École Doctorale Carnot-Pasteur

THÈSE DE DOCTORAT

Spécialité : **Mathématiques et Applications** (option : **Statistique**)

présentée par

Sobom Matthieu SOMÉ

Estimation non-paramétrique par noyaux associés multivariés et applications

---

dirigée par **Professeur Célestin C. KOKONENDJI**

Soutenue le **16 Novembre 2015**

Après avis des Rapporteurs :

Sophie Dabo-Niang, Professeur à l'Université de Lille 3, France
Laurent Gardes, Professeur à l'Université de Strasbourg, France

Devant le Jury composé du Directeur de thèse, des Rapporteurs et :

Hervé Cardot, Président, Professeur à l'Université de Bourgogne, France
Smail Adjabi, Examinateur, Professeur à l'Université de Bejaïa, Algérie
Tristan Senga Kiessé, Examinateur, Chargé de Recherche à l'INRA Rennes, France

Remerciements

Je tiens d'abord à remercier mon directeur de thèse, Monsieur Célestin C. Kokonendji, Professeur à l'Université Franche-Comté de Besançon, pour m'avoir donné l'opportunité de mener ce travail de recherche, pour sa confiance, ses encouragements et ses observations critiques, dans l'élaboration de ce travail. Je le remercie de m'avoir fait découvrir la méthode des noyaux associés, déjà à mon DEA, et de m'avoir permis de mener mes recherches en thèse, tout en m'accordant une liberté dans la réalisation de mes travaux.

Je tiens à remercier Madame Sophie Dabo-Niang, Professeur à l'Université Charles de Gaulle Lille 3 et Monsieur Laurent Gardes, Professeur à l'Université de Strasbourg, de l'honneur qu'ils me font d'avoir accepté d'être les rapporteurs de cette thèse et pour l'intérêt qu'ils ont porté à mon travail. Je remercie également Monsieur Hervé Cardot, Professeur à l'Université de Bourgogne, de présider ce jury de thèse. Un grand merci à Messieurs Smail Adjabi, Professeur à l'Université de Bejaïa en Algérie, et Tristan Senga Kiessé, Chargé de Recherche à l'INRA Rennes d'avoir accepté de siéger à ce jury.

Mes remerciements vont également à tous les membres du Laboratoire de Mathématiques de Besançon (LMB) qui m'ont accueilli chaleureusement et permis d'effectuer cette thèse dans une ambiance vraiment sereine. Un grand merci à tous mes collègues doctorants pour tous les moments passés ensemble.

Enfin, je tiens à remercier ma famille et mes amis pour m'avoir soutenu durant toutes ces années.

Cette thèse a été financée par l'Etat du Burkina Faso au travers du *Centre de Gestion des Etudiants Burkinabès*.

Productions scientifiques

I. Pour publication dans des revues à comité de lecture

[I.1] Kokonendji, C.C. & **Somé, S.M.** (2015). On multivariate associated kernels for smoothing some density function, `arXiv:1502.01173`. Submitted to *Journal of Nonparametric Statistics*, GNST-2014-08-14 (Under review).

[I.2] **Somé, S.M.** & Kokonendji, C.C. (2015). Effects of associated kernels in nonparametric multiple regressions, `arXiv:1502.01488`. Submitted to *Journal of Statistical Theory and Practice*, UJSP-2015-0077.R1 (In revision).

[I.3] **Somé, S.M.**, Kokonendji, C.C. & Ibrahim, M. (2015). Associated kernel discriminant analysis for mixed data. Submitted to *Journal of Statistical Computation and Simulation*, GSCS-2015-0430 (Under review).

[I.4] Wansouwé, W.E., **Somé, S.M.** & Kokonendji, C.C. (2015). `Ake:` an `R` package for discrete and continuous associated kernel estimations. Submitted to *Journal of Statistical Software*, JSS 1597 (Under review).

II. Autre publication

Wansouwé, W.E., **Somé, S.M.** & Kokonendji, C.C. (2015). `Ake` : Associated kernel estimations, URL `http://cran.r-project.org/web/packages/Ake/`.

III. Actes de congrès avec comité de lecture

[III.1] Kokonendji, C.C., **Somé, S.M.** & Ibrahim, M. (2015). Associated kernel discriminant analysis for multivariate mixed data. *Proceedings of the 60th World Statistics Congress of the International Statistical Institute (ISI)*, 5 pages, Rio de Janeiro 26-31 Juillet 2015, à paraître. URL http ://www.isi2015.org/.

[III.2] **Somé, S.M.** & Kokonendji, C.C. (2015). Analyse discriminante par noyaux associés pour données mixtes. *Actes des 47èmes Journées de Statistique de la Société Française de Statistiques (SFdS)*, 6 pages, Lille 01-05 Juin 2015.
URL http ://papersjds15.sfds.asso.fr/submission54.pdf.

[III.3] **Somé, S.M.**, Senga Kiéssé, T. & Kokonendji, C.C. (2014). Régression non-paramétrique multiple par noyaux associés. *Actes des 46èmes Journées de Statistique de la Société Française de Statistiques (SFdS)*, 6 pages, Rennes 02-06 Juin 2014.
URL http ://papersjds14.sfds.asso.fr/submission81.pdf.

[III.4] **Somé, S.M.**, Libengué, F.G. & Kokonendji, C.C. (2013). Estimation de densité par noyau beta bivarié avec structure de corrélation. *Actes des 45èmes Journées de*

Statistique de la Société Française de Statistiques (SFdS), 6pages, Toulouse 27-31 Mai 2013. URL `http://papersjds13.sfds.asso.fr/submission47.pdf`.

[III.5] Libengué, F.G., **Somé, S.M.** & Kokonendji, C.C. (2013). Estimation par noyaux associés mixtes d'un modèle de mélange. *Actes des 45èmes Journées de Statistique de la Société Française de Statistiques (SFdS)*, 6pages, Toulouse 27-31 Mai 2013. URL `http://papersjds13.sfds.asso.fr/submission82.pdf`.

IV. Autres communications orales et séminaires universitaires

[IV.1] **Somé, S.M.** (2014a). Effects of associated kernels on multiple regression. *Séminaire de Probabilités et Statistique*. Pau 2014.12.04.

[IV.2] **Somé, S.M.** (2014b). Nonparametric multiple regression by associated kernels *4èmes Journées Internationales du centre d'excellence de l'UEMOA : Laboratoire d'Analyse Numérique d'Informatique et de BIOMathématique (LANIBIO) de l'Université de Ouagadougou*. Ouagadougou 2014.07.29.

[IV.3] **Somé, S.M.** (2014c). Régression multiple non-paramétrique par noyaux associés. *15èmes Journées de L'Ecole Doctorale CARNOT-PASTEUR des universités de Bourgogne et de Franche-Comté*. Besançon 2014.05.27.

[IV.4] **Somé, S.M.** (2013a). Estimation par noyaux associés mixtes d'un modèle de mélange. *5èmes Rencontres des Jeunes Statisticiens*. Aussois 2013.08.27.

[IV.5] **Somé, S.M.** (2013b). Estimation de densité par noyau beta bivarié avec structure de corrélation. *3èmes Journées de Rencontre Dijon-Besançon en Probabilités et Statistique*. Besançon 2013.04.05.

Résumé

Dans ce travail, l'approche non-paramétrique par noyaux associés mixtes multivariés est présentée pour les fonctions de densités, de masse de probabilité et de régressions à supports partiellement ou totalement discrets et continus. Pour cela, quelques aspects essentiels des notions d'estimation par noyaux continus (dits classiques) multivariés et par noyaux associés univariés (discrets et continus) sont d'abord rappelés. Les problèmes de supports sont alors révisés ainsi qu'une résolution des effets de bords dans les cas des noyaux associés univariés. Le noyau associé multivarié est ensuite défini et une méthode de leur construction dite *mode-dispersion multivarié* est proposée. Il s'ensuit une illustration dans le cas continu utilisant le noyau bêta bivarié avec ou sans structure de corrélation de type Sarmanov. Les propriétés des estimateurs telles que les biais, les variances et les erreurs quadratiques moyennes sont également étudiées. Un algorithme de réduction du biais est alors proposé et illustré sur ce même noyau avec structure de corrélation. Des études par simulations et applications avec le noyau bêta bivarié avec structure de corrélation sont aussi présentées. Trois formes de matrices des fenêtres, à savoir, pleine, Scott et diagonale, y sont utilisées puis leurs performances relatives sont discutées. De plus, des noyaux associés multiples ont été efficaces dans le cadre de l'analyse discriminante. Pour cela, on a utilisé les noyaux univarié binomial, catégoriel, triangulaire discret, gamma et bêta. Par la suite, les noyaux associés avec ou sans structure de corrélation ont été étudiés dans le cadre de la régression multiple. En plus des noyaux univariés ci-dessus, les noyaux bivariés avec ou sans structure de corrélation ont été aussi pris en compte. Les études par simulations montrent l'importance et les bonnes performances du choix des noyaux associés multivariés à matrice de lissage pleine ou diagonale. Les travaux ont aussi donné lieu à la création d'un package R pour l'estimation de fonctions univariés de densités, de masse de probabilité et de régression. Plusieurs méthodes de sélections de fenêtres optimales y sont implémentées avec une interface facile d'utilisation. Tout au long de ce travail, la sélection des matrices de lissage se fait généralement par validation croisée et parfois par les méthodes bayésiennes. Enfin, des compléments sur les constantes de normalisations des estimateurs à noyaux associés des fonctions de densité et de masse de probabilité sont présentés.

Mots clés : Analyse discriminante, corrélation de Sarmanov, effet de bords, fonction de masse de probabilité, matrice de dispersion, matrice de lissage, méthode bayésienne adaptative, noyau classique, régression multiple non-paramétrique, validation croisée profilée.

Abstract

This work is about nonparametric approach using multivariate mixed associated kernels for densities, probability mass functions and regressions estimation having supports partially or totally discrete and continuous. Some key aspects of kernel estimation using multivariate continuous (classical) and (discrete and continuous) univariate associated kernels are recalled. Problem of supports are also revised as well as a resolution of boundary effects for univariate associated kernels. The multivariate associated kernel is then defined and a construction by *multivariate mode-dispersion* method is provided. This leads to an illustration on the bivariate beta kernel with Sarmanov's correlation structure in continuous case. Properties of these estimators are studied, such as the bias, variances and mean squared errors. An algorithm for reducing the bias is proposed and illustrated on this bivariate beta kernel. Simulations studies and applications are then performed with bivariate beta kernel. Three types of bandwidth matrices, namely, full, Scott and diagonal are used. Furthermore, appropriated multiple associated kernels are used in a practical discriminant analysis task. These are the binomial, categorical, discrete triangular, gamma and beta. Thereafter, associated kernels with or without correlation structure are used in multiple regression. In addition to the previous univariate associated kernels, bivariate beta kernels with or without correlation structure are taken into account. Simulations studies show the performance of the choice of associated kernels with full or diagonal bandwidth matrices. This is followed by an R package, created in univariate case, for densities, probability mass functions and regressions estimations. Several smoothing parameter selections are implemented via an easy-to-use interface. Throughout the paper, bandwidth matrix selections are generally obtained using cross-validation and sometimes Bayesian methods. Finally, some additionnal informations on normalizing constants of associated kernel estimators are presented for densities or probability mass functions.

Keywords: Adaptive Bayesian method, bandwidth matrix, boundary effects, classical kernel, correlation of Sarmanov, discriminant analysis, dispersion matrix, nonparametric multiple regression, probability mass function, profile cross-validation, smothing matrix.

