\mathbf{x}; \mu, \mu_0) = \exp\left(-\mu_0 \max(x_1, \dots, x_d) - \sum_{j=1}^d \mu_j x_j\right),$$

its pdf can be written as

$$p_d(\mathbf{x}; \boldsymbol{\mu}, \mu_0) = \begin{cases} S(\mathbf{x}; \boldsymbol{\mu}, \mu_0) (\mu_0 + \mu_\ell) \prod_{j=1, j \neq \ell}^d \mu_j & \text{if } x_\ell := \max(x_1, \dots, x_d) \text{ and } x_\ell \neq x_j, j \neq \ell \\ S(\mathbf{x}; \boldsymbol{\mu}, \mu_0) \mu_0 \mu_{j_1} \cdots \mu_{j_k} & \text{if } x_{j_1}, \dots, x_{j_k} < x_{\ell_{k+1}} = \dots = x_{\ell_d} \\ S(\mathbf{x}; \boldsymbol{\mu}, \mu_0) \mu_0 & \text{if } x_1 = \dots = x_d > 0. \end{cases}$$

It is not absolutely continuous with respect to the Lebesgue measure in  $\mathbb{T}_d^+$  and has singularities corresponding to the cases where two or more of the  $x_j$ 's are equal. Karlis [60] has proposed a maximum likelihood estimation of parameters via an EM algorithm. Finally, Kokonendji et al. [16] have calculated

$$\text{GVI}(\mathbf{X}) = 1 + \frac{\mu_0 \sum_{j=1}^d (\mu_j + \mu_0)^{-1} \{ \sum_{\ell \neq j} (\mu_j + \mu_\ell + \mu_0)^{-1} (\mu_\ell + \mu_0)^{-1} \}}{\{ (\mu_1 + \mu_0)^{-2} + \dots + (\mu_d + \mu_0)^{-2} \}^2} \geq 1 \quad (\Leftrightarrow \mu_0 \geq 0).$$

and

$$\text{MVI}(\mathbf{X}) = \frac{\sum_{j=1}^d (\mu_j + \mu_0)^{-4}}{\sum_{j=1}^d (\mu_j + \mu_0)^{-4} + 2 \sum_{1 \leq j < \ell \leq d} (\mu_j + \mu_0)^{-2} (\mu_\ell + \mu_0)^{-2}} < 1.$$

Hence, the Marshall–Olkin exponential model  $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \mu_0)$  is always under-varied with respect to the MVI and over- or equi-varied with respect to GVI. If  $\mu_0 = 0$  then  $\mathcal{E}_d(\boldsymbol{\mu}, \mu_0)$  is reduced to the uncorrelated  $\mathcal{E}_d(\boldsymbol{\mu})$  with  $\text{GVI}(\mathbf{Y}) = 1$ . However, the assumption of non-negative correlations between components is sometimes unrealistic for some analyzes.

**Appendix B. Proofs of Proposition 2, Proposition 3 and Proposition 5**

**Proof of Proposition 2.** From Definition 1, we get (see also [10] for more details)

$$\begin{aligned} \mathbb{E}[\widetilde{f}_n(\mathbf{x})] - f(\mathbf{x}) &= \mathbb{E}[\mathbf{K}_{\mathbf{x}, \mathbf{H}_n}(\mathbf{X}_j)] - f(\mathbf{x}) = \int_{\mathbb{S}_{\mathbf{x}, \mathbf{H}_n} \cap \mathbb{T}_d^+} \mathbf{K}_{\mathbf{x}, \mathbf{H}_n}(\mathbf{u}) f(\mathbf{u}) \nu(d\mathbf{u}) - f(\mathbf{x}) \\ &= \mathbb{E}[f(\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n})] - f(\mathbf{x}). \end{aligned} \tag{A2}$$