Table des matières

Liste des tableaux

Liste des figures

Introduction générale

Un des problèmes habituellement rencontrés en statistique est celui de l'estimation de fonctions multivariés. Il s'agit d'un problème bien connu dans de nombreux domaines où l'on a souvent à modéliser des phénomènes complexes. On rencontre ce genre de questions en économie, médecine, sociologie, environnement, marketing, etc. Dans le cas multivarié, ces derniers sont souvent décrits par des vecteurs aléatoires à valeurs réelles et à support connu dans $\mathbb{T}_d \subseteq \mathbb{R}^d$ avec $d \in \{1, 2, \ldots\}$. Cet ensemble peut être formé d'axes univariés continus, discrets ou mixtes (à la fois discrets et continus). On parlera d'*échelles de temps* ("time scales" en anglais) dans le cas d'axes mixtes univariés. On peut se référer à Hilger (1990), Agarwal & Bohner (1999), Bohner & Peterson (2001, 2003) pour d'autres développements. Il faut noter que le support $\mathbb{T}_d$ est parfois relatif à des données fonctionnelles. Le lecteur pourra se référer, par exemple, à Cardot *et al.* (2015), Dabo-Niang & Yao (2013) et Amiri *et al.* (2014) pour des utilisations en sondages, analyse spatiale et régression, respectivement.

On considère une fonction multivariée f à estimer à partir des n observations du vecteur aléatoire réel (v.a.r.) dans $\mathbb{T}_d$ ($\subseteq \mathbb{R}^d$). Cette fonction f peut être une fonction de densité (donc continue), de masse de probabilité ou de régression (continue et/ou discrète). Puisqu'on ne dispose pas de manière générale de modèles paramétriques, cette thèse propose l'approche non-paramétrique pour estimer la fonction f par la méthode des noyaux dits *associés multivariés* afin de respecter, à la fois, les différentes structures topologiques de $\mathbb{T}_d$ et les structures de corrélations entre les variables d'intérêts. Dans le cas d'une fonction de densité, l'estimateur à noyau associé multivarié est donné par :

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i), \quad \forall \mathbf{x} \in \mathbb{T}_d, \tag{1}$$

où $K_{\mathbf{x},\mathbf{H}}(\cdot)$ sera le "noyau associé" lequel est intrinsèquement lié à la cible $\mathbf{x}$ et à la matrice des fenêtres $\mathbf{H}$. Les noyaux associés généralisent les "noyaux classiques" de Rozenblatt (1956) et Parzen (1962). Ces derniers sont très populaires dans la littérature et approprié pour le support continu $\mathbb{T}_d = \mathbb{R}^d$. Ils servent entre autre à l'estimation de quantiles de régression, de distributions à queues lourdes et aussi en analyse discriminante. Le lecteur consultera respectivement Daouia *et al.* (2013), Markovich (2008) et Hall & Kang (2005) pour ces situations. Plusieurs autres méthodes non-paramétriques existent et peuvent être utilisées pour ces problèmes d'estimations. On peut citer les travaux de Antoniadis (1997) et Vidakovic (1999) pour les ondelettes, Wahba (1990) pour les splines et plus généralement Wassermann (2006) pour une revue de ces méthodes non-paramétriques. De manière générale, le v.a.r. $\mathbf{X}$ à un support pouvant s'écrire

$$\mathbb{T}_d = ([t_0, t_1] \cup \{t_2, \ldots, t_k\} \cup]t_{k+1}, +\infty[)^d \subseteq \mathbb{R}^d, \quad \forall d \in \{1, 2, \ldots\}, \tag{2}$$

où $t_j \in [-\infty, +\infty]$ avec $t_j \leq t_{j+1}$. Des cas usuels de (2) traités dans cette thèse sont $\mathbb{R}^d$, $\mathbb{N}^d$, $\mathbb{R}^{d-\ell} \times \mathbb{N}^\ell$. Un autre exemple de (2) concerne l'estimation de densités où la v.a.r. d'intérêt $\mathbb{T}_1$ est constituée à la fois de parties continues (intervalles) et discrètes. On parlera alors d'*échelles de temps* ("time scales" en anglais). Cette notion a été introduite par Stefan Hilger dans sa thèse en 1988 pour unifier l' Analyse topologique des continues avec des discrètes ; voir Hilger (1990), Agarwal & Bohner (1999), Bohner & Peterson (2001, 2003), Sanyal (2008) pour d'autres développements. En univarié, ces outils ont été utilisés pour developper la méthode des noyaux associés mixtes pour l'estimation de telles densités. On peut se référer à Libengué *et al.* (2013) et aussi à Libengué (2013) pour plus de détails.

L'objectif principal de cette thèse est d'étendre au multivarié les estimateurs à noyaux associés univariés discrets de Senga Kiéssé (2008) et univariés continus de Libengué (2013). En multivarié, les noyaux associés doivent conserver leur aptitude à respecter le support $\mathbb{T}_d$ et aussi à atteindre n'importe quelle cible. Dans cet environnement multidimensionnel, on étudiera l'effet de la corrélation pour des fonctions de densité et de régression. Sans perte de généralité, on s'intéressera à des fonctions par rapport à la mesure de référence du support $\mathbb{T}_d$ de (2), principalement Lebesgue et dénombrement sur chacun des axes. La question de l'analyse discriminante par noyaux associés sera aussi abordée. En univarié, on proposera une vulgarisation de la méthode des noyaux associés sous le logiciel `R` (R Development Core Team, 2015). Comptant sur la bonne compréhension des lecteurs, les Chapitres 2 à 5 sont rédigés en anglais puisqu'étant des tapuscrits soumis pour publication. Pour éviter le gaspillage, chacun de ces chapitres a été rédigé sous une forme "prête pour publication" et contient sa propre introduction, **au prix de quelques redites**.

Le Chapitre 1 rappelle l'état de l'art concernant la méthode à noyaux classiques multivariés, ainsi qu'associés univariés (discrets et continus) pour l'estimation de fonctions de densité, de masse de probabilité et de régression. Les propriétés essentielles des estimateurs à noyaux classiques sur $\mathbb{T}_d = \mathbb{R}^d$ sont rappelés. On souligne les problèmes liés aux différentes structures de supports ainsi qu'aux formes de la matrice de lissage. Un clin d'oeil concernant l'application des estimateurs à noyaux classiques multivariés est aussi effectué. Un récapitulatif des "lissages discrets et continus" univariés sur $\mathbb{T}_1$ ($\subseteq \mathbb{N}$ et $\mathbb{R}$) par des noyaux associés est donné. Différentes propriétés y sont revues ; en particulier une résolution pratique des problèmes de bords pour le noyau triangulaire discret ainsi qu'un algorithme de réduction pour l'estimateur à noyaux associés continus. L'autre partie de ce chapitre résume brièvement l'utilisation des noyaux classiques et associés non classiques en régression multiple.

Dans la Chapitre 2, une méthode de construction des noyaux associés multivariés prenant en compte différentes formes de la matrice de lissage est proposée dans le cas continu ainsi qu'une illustration sur le noyau bêta bivarié avec structure de corrélation. On élabore aussi les propriétés caractéristiques de ces estimateurs de densité et un algorithme de réduction du biais est proposé et illustré sur ce noyau bêta bivarié avec corrélation. Le problème de sélection de matrice des fenêtres optimales y est également abordé.

Dans le Chapitre 3, les noyaux associés multivariés utilisant le noyau bêta bivarié avec corrélation ainsi que plusieurs noyaux associés multiples composés d'univariés

bêta, gamma, binomial et triangulaire discret sont aussi utilisés. Ces noyaux associés multiples ont fait l'objet d'étude dans le cadre pratique de l'analyse discriminante. Pour ce problème de classification supervisée, une sélection de matrice par validation croisée profilée a été introduite pour des v.a.r. mixtes.

Au travers des simulations et applications, le Chapitre 4 montre la pertinence et l'efficacité des noyaux associés multivariés (avec ou sans structure de corrélation) dans le problème de régression multiple. En plus des noyaux univariés ci-dessus, les noyaux Epanexchnikov et bêta bivarié avec ou sans structure de corrélation y seront aussi utilisés.

Dans le Chapitre 5, un package `R` a été créé pour vulgariser les méthodes à noyaux associés univariés (discrets et continus) pour l'estimation de fonction de masse de probabilité, de densité et de régression. Plusieurs méthodes de sélection de paramètre de lissage univarié y ont été rappelées et implémentées ainsi que la nouvelle approche bayésienne adaptative pour l'estimateur à noyau gamma.

Le Chapitre 6 fournit quelques discussions et compléments des différents résultats. On s'intéressera en particulier à la constante de normalisation pour les fonctions de densités multivariées. Aussi, une approche bayésienne adaptative est proposée pour la sélection de matrice de lissages optimales pour l'estimateur de densité à noyau gamma multiple.

Enfin, cette thèse se termine par une conclusion et des perspectives. Une liste des références est alors donnée. Une annexe contenant les codes `R` des Chapitres 2 à 5 est également fournie.

Chapitre 1

Contexte et contributions

Cette section rappelle les méthodes d'estimations par noyaux continus multivariés ainsi que associés discrets et continues univariés. On s'intéressera d'abord aux problèmes d'estimations de fonctions de densité (f.d.p.) ou de masse de probabilité (f.m.p.). Pour ces derniers, les estimateurs à noyaux continus (classiques) et associés (discrets et continus) seront présentés avec leurs propriétés de base et les différentes méthodes de choix de matrices des fenêtres. Les problèmes de régression multiple et d'analyse discriminante par les noyaux seront succinctement décrits.

1.1 Estimation par noyaux de densité et de fonction masse de probabilité

Tout au long de cette partie, on suppose que les variables à observer $\mathbf{X}_1, \ldots, \mathbf{X}_n$ sont des vecteurs aléatoires indépendants et identiquement distribués (i.i.d.) de même loi de densité f par rapport à la mesure de référence (Lebesgue ou comptage, voire mixte) sur $\mathbb{T}_d \subseteq \mathbb{R}^d$ avec $d \geq 1$.

1.1.1 Cas des noyaux classiques multivariés et analyse discriminante

Cette section présente la méthode d'estimation de densité par noyaux continus de Rozenblatt (1956) et Rosenblatt (1962) et qu'on appellera ici "noyau classique". On suppose que les observations $\mathbf{X}_i$ sont à valeurs réelles et que f est la densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d =: \mathbb{T}_d$.

Définition 1.1.1 *Une fonction $\mathcal{K}$ de support continu $\mathbb{S}_d (\subseteq \mathbb{R}^d)$ est dite **noyau classique** (donc continu) si elle est une densité de probabilité symétrique (i.e. $\mathcal{K}(-\mathbf{u}) = \mathcal{K}(\mathbf{u})$), de vecteur moyen $\boldsymbol{\mu}_{\mathcal{K}}$ nul ($\boldsymbol{\mu}_{\mathcal{K}} := \int_{\mathbb{S}_d} \mathbf{u}\mathcal{K}(\mathbf{u})d\mathbf{u} = \mathbf{0}$), de matrice de variance-covariance $\boldsymbol{\Sigma}_{\mathcal{K}}$ $\left(\boldsymbol{\Sigma}_{\mathcal{K}} := \int_{\mathbb{S}_d} \mathbf{u}\mathbf{u}^\top \mathcal{K}(\mathbf{u})d\mathbf{u}\right)$ et de carré intégrable $\left(\int_{\mathbb{S}_d} \mathcal{K}^2(\mathbf{u})d\mathbf{u} < +\infty\right)$.*

Précisons ici qu'en tant que densité de probabilité, le noyau classique $\mathcal{K}$ est positif et de masse totale égale à l'unité (i.e. pour tout élément $\mathbf{u}$ de $\mathbb{S}_d$, $\mathcal{K}(\mathbf{u}) \geq \mathbf{0}$ et $\int_{\mathbb{S}_d} \mathcal{K}(\mathbf{u})d\mathbf{u} =$

1).

Définition 1.1.2 *Soit $\mathbf{H}_n$ une matrice des fenêtres (i.e. symétrique, définie positive et $\mathbf{H}_n \to \mathbf{0}_d$ quand $n \to +\infty$) et $\mathcal{K}$ la fonction noyau vérifiant la Définition 1.1.1. L'estimateur à noyau classique multivarié de f peut être défini en un point $\mathbf{x} \in \mathbb{T}_d$ par*

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n \det \mathbf{H}_n} \sum_{i=1}^{n} \mathcal{K}(\mathbf{H}_n^{-1}(\mathbf{x} - \mathbf{X}_i)), \tag{1.1}$$

ou

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n(\det \mathbf{H}_n)^{1/2}} \sum_{i=1}^{n} \mathcal{K}(\mathbf{H}_n^{-1/2}(\mathbf{x} - \mathbf{X}_i)). \tag{1.2}$$

On peut unifier cette définition par

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_n}(\mathbf{x} - \mathbf{X}_i), \tag{1.3}$$

où $\mathcal{K}_{\mathbf{H}_n}(\cdot) = [1/(\det \mathbf{H}_n)]\,\mathcal{K}(\mathbf{H}_n^{-1}\cdot)$ pour (1.1) et $\mathcal{K}_{\mathbf{H}_n}(\cdot) = \left[1/(\det \mathbf{H}_n)^{-1/2}\right]\mathcal{K}(\mathbf{H}_n^{-1/2}\cdot)$ pour (1.2). Notons que la formulation (1.1) est la plus connue alors que (1.2) a été introduite par Wand & Jones (1993) pour faciliter une technique de choix de matrice des fenêtres. Pour ce dernier, on observe qu'en univarié la matrice de lissage d'ordre 1×1 est $\mathbf{H}_n = h_n^2$; ce qui motivera une considération des matrices de dispersion pour servir de matrices des fenêtres dans (1). Aussi, on peut remarquer que l'estimateur (1.3) en un point donné $\mathbf{x} \in \mathbb{T}_d$ est la moyenne des valeurs des fonctions noyaux centrées sur chaque observation $\mathbf{X}_i$ de l'échantillon, et à comparer plus tard à (1).

Dans (1.3) ainsi que (1), le paramètre de lissage $\mathbf{H}_n$ a une influence certaine. Lorsqu'il tend vers la matrice nulle $\mathbf{0}_d$, la fonction noyau s'apparente à la fonction Dirac et plus on manque d'informations (sous lissage) ; tandis que, lorsque $\mathbf{H}_n$ devient grand, plus on a d'informations en trop (sur lissage). Il est bien connu que le choix de $\mathbf{H}_n$ est d'une importance capitale dans le processus, contrairement aux choix de noyaux. Ainsi, l'obtention d'une estimation plus proche de la réalité implique donc un choix optimal de $\mathbf{H}_n$; ceci fera aussi l'objet de ce chapitre. Dans toute la suite, sauf mention expresse du contraire, la matrice des fenêtres $\mathbf{H}_n$ sera remplacée par $\mathbf{H}$.

Le type d'orientation de la fonction noyau classique $\mathcal{K}$ est contrôlé par la paramétrisation de la matrice de lissage. Wand & Jones (1993) ont étudié les différentes formes de paramétrisations de la matrice de lissage bivarié $\mathbf{H}$. Les auteurs ont dénombré trois formes principales et trois autres hybrides :

(i) la classe des matrices "pleines" symétriques définies positives $\mathbf{H} = \begin{pmatrix} h_1 & h_{12} \\ h_{12} & h_2 \end{pmatrix}$

(ii) la classe des matrices diagonales $\mathbf{H} = \operatorname{diag}(h_1, h_2)$

(iii) la classe formée du produit d'une constante positive et de la matrice identité $\mathbf{H} = h\mathbf{I}_2$

(iv) la classe formée du produit d'une constante positive et de la matrice de variance-covariance empirique $h\mathbf{S} = \begin{pmatrix} hS_1^2 & hS_{12} \\ hS_{12} & hS_2^2 \end{pmatrix}$

(v) la classe formée d'une constante positive et de la diagonale de $\boldsymbol{S}$: $\begin{pmatrix} hS_1^2 & 0 \\ 0 & hS_2^2 \end{pmatrix}$

(vi) la classe des matrices obtenues en utilisant le coefficient de corrélation ρ_{12} pour déterminer la rotation : $\begin{pmatrix} h_1^2 & \rho h_1 h_2 \\ \rho h_1 h_2 & h_2^2 \end{pmatrix}$.

Ces différentes paramétrisations sont décrites et illustrées par la Figure 1.1. Pour une utilisation générale, la forme (iii) est la moins intéressante parce que trop restrictive. Comme le montre la Partie (b) de la Figure 1.1, les contours de la densité cible sont des ellipses pendant que celles du noyau sont circulaires. Les formes (iv) et (v) imposent au noyau de s'aligner suivant la matrice de variance-covariance de la cible f et ne sont donc pas conseiller de manière générale (voir Partie (c) de la Figure 1.1). La troisième forme hybride (vi) dépend du choix judicieux du paramètre de corrélation du noyau. La Partie (d) de la Figure 1.1 montre bien que le noyau classique est orienté suivant la matrice de corrélation et n'est donc pas généralement utilisé. La forme diagonale (ii) est la plus utilisée mais reste inappropriée pour des densités cibles à forte structure de corrélation comme la Partie (a) de la Figure 1.1. Le noyau classique utilisant cette forme garde la même orientation selon les axes quand bien même la densité cible f est fortement corrélée (i.e. orientée de manière oblique). Puisqu'on s'intéresse à toutes les formes de corrélation de f, la plus générale est (i) qui a une matrice de lissage pleine.

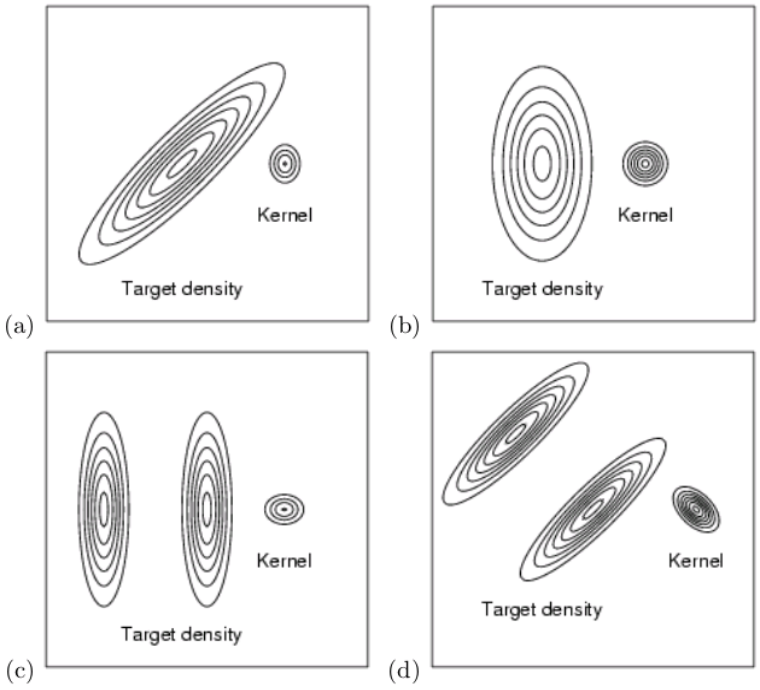


FIGURE 1.1 – Paramétrisations de matrice de lissage (Duong, 2004) : densité cible et forme du noyau associé classique.

Erreur Quadratique Moyenne

On étudie ici l'Erreur Quadratique Moyenne ponctuelle, en anglais "pointwise Mean Squared Error" (MSE) pour l'estimateur $\widehat{f_n}$ défini en (1.1), et donnée par

$$\text{MSE}(\mathbf{x}) = \mathbb{E}\left[\left\{\widehat{f_n}(\mathbf{x}) - f(\mathbf{x})\right\}^2\right], \quad \mathbf{x} \in \mathbb{T}_d. \tag{1.4}$$

Un develeppement de cette expression produit les variances et biais ponctuels de l'estimateur $\widehat{f_n}$ notés respectivement $\text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\}$ et $\text{Biais}\left\{\widehat{f_n}(\mathbf{x})\right\}$:

$$\begin{aligned}
\text{MSE}(\mathbf{x}) &= \mathbb{E}\left[\left\{\widehat{f_n}(\mathbf{x})\right\}^2\right] - 2\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} f(\mathbf{x}) + \left[\mathbb{E}\left\{f(\mathbf{x})\right\}\right]^2 \\
&= \mathbb{E}\left[\left\{\widehat{f_n}(\mathbf{x})\right\}^2\right] - \left[\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\}\right]^2 + \left[\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\}\right]^2 - 2f(\mathbf{x})\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} + f^2(\mathbf{x}) \\
&= \mathbb{E}\left[\left\{\widehat{f_n}(\mathbf{x})\right\}^2\right] - \left[\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\}\right]^2 + \left(\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} - f(\mathbf{x})\right) \\
&= \text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} + \text{Biais}^2\left\{\widehat{f_n}(\mathbf{x})\right\},
\end{aligned} \tag{1.5}$$

où

$$\text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} = \mathbb{E}\left(\left[\widehat{f_n}(\mathbf{x}) - \mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\}\right]^2\right) \tag{1.6}$$

et

$$\text{Biais}\left\{\widehat{f_n}(\mathbf{x})\right\} = \mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} - f(\mathbf{x}). \tag{1.7}$$

Le terme $\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\}$ dans (1.6) et (1.7) montre bien l'effet qu'une augmentation/réduction peut avoir sur le biais. Les expressions des biais et variance peuvent être affinées en utilisant le noyau classique $\mathcal{K}$:

$$\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} = \int_{\mathbb{T}_d} \mathcal{K}_{\mathbf{H}}(\mathbf{t}-\mathbf{x}) f(\mathbf{t}) d\mathbf{t} = \frac{1}{\det \mathbf{H}} \int_{\mathbb{T}_d} \mathcal{K}\left\{\mathbf{H}^{-1}(\mathbf{t}-\mathbf{x})\right\} f(\mathbf{t}) d\mathbf{t} \tag{1.8}$$

et

$$\begin{aligned}
\text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} &= \frac{1}{n}\text{Var}\left\{\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}}(\mathbf{t}-\mathbf{x})\right\} = \frac{1}{n(\det \mathbf{H})^2}\left[\mathbb{E}\left\{\mathcal{K}^2\left\{\mathbf{H}^{-1}(\mathbf{t}-\mathbf{x})\right\}\right\} - \mathbb{E}^2\left\{\mathcal{K}\left\{\mathbf{H}^{-1}(\mathbf{t}-\mathbf{x})\right\}\right\}\right], \\
&= \frac{1}{n(\det \mathbf{H})^2}\int_{\mathbb{T}_d} \mathcal{K}^2\left\{\mathbf{H}^{-1}(\mathbf{t}-\mathbf{x})\right\} f(\mathbf{t}) d\mathbf{t} - \frac{1}{n(\det \mathbf{H})^2}\mathbb{E}^2\left[\mathcal{K}\left\{\mathbf{H}^{-1}(\mathbf{t}-\mathbf{X}_1)\right\}\right].
\end{aligned} \tag{1.9}$$