Next, using (A2), by a Taylor expansion of the function  $f(\cdot)$  over the points  $\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}$  and  $\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]$ , we get

$$\begin{aligned} f(\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}) &= f(\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) + \langle \nabla f(\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]), (\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) \rangle \\ &\quad + \frac{1}{2} \langle \mathcal{H}f(\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) (\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]), (\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) \rangle \\ &\quad + \|\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]\|^2 o(1) \\ &= f(\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) + \langle \nabla f(\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]), (\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) \rangle \\ &\quad + \frac{1}{2} \text{tr}[\mathcal{H}f(\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) (\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) (\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}])^\top] \\ &\quad + \text{tr}[(\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) (\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n} - \mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}])^\top] o(1), \end{aligned} \tag{A3}$$

where  $o(1)$  is uniform in a neighborhood of  $\mathbf{x}$ . Therefore, taking the expectation in both sides of (A3) and then substituting the result in (A2), we get

$$\begin{aligned} \mathbb{E}[\widetilde{f}_n(\mathbf{x})] - f(\mathbf{x}) &= f(\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) - f(\mathbf{x}) + \frac{1}{2} \text{tr}[\mathcal{H}f(\mathbb{E}[\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n}]) \text{var}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n})] + o\{\text{tr}[\text{var}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}_n})]\} \\ &= f(\mathbf{x} + \mathbf{A}) - f(\mathbf{x}) + \frac{1}{2} \text{tr}[\mathcal{H}f(\mathbf{x} + \mathbf{A}) \mathbf{B}(\mathbf{x}, \mathbf{H}_n)] + o\{\text{tr}[\mathbf{B}(\mathbf{x}, \mathbf{H}_n)]\}, \end{aligned}$$

where  $o\{\text{tr}[\mathbf{B}(\mathbf{x}, \mathbf{H}_n)]\}$  is uniform in a neighborhood of  $\mathbf{x}$ . The second Taylor expansion of the function  $f(\cdot)$  over the points  $\mathbf{x}$  and  $\mathbf{x} + \mathbf{A}(\mathbf{x}, \mathbf{H}_n)$  allows to conclude the bias (10).

About the variance term,  $f$  being bounded, we have  $\mathbb{E}[\mathbf{K}_{\mathbf{x},\mathbf{H}_n}(\mathbf{X}_j)] = O(1)$ . It follows that:

$$\begin{aligned} \text{var}[\widehat{f}_n(\mathbf{x})] &= \frac{1}{n} \text{var}[\mathbf{K}_{\mathbf{x},\mathbf{H}_n}(\mathbf{X}_j)] \\ &= \frac{1}{n} \left[ \int_{\mathbb{S}_{\mathbf{x},\mathbf{H}_n} \cap \mathbb{T}_d^+} \mathbf{K}_{\mathbf{x},\mathbf{H}_n}^2(\mathbf{u}) f(\mathbf{u}) \nu(d\mathbf{u}) + O(1) \right] \\ &= \frac{1}{n} \int_{\mathbb{S}_{\mathbf{x},\mathbf{H}_n} \cap \mathbb{T}_d^+} \mathbf{K}_{\mathbf{x},\mathbf{H}_n}^2(\mathbf{u}) \left( \begin{array}{c} f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{u} \rangle \\ + \frac{1}{2} (\mathbf{x} - \mathbf{u})^T \mathcal{H} f(\mathbf{x}) (\mathbf{x} - \mathbf{u}) \\ + o(\|\mathbf{x} - \mathbf{u}\|^2) \end{array} \right) \nu(d\mathbf{u}) \\ &= \frac{1}{n} f(\mathbf{x}) \|\mathbf{K}_{\mathbf{x},\mathbf{H}_n}\|_2^2 + o\left[\frac{1}{n(\det \mathbf{H}_n)^r}\right]. \end{aligned}$$

□

**Proof of Proposition 3.** Since one has  $\text{Bias}[\widehat{f}_n(\mathbf{x})] = p_d(\mathbf{x}; \theta_0) \mathbb{E}[\widetilde{w}_n(\mathbf{x})] - f(\mathbf{x})$  and  $\text{var}[\widehat{f}_n(\mathbf{x})] = [p_d(\mathbf{x}; \theta_0)]^2 \text{var}[\widetilde{w}_n(\mathbf{x})]$ , it is enough to calculate  $\mathbb{E}[\widetilde{w}_n(\mathbf{x})]$  and  $\text{var}[\widetilde{w}_n(\mathbf{x})]$  following Proposition 2 applied to  $\widetilde{w}_n(\mathbf{x}) = n^{-1} \sum_{i=1}^n \mathbf{K}_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i) / p_d(\mathbf{X}_i; \theta_0)$  for all  $\mathbf{x} \in \mathbb{T}_d^+$ .