L'erreur moyenne quadratique peut être vu aussi comme une mesure la consistance de l'estimateur. Pour le prouver, on réécrit d'abord l'espérance de l'estimateur $\widehat{f_n}$ dans (1.8) et la variance dans (1.9) avec $\mathbf{t} = \mathbf{x} - \mathbf{H}\mathbf{u}$:

$$\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} = \int_{\mathbb{T}_d} \mathcal{K}(\mathbf{u}) f(\mathbf{x} - \mathbf{H}\mathbf{u}) d\mathbf{u} \tag{1.10}$$

et

$$\text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} = \frac{1}{n(\det \mathbf{H})}\int_{\mathbb{T}_d} \mathcal{K}^2(\mathbf{u}) f(\mathbf{x} - \mathbf{H}\mathbf{u}) d\mathbf{u} - \frac{1}{n(\det \mathbf{H})^2}\mathbb{E}^2\left[\mathcal{K}\left\{\mathbf{H}^{-1}(\mathbf{t}-\mathbf{X}_1)\right\}\right]. \tag{1.11}$$

On rappelle que sous les hypothèses de différentiabilité de tous ordres de f, le développement en séries de Taylor de $f(\mathbf{x}-\mathbf{Hu})$ au voisinage de $\mathbf{x}$ donné par

$$f(\mathbf{x}-\mathbf{Hu}) = f(\mathbf{x}) - \mathbf{u}^\top\mathbf{H}^\top\nabla f(\mathbf{x}) + \frac{1}{2}\mathbf{u}^\top\mathbf{H}^\top\nabla^2 f(\mathbf{x})\mathbf{Hu} + \mathrm{o}\left\{\mathbf{u}^\top\mathbf{H}^2\mathbf{u}\right\}. \tag{1.12}$$

La forme quadratique dans (1.12) est une matrice 1×1 qui est trivialement égale à sa trace. On rappelle aussi que l'opérateur trace satisfait trace($\mathbf{AB}$) = trace($\mathbf{BA}$) pour toutes matrices $\mathbf{A}$ et $\mathbf{B}$ de dimensions respectives $r\times s$ et $s\times r$. De ces résultats de la trace et en injectant (1.12) dans l'espérance (1.10) et la variance (1.11), on obtient alors :

$$\begin{aligned}\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} &= \int_{\mathbb{T}_d}\mathcal{K}(\mathbf{u})\left\{f(\mathbf{x}) - \mathbf{u}^\top\mathbf{H}^\top\nabla f(\mathbf{x}) + \frac{1}{2}\mathbf{u}^\top\mathbf{H}^\top\nabla^2 f(\mathbf{x})\mathbf{Hu} + \mathrm{o}\left(\mathbf{u}^\top\mathbf{H}^2\mathbf{u}\right)\right\}d\mathbf{u}\\ &= f(\mathbf{x}) + \frac{1}{2}\operatorname{trace}\left(\mathbf{\Sigma}_{\mathcal{K}}\cdot\mathbf{H}^\top\nabla^2 f(\mathbf{x})\mathbf{H}\right) + \mathrm{o}\left\{\operatorname{trace}\left(\mathbf{H}^2\right)\right\}\end{aligned} \tag{1.13}$$

et

$$\begin{aligned}\operatorname{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} &= \frac{1}{n\,(\det\mathbf{H})}\int_{\mathbb{T}_d}\mathcal{K}^2(\mathbf{u})\,f(\mathbf{x})d\mathbf{u} + R_n + \mathrm{o}\left(n^{-1}\,(\det\mathbf{H})\right)\\ &= \frac{1}{n\,(\det\mathbf{H})}f(\mathbf{x})\int_{\mathbb{T}_d}\mathcal{K}^2(\mathbf{u})\,d\mathbf{u} + \mathrm{o}\left(\frac{1}{n\,(\det\mathbf{H})}\right),\end{aligned} \tag{1.14}$$

avec

$$\begin{aligned}R_n &= \frac{1}{n\,(\det\mathbf{H})^2)}\left(-\mathbf{u}^\top\mathbf{H}^\top\nabla f(\mathbf{x}) + \frac{1}{2}\mathbf{u}^\top\mathbf{H}^\top\nabla^2 f(\mathbf{x})\mathbf{Hu} - \mathbb{E}^2\left[\mathcal{K}\left\{\mathbf{H}^{-1}(\mathbf{t}-\mathbf{X}_1)\right\}\right]^2\right)\\ &\simeq \mathrm{o}\left\{1/(n\ \det\mathbf{H})\right\}.\end{aligned}$$

De ces deux derniers résultats (1.14) et (1.13), on peut déduire la forme approximée et asymptotique du MSE notées AMSE par :

$$\operatorname{AMSE}(\mathbf{x}) = \frac{1}{n\,(\det\mathbf{H})}f(\mathbf{x})\int_{\mathbb{T}_d}\mathcal{K}^2(\mathbf{u})d\mathbf{u} + \frac{1}{4}\left[\operatorname{trace}\left\{\mathbf{\Sigma}_{\mathcal{K}}\cdot\mathbf{H}^\top\nabla^2 f(\mathbf{x})\mathbf{H}\right\}\right]^2. \tag{1.15}$$

Une mesure, globale, de l'efficacité de l'estimateur $\widehat{f_n}$ est obtenue en intégrant le MSE sur tout le support $\mathbb{T}_d$ de f. Il s'agit de "l'Erreur Quadratique Moyenne Intégrée", en anglais "Mean Integrated Squared Error" (MISE). En utilisant (1.5), elle s'écrit :

$$\begin{aligned}\operatorname{MISE}(n,\mathbf{H},\mathcal{K},f) &= \int_{\mathbb{T}_d}\operatorname{MSE}(\mathbf{x})d\mathbf{x}\\ &= \int_{\mathbb{T}_d}\operatorname{Var}\{\widehat{f_n}(\mathbf{x})\}d\mathbf{x} + \int_{\mathbb{T}_d}\operatorname{Biais}^2\{\widehat{f_n}(\mathbf{x})\}d\mathbf{x}.\end{aligned} \tag{1.16}$$

L'expression approchée du MISE notée AMISE est obtenue en intégrant le AMSE de (1.15) :

$$\operatorname{AMISE}(n,\mathbf{H},\mathcal{K},f) = \frac{1}{n\,(\det\mathbf{H})}\int_{\mathbb{T}_d}\mathcal{K}^2(\mathbf{u})d\mathbf{u} + \frac{1}{4}\int_{\mathbb{T}_d}\left[\operatorname{trace}\left\{\mathbf{\Sigma}_{\mathcal{K}}\cdot\mathbf{H}^\top\nabla^2 f(\mathbf{x})\mathbf{H}\right\}\right]^2 d\mathbf{x}. \tag{1.17}$$

Si on suppose de plus que $\mathbf{\Sigma}_{\mathcal{K}} = \mu_2(\mathcal{K})\mathbf{I}_d$ où $\mu_2(\mathcal{K}) > 0$ et $\mathbf{I}_d$ est la matrice unité d'ordre $d \times d$, alors l'AMISE (1.17) devient

$$\text{AMISE}(n, \mathbf{H}, \mathcal{K}, f) = \frac{1}{n\,(\det \mathbf{H})} \int_{\mathbb{T}_d} \mathcal{K}^2(\mathbf{u}) d\mathbf{u} + \frac{1}{4}\mu_2(\mathcal{K})^2 \int_{\mathbb{T}_d} \left[\text{trace}\left\{\mathbf{H}^\top \nabla^2 f(\mathbf{x})\mathbf{H}\right\}\right]^2 d\mathbf{x}. \tag{1.18}$$

Cette version (1.18) est différente de celle de Duong (2004) utilisant l'estimateur $\widehat{f_n}$ de (1.2). En effet, en remplaçant l'estimateur $\widehat{f_n}$ de (1.2) par celui de (1.1) dans les formules (1.8) à (1.18), on obtient alors

$$\text{AMISE}(n, \mathbf{H}, \mathcal{K}, f) = \frac{1}{n\,(\det \mathbf{H})^{1/2}} \int_{\mathbb{T}_d} \mathcal{K}^2(\mathbf{u}) d\mathbf{u} + \frac{1}{4}\mu_2(\mathcal{K})^2 (\text{vech}^\top \mathbf{H}) \mathbf{\Psi}_4 (\text{vech}\, \mathbf{H}), \tag{1.19}$$

où $\mathbf{\Psi}_4$ est la matrice carrée de taille $d(d+1)/2$ donnée par

$$\mathbf{\Psi}_4 = \int_{\mathbb{T}_d} \text{vech}\left(2\nabla^2 f(\mathbf{x}) - \text{diag}\, \nabla^2 f(\mathbf{x})\right) \text{vech}^\top \left(2\nabla^2 f(\mathbf{x}) - \text{diag}\, \nabla^2 f(\mathbf{x})\right) d\mathbf{x} \tag{1.20}$$

sous les mêmes conditions de régularité. On peut se référer à Henderson & Searle (1979) et à Magnus & Neudecker (1988) pour les notions de "vector half" vech. Cet opérateur transforme les éléments triangulaires inférieures d'une matrice d'ordre $d \times d$ en vecteur. Par exemple,

$$\text{vech}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a \\ c \\ d \end{pmatrix}.$$

Aussi, la fonction $\mathbf{\Psi}_4$ peut être définie explicitement en fonction d'éléments individuels de la matrice comme suit. Considérons $\mathbf{r} = (r_1, \ldots, r_d)^\top$ où les $r_1, \ldots, r_d$ sont des entiers positifs et $|\mathbf{r}| = r_1 + \cdots + r_d$. Alors, la dérivée partielle d'ordre $\mathbf{r}$ de f est

$$f^{(\mathbf{r})}(\mathbf{x}) = \frac{\partial^{|\mathbf{r}|}}{\partial_{x_1}^{r_1} \cdots \partial_{x_d}^{r_d}} f(\mathbf{x}) \tag{1.21}$$

et les dérivées fonctionnelles de $\mathbf{\Psi}_4$ sont

$$\psi_r = \int_{\mathbb{T}_d} f^{(\mathbf{r})}(\mathbf{x} f(\mathbf{x}) d\mathbf{x}. \tag{1.22}$$

Ceci implique que chaque élément de la matrice $\mathbf{\Psi}_4$ est une fonctionnelle ψ_r.

Exemples de noyaux classiques

Le plus utilisé des noyaux classiques multivariés est la loi normale multivariée, donnée par $\mathcal{N}(\boldsymbol{\nu}, \mathbf{\Sigma})$ et où $\boldsymbol{\nu}$ est le vecteur moyenne et $\mathbf{\Sigma}$ la matrice de variance covariance :

$$N(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}(\det \mathbf{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\nu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\nu})\right\}, \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{1.23}$$

En plus de la loi normale multidimensionnelle, on peut construire plusieurs autres noyaux classiques multivariés à partir des univariés. Quelques uns sont rappelés dans la Table 1.1. Tous ces noyaux univariés vérifient la Définition 1.1.1 et ont pour support $\mathbb{S}_1 = \mathbb{R}$ ou $[-1, 1]$.