Indeed, one successively has

$$\begin{aligned} \mathbb{E}[\widetilde{w}_n(\mathbf{x})] &= \mathbb{E}[\mathbf{K}_{\mathbf{x},\mathbf{H}_n}(\mathbf{X}_1) / p_d(\mathbf{X}_1; \theta_0)] = \int_{\mathbb{S}_{\mathbf{x},\mathbf{H}_n} \cap \mathbb{T}_d^+} \mathbf{K}_{\mathbf{x},\mathbf{H}_n}(\mathbf{u}) [p_d(\mathbf{u}; \theta_0)]^{-1} f(\mathbf{u}) \nu(d\mathbf{u}) = \mathbb{E}[w(\mathcal{Z}_{\mathbf{x},\mathbf{H}_n})] \\ &= w(\mathbf{x}) + \langle \nabla w(\mathbf{x}), \mathbf{A}(\mathbf{x}, \mathbf{H}_n) \rangle + \frac{1}{2} (\text{tr}\{\mathcal{H}w(\mathbf{x})[\mathbf{B}(\mathbf{x}, \mathbf{H}_n) + \mathbf{A}(\mathbf{x}, \mathbf{H}_n)^T \mathbf{A}(\mathbf{x}, \mathbf{H}_n)]\}) \\ &\quad + o(\text{tr}[\mathbf{B}(\mathbf{x}, \mathbf{H}_n)]), \end{aligned}$$

which leads to the announced result of  $\text{Bias}[\widehat{f}_n(\mathbf{x})]$ . As for  $\text{var}[\widetilde{w}_n(\mathbf{x})]$ , one also write

$$\begin{aligned} \text{var}[\widetilde{w}_n(\mathbf{x})] &= \frac{1}{n} \text{var}[\mathbf{K}_{\mathbf{x},\mathbf{H}_n}(\mathbf{X}_1) / p_d(\mathbf{X}_1; \theta_0)] \\ &= \frac{1}{n} \left[ \int_{\mathbb{S}_{\mathbf{x},\mathbf{H}_n} \cap \mathbb{T}_d^+} \mathbf{K}_{\mathbf{x},\mathbf{H}_n}^2(\mathbf{u}) [p_d(\mathbf{u}; \theta_0)]^{-2} f(\mathbf{u}) \nu(d\mathbf{u}) + O(1) \right] \\ &= \frac{1}{n} f(\mathbf{x}) [p_d(\mathbf{x}; \theta_0)]^{-2} \|\mathbf{K}_{\mathbf{x},\mathbf{H}_n}\|_2^2 + o\left[\frac{1}{n(\det \mathbf{H}_n)^r}\right] \end{aligned}$$

and the desired result of  $\text{var}[\widehat{f}_n(\mathbf{x})]$  is therefore deduced. □

**Proof of Proposition 5.** We have to adapt Theorem 2.1 of Somé and Kokonendji [13] to this semiparametric estimator  $\widehat{f}_n$  in (16). First, the leave-one-out associated kernel estimator (14) becomes

$$\widehat{f}_{n,\mathbf{H}_i,-i}(\mathbf{X}_i) := \frac{p_d(\mathbf{X}_i; \widehat{\theta}_n)}{n-1} \sum_{\ell=1, \ell \neq i}^n \frac{1}{p_d(\mathbf{X}_\ell; \widehat{\theta}_n)} \mathbf{K}_{\mathbf{X}_i, \mathbf{H}_i}(\mathbf{X}_\ell).$$

Then, the posterior distribution deduced from (13) is expressed as

$$\pi(\mathbf{H}_i | \mathbf{X}_i) := \pi(\mathbf{H}_i) \widehat{f}_{n,\mathbf{H}_i,-i}(\mathbf{X}_i) \left[ \int_{\mathcal{M}} \widehat{f}_{n,\mathbf{H}_i,-i}(\mathbf{X}_i) \pi(\mathbf{H}_i) d\mathbf{H}_i \right]^{-1}$$

and which leads to the result of Part (i) via Theorem 2.1 (i) in [13] for details. Consequently, we similarly deduce the adaptive Bayesian estimator  $\widehat{\mathbf{H}}_i = \text{diag}_d(\widehat{h}_{im})$  of Part (ii). □