Noyaux	Support	Densité	Efficacité
Epanechnikov	$[-1,1]$	$\frac{3}{4}(1-x^2)$	1
Cosinus	$[-1,1]$	$\frac{\pi}{4}cos(\frac{\pi}{2}x)$	0,999
Biweight	$[-1,1]$	$\frac{15}{16}(1-x^2)^2$	0,994
Triweight	$[-1,1]$	$\frac{35}{32}(1-x^2)^3$	0,987
Triangulaire	$[-1,1]$	$1-\|x\|$	0,986
Gaussien	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$	0,946
Unifiorme	$[-1,1]$	$\frac{1}{2}$	0,930
Double Epanechnikov	$[-1,1]$	$3\|x\|(1-\|x\|)$	0,816
Double Exponentielle	$\mathbb{R}$	$\frac{1}{2}\exp\{-\frac{1}{2}\|x\|\}$	0,759

TABLE 1.1 – Quelques noyaux classiques.

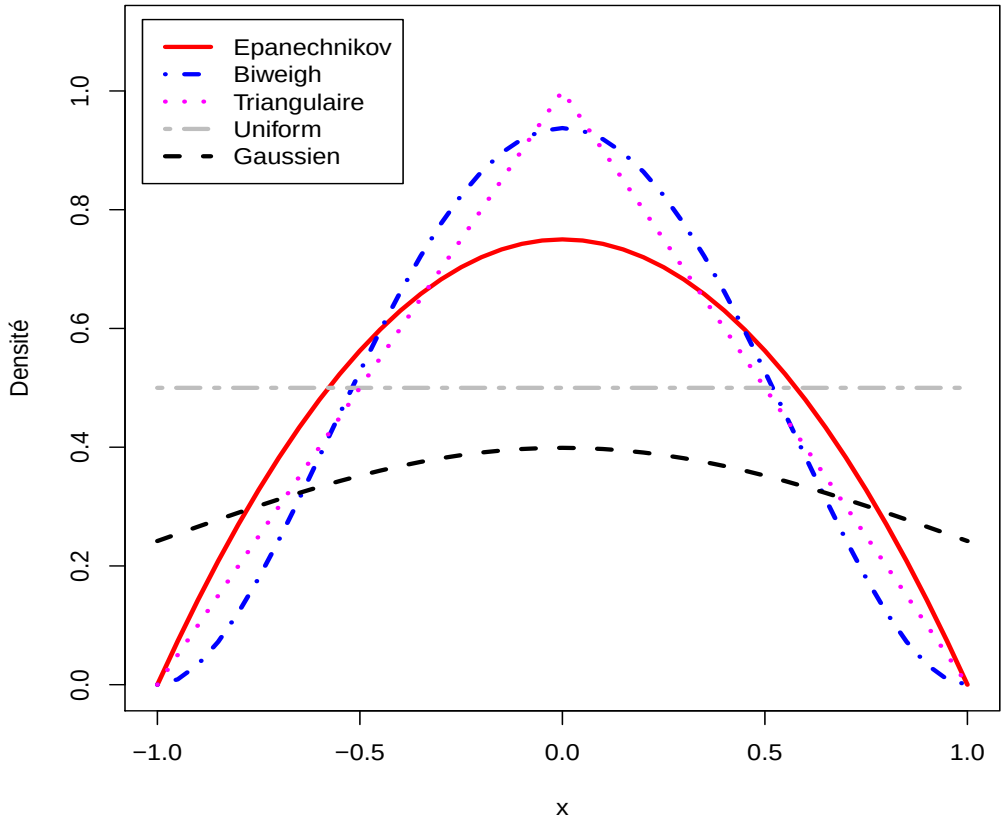


FIGURE 1.2 – Comportements de quelques noyaux classiques de même cible $x = 0$.

Choix de matrice des fenêtres

Dans ce qui suit, on rappelle trois méthodes pratiques de choix de matrice de fenêtres $\mathbf{H}$, à savoir la méthode de substitution ("Plug-in" en anglais), la méthode de validation croisée ainsi que l'approche bayésienne adaptative.

Méthode Plug-in

En univarié, la plus connue est due à Sheater & Jones (1991) dont une extension en multivarié est fournie par Wand & Jones (1994). Dans le cadre de l'estimateur à noyau gaussien sphérique de Duong (2004), obtenu avec (1.2), le critère de la méthode plug-in est donné par

$$PI(\mathbf{H}) = n^{-1}(4\pi)^{-d/2}(\det \mathbf{H})^{-1/2} + \frac{1}{4}(\mathrm{vech}^{\top}\, \mathbf{H})\widehat{\mathbf{\Psi}}_4(\mathrm{vech}\, \mathbf{H}), \tag{1.24}$$

où $\widehat{\mathbf{\Psi}}_4$ est une estimation de $\mathbf{\Psi}_4$ donnée en (1.20). En pratique, l'estimation $\widehat{\mathbf{\Psi}}_4$ utilise une estimation de ψ_r de (1.22) donnée par

$$\widehat{\psi_r}(\mathbf{G}) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f_n^{(\mathbf{r})}}(\mathbf{X}_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{K}_{\mathbf{G}}^{(\mathbf{r})}\left(\mathbf{X}_i - \mathbf{X}_j\right)$$

où $\mathbf{G} = g\mathbf{I}_2$ avec $g > 0$ est une matrice pilote.

Méthode de la validation croisée par les moindres carrées

Le principe de base de cette méthode est de minimiser par rapport à $\mathbf{H}$ un estimateur du MISE. La matrice des fenêtres optimale notée $\mathbf{H}_{cv}$, est donnée par

$$\mathbf{H}_{cv} = \arg\min_{\mathcal{H}} \mathrm{LSCV}(\mathbf{H}),$$

avec $\mathcal{H}$ l'espace des matrices de lissage symétriques définies positives et

$$\mathrm{LSCV}(\mathbf{H}) = \int_{\mathbb{T}_d} \left\{\widehat{f_n}(\mathbf{x})\right\}^2 d\mathbf{x} - \frac{2}{n}\sum_{i=1}^{n} \widehat{f_{n,-i}}(\mathbf{X}_i), \tag{1.25}$$

où $\widehat{f_{n,-i}}(\mathbf{X}_i) = (n-1)^{-1}\sum_{j\neq i} \mathcal{K}_{\mathbf{H}}\left(\mathbf{x} - \mathbf{X}_j\right)$ est l'estimateur calculé à partir de l'échantillon privé de l'observation $\mathbf{X}_i$. Cette méthode est juste la généralisation en multivarié des cas univariés de Bowman (1984) et Rudemo (1982). La difficulté ici réside dans le fait que la matrice est de dimension $d \times d$. Dans le cas le plus général de la matrice de lissage pleine, ceci se traduit par une minimisation sur $d(d+1)/2$ paramètres.

La validation croisée précédente est très coûteuse en temps de calcul. Duong (2004) a proposé, pour le noyau gaussien multivarié, des versions plus rapides qui utilisent des matrices de lissages pilotes. Ces versions sont basées sur la forme exacte de $\int_{\mathbb{T}_d} \mathcal{K}^2(\mathbf{u})d\mathbf{u}$ dans (1.25). On pourra se référer Duong (2007) pour des implémentations de ces méthodes sous le logiciel R.

Approche bayésienne adaptative

Soit $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ un échantillon de vecteurs aléatoires i.i.d. de densité inconnue f sur $\mathbb{T}_d$. L'estimateur $\widehat{f_n}$ de f est celui de (1.3). L'approche bayésienne adaptative consiste à associer une matrice de lissage $\mathbf{H}_i$ à chaque observation $\mathbf{X}_i$ et traite donc $\mathbf{H}_i$ comme une quantité aléatoire de loi *a priori* $\pi(\cdot)$. L'information apportée par les observations pour $\mathbf{H}_i$ est obtenue par une estimation par validation croisée de $f(\mathbf{X}_i)$:

$$\widehat{f}\Big(\mathbf{X}_i|\{\mathbf{X}_{-i}\}, \mathbf{H}_i\Big) = \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} \mathcal{K}_{\mathbf{H}_i}(\mathbf{X}_i - \mathbf{X}_j), \tag{1.26}$$

où $\{\mathbf{X}_{-i}\}$ est l'ensemble des observations recueillies en excluant $\mathbf{X}_i$, et $\mathcal{K}_{\mathbf{H}_i}$ est le noyau classique paramétré de (1.3).

À partir de la formule de Bayes et de (1.26), la loi *a posteriori* pour chaque $\mathbf{H}_i$ prend la forme suivante

$$\pi(\mathbf{H}_i|\mathbf{X}_i) = \pi(\mathbf{H}_i) \sum_{j=1, j\neq i}^{n} \mathcal{K}_{\mathbf{H}_i}(\mathbf{X}_i - \mathbf{X}_j) \left[\int_{\mathcal{H}} \pi(\mathbf{H}_i) \sum_{j=1, j\neq i}^{n} \mathcal{K}_{\mathbf{H}_i}(\mathbf{X}_i - \mathbf{X}_j)\, d\mathbf{H}_i \right]^{-1}, \tag{1.27}$$

où $\mathcal{H}$ est l'espace des matrices symétriques et définies positives.

Par l'estimateur de Bayes sous les pertes quadratique et entropique, le meilleur estimateur des $\mathbf{H}_i$ est la moyenne de $\pi(\mathbf{H}_i|\mathbf{X}_i)$ donnée respectivement par

$$\widetilde{\mathbf{H}}_i^{qua} = \mathbb{E}\Big(\mathbf{H}_i|\mathbf{X}_i\Big) = \int_{\mathcal{H}} \mathbf{H}_i \pi(\mathbf{H}_i|\mathbf{X}_i)\, d\mathbf{H}_i \tag{1.28}$$

et

$$\widetilde{\mathbf{H}}_i^{ent} = \Big[\mathbb{E}\Big(\mathbf{H}_i^{-1}|\mathbf{X}_i\Big)\Big]^{-1} = \left(\int_{\mathcal{H}} \mathbf{H}_i^{-1} \pi(\mathbf{H}_i|\mathbf{X}_i)\, d\mathbf{H}_i \right)^{-1}. \tag{1.29}$$

Les expressions (1.28) et (1.29) donnent dans certains cas des résultats explicites grâce à l'usage de lois *a priori* conjuguées ; on peut se référer à Zougab *et al.* (2014) utilisant le noyau gaussien multivarié et Brewer (2000) pour le cas gaussien univarié.

L'estimateur à noyau classique $\widehat{f_n}$ de (1.3), et de matrice de lissage bayésienne adaptative $\widetilde{\mathbf{H}}_i$ est

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_{\widetilde{\mathbf{H}}_i}(\mathbf{x} - \mathbf{X}_i),$$

où $\widetilde{\mathbf{H}}_i = \widetilde{\mathbf{H}}_i^{qua}$ pour (1.28) et $\widetilde{\mathbf{H}}_i = \widetilde{\mathbf{H}}_i^{ent}$ pour (1.29).

Analyse discriminante par noyaux classiques

Le problème d'estimation de densité par la méthode des noyaux est très important dans l'analyse exploratoire des données. En analyse discriminante, par exemple, il s'agit d'utiliser une règle de décision pour classer des observations $\mathbf{x} = (x_1, \ldots, x_d)^\top$ appartenant à un ensemble $\mathbb{T}_d$ ($\subseteq \mathbb{R}^d$) dans une des J classes prédéfinies. Spécifiquement,

la règle de Bayes optimale affecte une observation à la classe ayant la plus grande probabilité. Elle est donnée par :

$$\text{Allouer } \mathbf{x} \text{ au groupe } j_0 \text{ où } j_0 = \arg\max_{j\in\{1,\dots,J\}} \pi_j f_j(\mathbf{x}), \tag{1.30}$$

où les π_j sont les probabilités *a priori* et les $f_j(\mathbf{x})$ sont des fonctions de densités liées aux classes respectives ($j = 1, \dots, J$). On pourra se référer à Habbema *et al.* (1979), Hall & Kang (2005), Hall & Wand (1988), Gosh & Chaudhury (2004) et Gosh & Hall (2008) pour différents travaux sur le sujet, notamment l'estimation par noyaux classiques des densités f_j.