## References

1. Somé, S.M.; Kokonendji, C.C.; Ibrahim, M. Associated kernel discriminant analysis for multivariate mixed data. *Electr. J. Appl. Statist. Anal.* **2016**, *9*, 385–399.
2. Libengué Dobélé-Kpoka, F.G.B. Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications. Ph.D. Thesis, Université de Franche-Comté, Besançon, France, 2013.
3. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Discrete Multivariate Distributions*; Wiley: New York, NY, USA, 1997.
4. Kotz, S.; Balakrishnan, N.; Johnson, N.L. *Continuous Multivariate Distributions—Models and Applications*, 2nd ed.; Wiley: New York, NY, USA, 2000.
5. Balakrishnan, N.; Lai, C.-D. *Continuous Bivariate Distributions*, 2nd ed.; Springer: New York, NY, USA, 2009.
6. Belaid, N.; Adjabi, S.; Kokonendji, C.C.; Zougab, N. Bayesian local bandwidth selector in multivariate associated kernel estimator for joint probability mass functions. *J. Statist. Comput. Simul.* **2016**, *86*, 3667–3681. [[CrossRef](#)]
7. Belaid, N.; Adjabi, S.; Kokonendji, C.C.; Zougab, N. Bayesian adaptive bandwidth selector for multivariate discrete kernel estimator. *Commun. Statist. Theory Meth.* **2018**, *47*, 2988–3001. [[CrossRef](#)]
8. Funke, B.; Kawka, R. Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods. *Comput. Statist. Data Anal.* **2015**, *92*, 148–162. [[CrossRef](#)]
9. Hirukawa, M. *Asymmetric Kernel Smoothing—Theory and Applications in Economics and Finance*; Springer: Singapore, 2018.
10. Kokonendji, C.C.; Somé, S.M. On multivariate associated kernels to estimate general density functions. *J. Korean Statist. Soc.* **2018**, *47*, 112–126. [[CrossRef](#)]
11. Ouimet, F. Density estimation using Dirichlet kernels. *arXiv* **2020**, arXiv:2002.06956v2.
12. Somé, S.M.; Kokonendji, C.C. Effects of associated kernels in nonparametric multiple regressions. *J. Statist. Theory Pract.* **2016**, *10*, 456–471. [[CrossRef](#)]
13. Somé, S.M.; Kokonendji, C.C. Bayesian selector of adaptive bandwidth for multivariate gamma kernel estimator on  $[0, \infty)^d$ . *J. Appl. Statist.* **2021**, forthcoming. [[CrossRef](#)]
14. Zougab, N.; Adjabi, S.; Kokonendji, C.C. Comparison study to bandwidth selection in binomial kernel estimation using Bayesian approaches. *J. Statist. Theory Pract.* **2016**, *10*, 133–153. [[CrossRef](#)]
15. Zougab, N.; Harfouche, L.; Ziane, Y.; Adjabi, S. Multivariate generalized Birnbaum—Saunders kernel density estimators. *Commun. Statist. Theory Meth.* **2018**, *47*, 4534–4555. [[CrossRef](#)]
16. Kokonendji, C.C.; Touré, A.Y.; Sawadogo, A. Relative variation indexes for multivariate continuous distributions on  $[0, \infty)^k$  and extensions. *AStA Adv. Statist. Anal.* **2020**, *104*, 285–307. [[CrossRef](#)]
17. Kokonendji, C.C.; Puig, P. Fisher dispersion index for multivariate count distributions: A review and a new proposal. *J. Multiv. Anal.* **2018**, *165*, 180–193. [[CrossRef](#)]