De manière générale, la règle de discrimination par noyaux classiques s'obtient en remplaçant dans (1.30) f_j par une estimation $\widehat{f_j}$ et π_j par la taille d'échantillon empirique n_j/n avec $\sum_{j=1}^{J} n_j = n$:

$$\text{Allouer } \mathbf{x} \text{ au } \textit{groupe } \widehat{j_0}, \text{ où } \widehat{j_0} = \arg\max_{j\in\{1,\dots,J\}} \widehat{\pi}_j \widehat{f_j}(\mathbf{x}). \tag{1.31}$$

On pourra se référer à Duong (2004, 2007) pour plus de détails. L'auteur y calcule le taux d'erreur de classification ("*misclassification rate*" en anglais) en comptant le nombre d'observations mal classés pour des données test $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ données :

$$\widehat{\mathrm{MR}} = 1 - m^{-1} \sum_{k=1}^{m} \mathbb{1}\{\mathbf{Y}_k \text{ est bien classé suivant (1.31)}\}. \tag{1.32}$$

Problèmes aux bords

De ce qui précède, les noyaux classiques, de par leurs propriétés, présente de nombreux avantages. Cependant, du fait de sa symétrie, elle devient inapproprié lorsqu'il s'agit d'estimer des densités à support partiellement ou totalement borné. En effet, elle assigne des poids à l'extérieur du support de la densité à estimer lorsque le lissage se fait aux points de bord. Cela crée ainsi les biais dits de bord qui rendent l'estimateur non consistant. En multivarié, les conséquences des problèmes de bords sont bien plus sévères en ce sens que les régions de bords augmentent avec la dimension d. Les solutions aux problèmes de bords ont été moins bien étudiées que dans le cas univarié.

Par exemple, Muller & Stadmüller (1999) ont proposé des noyaux de bords définies sur des supports arbitraires où les noyaux multivariés sont obtenus en minimisant un problème variationnel. On peut aussi se référer à Epanechnikov (1969). Bien que ces estimateurs aient des propriétés intéressantes, ils ne sont pas aisés à mettre en oeuvre. Pour le cas univarié, on pourra se référer, entre autre, aux méthodes dite des données reflétées ("reflection data method" en anglais) de Schuster (2000), Silverman (1986) et Cline & Hart (1991) et à celle des noyaux de bord ("boundary kernels method" en anglais) de Gasser & Müler (1979), Gasser *et al.* (1985) et Zhang & Karunamuni (2010).

Récemment, certains auteurs ont proposé dans le cas des densités à support compact, l'utilisation de noyaux dont le support coïncide avec celui de la densité à estimer. Ceci a efficacement résolu le problème des effets de bord puisque les noyaux utilisés ici sont généralement asymétriques et peuvent changer de forme selon la position du

point d'estimation. C'est notamment le cas, en univarié, de Chen (1999, 2000a) avec les noyaux bêta et gamma pour estimer les densités à support respectivement $[0,1]$ et $[0,+\infty[$, puis de Scaillet (2004) avec les noyaux inverse gaussien et sa réciproque pour les densités à support $]0,+\infty[$. Toujours dans le cas continu, Libengué (2013) a proposé une méthode de construction de noyaux associés continus qui englobe, entre autres, les noyaux de Chen (1999, 2000a) et de Scaillet (2004). Le cas discret a reçu l'attention de Senga Kiessé (2008) et Kokonendji & Senga Kiessé (2011) qui ont proposé l'utilisation des noyaux associés univariés discrets pour le lissage des données catégorielles ou de dénombrement.

En multivarié continu, Bouerzmani & Rombouts (2010) ont proposé des estimateurs à noyaux produits composés des univariés (bêta et gamma) de Chen (1999, 2000a). Dans le cas discret, les noyaux catégoriels sont les plus connus et adaptés pour, entre autres, les problèmes de classification (Aitchison & Aitken, 1976). Li & Racine (2007) ont utilisé ce produit de noyaux univariés auquel il a ajouter des noyaux univariés gaussiens pour les données continues. Tous ces noyaux multivariés ont une matrice de lissage diagonale alors que la rôle des matrices de lissages pleines a été démontré dans certaines situations (voir, par exemple, Figure 1.1).

1.1.2 Cas des noyaux associés univariés (discrets et continus)

Cette section présente l'essentiel des travaux d'Aictchison & Aitken (1976), Senga Kiéssé (2008), Kokonendji & Senga Kiéssé (2011) et Zougab *et al.* (2012) sur les noyaux associés univariés discrets. L'essentiel des travaux sur les noyaux univariés continus de Chen (1999, 2000ab) et Libengué (2013) est également présenté. On désignera par $\mathbb{T}_1$ le support de la fonction de densité (f.d.p.) ou de masse de probabilité (f.m.p.) à estimer. L'ensemble $\mathbb{T}_1$ sera soit continu $[a,b]$ ou $[0,+\infty[$ pour tout réels $a < b$ ou soit discret $\{0,1,\ldots,N\}$ pour tout entier N.

Définition 1.1.3 *On appelle type de noyau K_θ discret (resp. continu) toute fonction de masse de probabilité (resp. de densité) paramétrée par $\theta \in \Theta \subseteq \mathbb{R}^2$, de support $\mathbb{S}_\theta \subseteq \mathbb{Z}$ (resp. $\subseteq \mathbb{R}$) et de carré sommable (resp. intégrable).*

On ne considérera dans la suite que les types de noyaux discrets (resp. continus) pour la construction des noyaux associés discrets (resp. continus).

Définition 1.1.4 *(Kokonendji & Senga Kiéssé, 2011 ; Libengué, 2013) Soit $\mathbb{T}_1$ ($\subseteq \mathbb{R}$) le support de la f.m.p. ou de la f.d.p. à estimer, une cible $x \in \mathbb{T}_1$ et un paramètre de lissage h. Une f.m.p. (resp. f.d.p.) $K_{x,h}(\cdot)$ de support $\mathbb{S}_{x,h}$ ($\subseteq \mathbb{R}$) est appellée "noyau associé" lorsque les conditions suivantes sont satisfaites :*

$$x \in \mathbb{S}_{x,h}, \tag{1.33}$$

$$\mathbb{E}\left(\mathcal{Z}_{x,h}\right) = x + a(x,h), \tag{1.34}$$

$$\mathrm{Var}\left(\mathcal{Z}_{x,h}\right) = b(x,h), \tag{1.35}$$

où $\mathcal{Z}_{x,h}$ est la variable aléatoire de f.m.p. (resp. f.d.p.) $K_{x,h}$ et les quantités $a(x,h)$ et $b(x,h)$ tendent vers 0 quand h tend vers 0.

Notons que si l'on dispose d'un noyau associé discret ou continu, l'estimateur correspondant peut être facilement déduit. Si tel n'est pas le cas, il est possible de construire un noyau associé en utilisant le type de noyau approprié (continu ou discret). Le type de noyau utilisé doit avoir au moins le même nombre de paramètres que le nombre de composantes dans le couple (x,h) comme paramètres du noyau associé à construire ; voir aussi Définition 1.1.3. Dans le cas discret, il n'y a pas encore de méthode générale de construction. Par contre, dans le cas continu, Libengué (2013) a proposé une méthode dite "mode-dispersion" qui est rappelé ci-dessous.

Méthode de construction dans le cas continu

Principe 1.1.5 *(Libengué, 2013) Soit $K_{\theta(a,b)}$ un type de noyau uni-modal sur $\mathbb{S}_{\theta(a,b)}$, de mode $M(a,b)$ et de paramètre de dispersion $D(a,b)$. La méthode mode-dispersion permet la construction d'une fonction $K_{\theta(x,h)}$, dépendant de x et h en résolvant en a et b le système*

$$\begin{cases} M(a,b) = x \\ \\ D(a,b) = h. \end{cases}$$

Alors $\theta(x,h) = \theta\left(a(x,h), b(x,h)\right)$ où $a(x,h)$ et $b(x,h)$ sont solutions du système précédent, pour $h > 0$ et $x \in \mathbb{T}_1$ (support de la densité f à estimer).

On pourra se référer à Jørgensen (1997, 2013) et Jørgensen & Kokonendji (2011, 2015) pour de plus amples informations sur les notions de dispersion uni- et multivarié. La proposition suivante assure que la fonction noyau $K_{\theta(x,h)}$ issue de la méthode mode-dispersion est bien un noyau associé continu. En d'autres termes, il est montré que $K_{\theta(x,h)}$ satisfait les conditions de la Définition 1.1.4.

Proposition 1.1.6 *(Libengué, 2013) Soit $\mathbb{T}_1$ le support de la densité f, à estimer. Pour tout $x \in \mathbb{T}_1$ et $h > 0$, la fonction noyau $K_{\theta(x,h)}$ construite par la méthode mode-dispersion, de support $\mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}$, vérifie*

$$x \in \mathbb{S}_{\theta(x,h)}, \tag{1.36}$$

$$\mathbb{E}(\mathcal{Z}_{\theta(x,h)}) - x = A_\theta(x,h), \tag{1.37}$$

$$\mathrm{Var}(\mathcal{Z}_{\theta(x,h)}) = B_\theta(x,h), \tag{1.38}$$

où $\mathcal{Z}_{\theta(x,h)}$ est une variable aléatoire de densité $K_{\theta(x,h)}$ puis $A_\theta(x,h)$ et $B_\theta(x,h)$ tendent vers 0 quand h tend vers 0.

La définition suivante présente l'estimateur à noyau associé (continu ou discret) obtenu soit de la Définition 1.1.4, soit de la Proposition 1.1.6 ; c'est à dire $K_{\theta(x,h)} \equiv K_{x,h}$.

Définition 1.1.7 *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires i.i.d. de f.m.p. (resp. f.d.p.) f sur $\mathbb{T}_1$. L'estimateur à noyau associé discret (resp. continu) $\widehat{f_n}$ de f est défini par*

$$\widehat{f_n}(x) \quad = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i), \quad \forall x \in \mathbb{T}_1, \tag{1.39}$$

où $h > 0$ est le paramètre de lissage et $K_{x,h}$ est le noyau associé discret (resp. continu) dépendant de x et h.

On rappelle dans la suite quelques propriétés de l'estimateur (1.39) à noyau associé (continu et discret). Les propriétés asymptotiques ont, quant à elles, été étudiées par Abdous & Kokonendji (2009) pour le cas discret et Libengué (2013) pour le cas continu.

Proposition 1.1.8 *(Kokonendji & Senga Kiéssé, 2011 ; Libengué, 2013) Soit $X_1, X_2, \ldots, X_n$ un échantillon i.i.d. de f.m.p. (resp. f.d.p.) f sur $\mathbb{T}_1$. Soit $\widehat{f_n} = \widehat{f_{n,h,\mathcal{K}}}$ l'estimateur* (1.39) *de f utilisant un noyau associé. Alors, pour tout $x \in \mathbb{T}_1$ et $h > 0$, on a*

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

où $\mathcal{Z}_{x,h}$ est la variable aléatoire liée à la f.m.p. (resp. f.d.p.) $K_{x,h}$ sur $\mathbb{S}_{x,h}$. De plus, pour une f.m.p. (resp. f.d.p.), on a respectivement $\widehat{f_n}(x) \in [0,1]$ (resp. $\widehat{f_n}(x) > 0$) pour tout $x \in \mathbb{T}_1$ et

$$\int_{x\in\mathbb{T}_1} \widehat{f_n}(x)\nu(dx) = \Lambda_n, \tag{1.40}$$

où $\Lambda_n = \Lambda(n; h, \mathcal{K})$ est une constante positive et finie $\int_{\mathbb{T}_1} K_{x,h}(t)\nu(dx) < +\infty$ pour tout $t \in \mathbb{T}_1$, et ν est une mesure de comptage ou de Lebesgue sur $\mathbb{T}_1$.

La proposition suivante rappelle quelques propriétés de l'estimateur à noyau associé (1.39).

Proposition 1.1.9 *(Kokonendji & Senga Kiéssé, 2011 ; Libengué, 2013) Soit une cible $x \in \mathbb{T}_1$ et une fenêtre de lissage $h \equiv h_n$. Dans le cas continu, si f est de classe $\mathscr{C}^2(\mathbb{T}_1)$, alors*

$$\text{Biais}\left\{\widehat{f_n}(x)\right\} = A(x,h)f'(x) + \frac{1}{2}\left\{A^2(x,h) + B(x,h)\right\}f''(x) + o(h^2). \tag{1.41}$$

On obtient des expressions similaires du biais (1.41) *dans le cas discret, sauf que f' et f'' sont respectivement les différences finies du premier et du second ordre.*

De plus, dans le cas continu, si f est bornée sur $\mathbb{T}_1$ alors il existe $r_2 = r_2(K_{x,h}) > 0$ le plus grand réel tel que $\left\|K_{x,h}\right\|_2^2 := \int_{\mathbb{S}_{x,h}} K_{x,h}^2(u)du \leq c_2(x)h^{-r_2}$, $0 \leq c_2(x) \leq +\infty$ et

$$\text{Var}\left\{\widehat{f_n}(x)\right\} = \frac{1}{n}f(x)\left\|K_{x,h}\right\|_2^2 + o\left(\frac{1}{nh^{r_2}}\right). \tag{1.42}$$

Dans le cas discret, le résultat (1.42) *devient*

$$\text{Var}\left\{\widehat{f_n}(x)\right\} = \frac{1}{n}f(x)[\{Pr(\mathcal{Z}_{x,h} = x)\}^2 - f(x)],$$

où $\mathcal{Z}_{x,h}$ est la variable aléatoire discrète de f.m.p. $K_{x,h}$.

Exemples de noyaux associés univariés

Les Tables 1.2 et 1.3 suivantes présentent quelques noyaux associés univariés (discrets et continus) de la littérature. On trouvera les versions détaillées de ces noyaux ainsi que leurs applications dans les travaux de Senga Kiessé (2008), Kokonendji & Senga Kiessé (2011), Libengué (2013) et Wansouwé *et al.* (2015d).

Dans la Table 1.2, il faut noter que le noyau gaussien inverse réciproque utilise la notation

$$\zeta(x,h) = (x^2 + xh)^{1/2}$$

et que celle du noyau gaussien inverse est

$$\eta(x,h) = (1 - 3xh)^{1/2}.$$

On pourra se référer à la Partie (b) de la Figure 5.1 pour une illustration graphique de ces noyaux associés continus.

Noyau associé	$K_{x,h}(u)$	$\mathbb{S}_{x,h}$
Bêta étendu	$\frac{(u-a)^{(x-a)/\{(b-a)h\}}(b-u)^{(b-x)/\{(b-a)h\}}}{(b-a)^{1+h^{-1}}B(1+(x-a)/(b-a)h,1+(b-x)/(b-a)h)}$	$[a,b]$
Gamma	$\frac{u^{x/h}}{\Gamma(1+x/h)\,h^{1+x/h}}\exp\left(-\frac{u}{h}\right)$	$[0,+\infty[$
Gaussien inverse réciproque	$\frac{1}{\sqrt{2\pi hu}}\exp\left\{-\frac{\zeta(x,h)}{2h}\left(\frac{u}{\zeta(x,h)}-2+\frac{\zeta(x,h)}{u}\right)\right\}$	$]0,+\infty[$
Lognormal	$\frac{1}{uh\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{1}{h}\log(\frac{u}{x})-h\right)^2\right\}$	$[0,+\infty[$
Gaussien	$\frac{1}{h\sqrt{2\pi}}\exp\left\{\frac{1}{2}\left(\frac{u-x}{h}\right)^2\right\}$	$\mathbb{R}$
Gamma inverse	$\frac{h^{1-1/(xh)}}{\Gamma(-1+1/(xh))}u^{-1/(xh)}\exp\left(-\frac{1}{hu}\right)$	$]0,+\infty[$
Gaussien inverse	$\frac{1}{\sqrt{2\pi hu}}\exp\left\{-\frac{\eta(x,h)}{2h}\left(\frac{u}{\eta(x,h)}-2+\frac{\eta(x,h)}{u}\right)\right\}$	$]0,+\infty[$

Table 1.2 – Quelques noyaux associés continus en univarié (e.g., Libengué 2013)

Dans la Table 1.3, la constante $P(a,h)$ de la est la constante de normalisation du noyau triangulaire discret ; elle est explicitement donnée par :

$$P(a,h) = (2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^a, \quad \forall a \in \mathbb{N}.$$

Le noyau catégoriel DiracDU est celui d'Aichison & Aitken (1976) tandis que celui de Wang-VanRyzin est tiré de Wang & Van Ryzin (1981). On se référera, par exemple, à la Figure 1.3 pour quelques représentations graphiques de ces noyaux associés discrets.

Noyau associé	$K_{x,h}(u)$	$\mathbb{S}_{x,h}$
Dirac	$\mathbb{1}_{\{u=x\}}$	$\mathbb{N}$
DiracDU	$(1-h)\,\mathbb{1}_{\{u=x\}} + \dfrac{h}{c-1}\mathbb{1}_{\{u\neq x\}}$	$\{0,1,...,c-1\}$
Binomial	$\dfrac{(x+1)!}{u!(x+1-u)}\left(\dfrac{x+h}{x+1}\right)^{u}\left(\dfrac{1-h}{x+1}\right)^{x+1-u}$	$\{0,1,...,x+1\}$
Binomial négatif	$\dfrac{(x+u)!}{u!x!}\left(\dfrac{x+h}{2x+1+h}\right)^{u}\left(\dfrac{x+1}{2x+1+h}\right)^{x+1}$	$\mathbb{N}$
Poisson	$\dfrac{(x+h)^{u}e^{-(x+h)}}{u!}$	$\mathbb{N}$
Triangulaire discret	$\dfrac{(a+1)^{h}-\lvert u-x\rvert^{h}}{P(a,h)}$	$\{x, x\pm 1,...,x\pm a\}$
Wang-VanRyzin	$(1-h)\,\mathbb{1}_{\{u=x\}} + \dfrac{1}{2}(1-h)\,h^{\lvert u-x\rvert}\mathbb{1}_{\{\lvert u-x\rvert\geqslant 1\}}$	$\mathbb{Z}$

TABLE 1.3 – Quelques noyaux associés discrets en univarié (e.g., Kokonendji & Senga Kiéssé 2011)

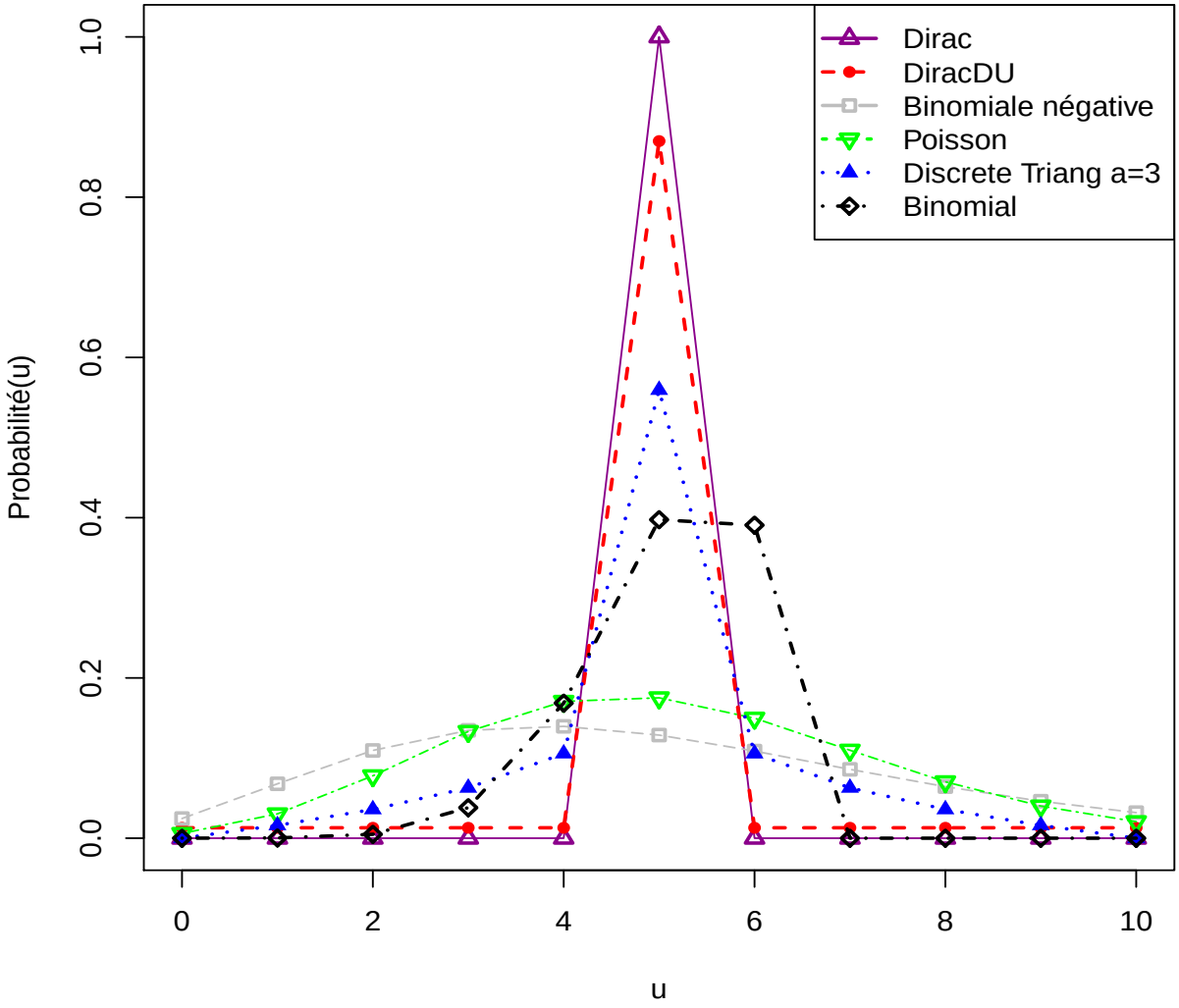


FIGURE 1.3 – Formes de quelques noyaux associés univariés discrets : Dirac, DiracDU, binomiale négative, Poisson, triangulaire discret $a = 3$ et binomial de même cible $x = 5$ et fenêtre de lissage $h = 0.13$.