18. Weiß, C.H. On some measures of ordinal variation. *J. Appl. Statist.* **2019**, *46*, 2905–2926. [[CrossRef](#)]
19. Hjort, N.L.; Glad, I.K. Nonparametric density estimation with a parametric start. *Ann. Statist.* **1995**, *23*, 882–904. [[CrossRef](#)]
20. Kokonendji, C.C.; Senga Kiessé, T.; Balakrishnan, N. Semiparametric estimation for count data through weighted distributions. *J. Statist. Plann. Infer.* **2009**, *139*, 3625–3638. [[CrossRef](#)]
21. Kokonendji, C.C.; Zougab, N.; Senga Kiessé, T. Poisson-weighted estimation by discrete kernel with application to radiation biodosimetry. In *Biomedical Big Data & Statistics for Low Dose Radiation Research—Extended Abstracts Fall 2015*; Ainsbury, E.A., Calle, M.L., Cardis, E., Einbeck, J., Gómez, G., Puig, P., Eds.; Springer Birkhäuser: Basel, Switzerland, 2017; pp. 115–120.
22. Senga Kiessé, T.; Zougab, N.; Kokonendji, C.C. Bayesian estimation of bandwidth in semiparametric kernel estimation of unknown probability mass and regression functions of count data. *Comput. Statist.* **2016**, *31*, 189–206. [[CrossRef](#)]
23. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*, 6th ed.; Pearson Prentice Hall: New Jersey, NJ, USA, 2007.
24. Touré, A.Y.; Dossou-Gbété, S.; Kokonendji, C.C. Asymptotic normality of the test statistics for the unified relative dispersion and relative variation indexes. *J. Appl. Statist.* **2020**, *47*, 2479–2491. [[CrossRef](#)]
25. Touré, A.Y.; Kokonendji, C.C. *Count and Continuous Generalized Variability Indexes*; The R Package GWI, R Foundation for Statistical Computing: Vienna, Austria, 2021.
26. Arnold, B.C.; Manjunath, B.G. Statistical inference for distributions with one Poisson conditional. *arXiv* **2020**, arXiv:2009.01296.
27. Abid, R.; Kokonendji, C.C.; Masmoudi, A. Geometric Tweedie regression models for continuous and semicontinuous data with variation phenomenon. *AStA Adv. Statist. Anal.* **2020**, *104*, 33–58. [[CrossRef](#)]
28. Kokonendji, C.C.; Senga Kiessé, T. Discrete associated kernels method and extensions. *Statist. Methodol.* **2011**, *8*, 497–516. [[CrossRef](#)]
29. Scott, D.W. *Multivariate Density Estimation—Theory, Practice, and Visualization*; Wiley: New York, NY, USA, 1992.
30. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK, 1986.
31. Zougab, N.; Adjabi, S.; Kokonendji, C.C. Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. *Comput. Statist. Data Anal.* **2014**, *75*, 28–38. [[CrossRef](#)]
32. Kokonendji, C.C.; Libengué Dobélé-Kpoka, F.G.B. Asymptotic results for continuous associated kernel estimators of density functions. *Afr. Diaspora J. Math.* **2018**, *21*, 87–97.
33. Kokonendji, C.C.; Senga Kiessé, T.; Zocchi, S.S. Discrete triangular distributions and non-parametric estimation for probability mass function. *J. Nonparam. Statist.* **2007**, *19*, 241–254. [[CrossRef](#)]

34. Wansouwé, W.E.; Somé, S.M.; Kokonendji, C.C. Ake: R Package Discrete continuous associated kernel estimations. *R J.* **2016**, *8*, 259–276. [CrossRef]
35. Somé, S.M. Bayesian selector of adaptive bandwidth for gamma kernel density estimator on  $[0, \infty)$ . *Commun. Statist. Simul. Comput.* **2021**, forthcoming. [CrossRef]
36. Ziane, Y.; Zougab, N.; Adjabi, S. Adaptive Bayesian bandwidth selection in asymmetric kernel density estimation for nonnegative heavy-tailed data. *J. Appl. Statist.* **2015**, *42*, 1645–1658. [CrossRef]
37. Kokonendji, C.C.; Belaid, N.; Abid, R.; Adjabi, S. Flexible semiparametric kernel estimation of multivariate count distribution with Bayesian bandwidth selection. *Statist. Meth. Appl.* **2021**, forthcoming.
38. Aitchison, J.; Aitken, C.G.G. Multivariate binary discrimination by the kernel method. *Biometrika* **1976**, *63*, 413–420. [CrossRef]
39. Kokonendji, C.C.; Zocchi, S.S. Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions. *Statist. Probab. Lett.* **2010**, *80*, 1655–1662. [CrossRef]
40. Huang, A.; Sippel, L.; Fung, T. A consistent second-order discrete kernel smoother. *arXiv* **2020**, arXiv:2010.03302.
41. Libengué Dobélé-Kpoka, F.G.B.; Kokonendji, C.C. The mode-dispersion approach for constructing continuous associated kernels. *Afr. Statist.* **2017**, *12*, 1417–1446. [CrossRef]
42. Huang, A. On arbitrarily underdispersed Conway-Maxwell-Poisson distributions. *arXiv* **2020**, arXiv:2011.07503.
43. Chen, S.X. Probability density function estimation using gamma kernels. *Ann. Inst. Statist. Math.* **2000**, *52*, 471–480. [CrossRef]
44. Hirukawa, M.; Sakudo, M. Family of the generalised gamma kernels: A generator of asymmetric kernels for nonnegative data. *J. Nonparam. Statist.* **2015**, *27*, 41–63. [CrossRef]
45. Mousa, A.M.; Hassan, M.K.; Fathi, A. A new non parametric estimator for Pdf based on inverse gamma distribution. *Commun. Statist. Theory Meth.* **2016**, *45*, 7002–7010. [CrossRef]
46. Scaillet, O. Density estimation using inverse and reciprocal inverse Gaussian kernels. *J. Nonparam. Statist.* **2004**, *16*, 217–226. [CrossRef]
47. Igarashi, G.; Kakizawa, Y. Re-formulation of inverse Gaussian, reciprocal inverse Gaussian, and Birnbaum-Saunders kernel estimators. *Statist. Probab. Lett.* **2014**, *84*, 235–246. [CrossRef]
48. Jin, X.; Kawczak, J. Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Ann. Econ. Financ.* **2003**, *4*, 103–124.
49. Marchant, C.; Bertin, K.; Leiva, V.; Saulo, H. Generalized Birnbaum-Saunders kernel density estimators and an analysis of financial data. *Comput. Statist. Data Anal.* **2013**, *63*, 1–15. [CrossRef]
50. Mombeni, H.A.; Masouri, B.; Akhoond, M.R. Asymmetric kernels for boundary modification in distribution function estimation. *REVSTAT* **2019**, forthcoming.
51. Salha, R.B.; Ahmed, H.I.E.S.; Alhoubi, I.M. Hazard rate function estimation using Weibull kernel. *Open J. Statist.* **2014**, *4*, 650–661. [CrossRef]
52. Erçelik, E.; Nadar, M. A new kernel estimator based on scaled inverse chi-squared density function. *Am. J. Math. Manag. Sci.* **2020**, forthcoming. [CrossRef]
53. Lafaye de Micheaux, P.; Ouimet, F. A study of seven asymmetric kernels for the estimation of cumulative distribution functions. *arXiv* **2020**, arXiv:2011.14893.
54. Kokonendji, C.C.; Touré, A.Y.; Abid, R. On general exponential weight functions and variation phenomenon. *Sankhyā A* **2020**, forthcoming. [CrossRef]
55. White, H. Maximum likelihood estimation of misspecified models. *Econometrica* **1982**, *50*, 1–26. [CrossRef]
56. R Core Team. R: *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: <http://cran.r-project.org/> (accessed on 28 January 2021).
57. Abdous, B.; Kokonendji, C.C.; Senga Kiessé, T. Semiparametric regression for count explanatory variables. *J. Statist. Plann. Inference* **2012**, *142*, 1537–1548. [CrossRef]
58. Cuenin, J.; Jørgensen, B.; Kokonendji, C.C. Simulations of full multivariate Tweedie with flexible dependence structure. *Comput. Statist.* **2016**, *31*, 1477–1492. [CrossRef]
59. Marshall, A.W.; Olkin, I. A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **1967**, *62*, 30–44. [CrossRef]
60. Karlis, D. ML estimation for multivariate shock models via an EM algorithm. *Ann. Inst. Statist. Math.* **2003**, *55*, 817–830. [CrossRef]