Un exemple, dans le cas continu, de lissage de données de "Old Faithful data" (Azzalini & Bowman, 1990) est fourni pour quelques noyaux associés de Table 1.2 par la Figure 5.2.

Choix de paramètre de lissage

Dans les travaux de Senga Kiessé (2008), Kokonendji & Senga Kiessé (2011) et Libengué (2013), les auteurs ont proposé deux méthodes pour le choix du paramètre de lissage dans le cas discret. Il s'agit des méthodes de validation croisée par les moindres

carrés et l'excès de zéro. Ces méthodes seront brièvement résumées dans les paragraphes suivants. De plus, on présente aussi les autres méthodes bayésiennes pour le choix de paramètres de lissage proposées par Zougab *et al.* (2012, 2013abc) et aussi par Kuruwita *et al.* (2010). Il est à noter que la méthode validation croisée et la méthode bayésienne adaptative sont les mêmes que celles données en Section 1.1.1 pour des noyaux classiques.

Excès de zéros

Considérons un échantillon $\underline{X} = (X_1, \ldots, X_n)$. Le choix de fenêtre repose ici sur une particularité des données de comptage (pour $\mathbb{T}_1 = \mathbb{N}$) qui est l'excès de zéros dans l'échantillon $\underline{X}$. A partir de l'expression

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \sum_{i=1}^{n} \mathbb{P}(\mathcal{Z}_{x,h} = y) f(y)$$

on identifie le nombre de zéros théoriques au nombre de zéros empirique n_0 en prenant $y = 0$ et $f(0) = 1$. Ainsi, une fenêtre adaptée $h_0 = h_0(n, K_\theta)$ de h revient à résoudre

$$\sum_{i=1}^{n} \mathbb{P}(\mathcal{Z}_{X_i,h_0} = 0) = n_0, \tag{1.43}$$

où n_0 est le nombre de zéros dans l'échantillon. Ce choix h_0 pour les noyaux associés discrets est en pratique comparable à la validation croisée. Pour plus de détails et des exemples, on pourra se référer à Kokonendji *et al.* (2007) de manière générale, et aussi à Marsh & Mukhopadhyay (1999) dans le cadre d'un modèle poissonien.

Approche bayésienne globale

Cette approche a été utilisée par Kuruwita *et al.* (2010) pour l'estimation de f.d.p. par le noyau lognormal puis par Zougab *et al.* (2013b) pour l'estimation d'une f.m.p. par noyaux associés discrets. L'estimateur $\widehat{f_n}$ à noyaux associés de f est celui de (1.39). L'idée consiste à considérer le paramètre de lissage h comme une variable aléatoire. Ensuite, on choisit une loi *a priori* $\pi(\cdot)$ pour h. L'information apportée par toutes les observations X_i pour ce h est obtenue à travers une estimation par vraisemblance conditionnelle de $X_1, \ldots, X_n$ sachant h :

$$L(X_1, \ldots, X_n|h) = \prod_{i=1}^{n} \left(\frac{1}{n-1} \sum_{j=1, j\neq i}^{n} K_{X_i,h}(X_j) \right), \tag{1.44}$$

où $[1/(n-1)] \sum_{j=1, j\neq i}^{n} K_{X_i,h}(X_j)$ est l'estimation par validation croisée de $f(X_i)$.

Ensuite, on détermine par le théorème de Bayes la loi *a posteriori* de h définie par

$$\pi(h|X_1, \ldots, X_n) = \frac{\pi(h) L(X_1, \ldots, X_n|h)}{cte(X_1, \ldots, X_n)} = \frac{\pi(h) \prod_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{X_i,h}(X_j)}{\int \pi(h) \prod_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{X_i,h}(X_j) dh}. \tag{1.45}$$

Pour l'estimateur bayésien global de h sous perte quadratique, le meilleur h est la moyenne de $\pi(h|X_1,\ldots,X_n)$ donnée par

$$\widehat{h}_{GBay} = \int h\,\pi(h|X_1,\ldots,X_n)dh. \tag{1.46}$$

L'estimateur $\widehat{f_n}$ de (1.39) utilisant la fenêtre de lissage bayésienne globale est alors

$$\widehat{f_n}(x) \;\; = \frac{1}{n}\sum_{i=1}^{n} K_{x,\widehat{h}_{GBay}}(X_i), \quad \forall x \in \mathbb{T}_1.$$

Dans la pratique, l'estimateur $\widehat{h}_{GBay}$ n'a pas toujours une forme explicite. Il est alors estimé en choisissant une loi *a priori* $\pi(h)$ de h de sorte que les méthodes de Monté Carlo pour les Chaînes de Markov (MCMC) puissent fonctionner. Par exemple, Zougab *et al.* (2013b) a considéré $\pi(h)$ proportionnelle à $1/(1+h^2)$ alors que Brewer (1998) a proposé la loi gamma. L'algorithme à marche aléatoire des MCMC suit les quatre étapes suivantes :

Algorithme 1.1.10 ***Etape 1.*** *Initialisation de* $h^{(0)}$

Etape 2. *Pour* $t \in \{1,\ldots,N\}$ *:*
a) Générer $\tilde{h} \sim \mathcal{N}(h^{(t)},\tau)$
b) Calculer la probabilité d'acceptation $\rho = \min\left\{1, \dfrac{\pi(\tilde{h}|X_1,\ldots,X_n)}{\pi(h^{(t)}|X_1,\ldots,X_n)}\right\}$ *puis considérer*

$$h^{(t+1)} = \begin{cases} \tilde{h} & si\ u < \rho, \quad u \sim \mathcal{U}_{[0,1]} \\ h^{(t)} & sinon. \end{cases}$$

Etape 3. $t = t+1$ *et aller à* ***Etape 2.***

Etape 4. *Calculer l'estimateur de Bayes :* $\widehat{h} = \dfrac{1}{N-N_0}\displaystyle\sum_{t=N_0+1}^{N} h^{(t)}$*, où* N_0 *est le temps de chauffe.*

Approche bayésienne locale

L'objectif de cette approche est d'estimer le paramètre de lissage $h \equiv h(x)$ en chaque point x du support $\mathbb{T}_1$ d'une f.m.p. f caractérisant une suite de variables aléatoires i.i.d. $X_1, X_2, \ldots, X_n$. On utilise l'estimateur $\widehat{f_n}$ à noyaux associés de f présenté en (1.39). Afin d'estimer localement h, on le considère comme une variable aléatoire. Soit $\pi(h)$ la loi *a priori* de $h \equiv h(x)$. Par la formule de Bayes, la loi *a posteriori* de h au point x prend la forme suivante

$$\pi(h|x) = \frac{f_h(x)\pi(h)}{\int f_h(x)\pi(h)dh}. \tag{1.47}$$

Comme le modèle f_h est inconnu, on le remplace par son estimateur à noyau associé discret $\widehat{f_h}$ de (1.39). Ainsi, une estimation ou approximation de la loi *a posteriori* de $\pi(h|x)$ devient

$$\widehat{\pi}(h|x;X_1,X_2,\ldots,X_n) = \frac{\widehat{f_h}(x)\pi(h)}{\int \widehat{f_h}(x)\pi(h)dh}. \tag{1.48}$$

En fait, $\widehat{f_h}(x)$ représente l'information nécessaire sur $h \equiv h(x)$ pour compléter la loi *a priori* $\pi(\cdot)$. L'utilisation de l'estimateur de Bayes sous la perte quadratique conduit au meilleur estimateur $\widehat{h_n}(x)$ de h au point x par

$$\widehat{h_n}(x) = \int h\,\widehat{\pi}(h|x; X_1, X_2, \ldots, X_n)dh. \tag{1.49}$$

L'estimateur à noyaux associé avec fenêtre de lissage bayésienne locale est alors

$$\widehat{f_n}(x) \quad = \frac{1}{n}\sum_{i=1}^{n} K_{x,\widehat{h_n}(x)}(X_i), \quad \forall x \in \mathbb{T}_1. \tag{1.50}$$

Lorsque la moyenne *a posteriori* (1.49) ne s'obtient pas explicitement, on peut utiliser les MCMC de l'Algorithme 1.1.10 de l'approche bayésienne globale précédente.

Résolution de problèmes de biais de bordure

Tous les noyaux asymétriques des Tables 1.3 et 1.2 changent de forme en fonction de la cible x et n'ont apparemment pas de problèmes aux bords. Cependant, le biais (1.41) dans les cas discret et continu est différent de celui des noyaux classiques (1.7) obtenu en (1.13). On pourra se référer dans le cas continu à Malec & Schienle (2014), Hirukawa & Sakudo (2014) et Igarashi & Kakizawa (2015) qui ont traité ce problème de biais de bords pour certains noyaux asymétriques.

Dans ce qui suit, on décrit l'algorithme de réduction du biais pour tout noyau associé univarié continu de Libengué (2013), ainsi que la méthode de résolution du biais de bordure dans le cas du noyau triangulaire discret (voir Kokonendji & Zocchi, 2010).

Cas continu : réduction du biais de l'estimateur

Dans le biais (1.41) de l'estimateur $\widehat{f_n}(x)$ à noyau associé, on remarquera la présence du terme de gradient $A(x,h)f'(x)$ qui augmente ce biais. S'inspirant des travaux de Zhang (2010) et Zhang & Karunamuni (2010), Libengué (2013) a proposé un algorithme pour éliminer le terme $A(x,h)f'(x)$ dans la majeure partie du support continu $\mathbb{T}_1$. La première étape consiste à définir les régions intérieures et de bords ; tandis que la seconde traite le noyau associé modifié qui fournit un biais réduit.

1. On divise le support $\mathbb{T}_1 = [t_1, t_2]$ en deux régions d'ordre $\alpha(h) > 0$ avec $\alpha(h) \to 0$ quand $h \to 0$;
   (i) *région intérieure* (la plus grande région pouvant contenir 95% des observations) notée $\mathbb{T}_{\alpha(h),0}$ et définie par l'intervalle
   $$\mathbb{T}_{\alpha(h),0} =]t_1 + \alpha(h), t_2 - \alpha(h)[,$$
   (ii) *régions de bords* (pouvant être vide) représentées par les deux intervalles $\mathbb{T}_{\alpha(h),-1}$ et $\mathbb{T}_{\alpha(h),+1}$ respectivement définis par
   $$\mathbb{T}_{\alpha(h),-1} = [t_1, t_1 + \alpha(h)] \text{ (à gauche)},$$
   et
   $$\mathbb{T}_{\alpha(h),+1} = [t_2 - \alpha(h), t_2] \text{ (à droite)}.$$

Les régions de bords constituent le complémentaire de la région intérieure et on note :

$$\mathbb{T}^{c}_{\alpha(h),0} = \mathbb{T}_{\alpha(h),-1} \cup \mathbb{T}_{\alpha(h),+1}.$$

2. On modifie le noyau associé $K_\theta(x,h)$ correspondant à $A(x,h)$ et $B(x,h)$ en une nouvelle fonction noyau $K_{\widetilde{\theta}(x,h)}$ correspondant à $\widetilde{A}(x,h) = \left(\widetilde{A}_{-1}(x,h), \widetilde{A}_0(x,h), \widetilde{A}_{+1}(x,h)\right)$ et $\widetilde{B}(x,h) = \left(\widetilde{B}_{-1}(x,h), \widetilde{B}_0(x,h), \widetilde{B}_{+1}(x,h)\right)$ de sorte que, pour un h fixé,

$$\widetilde{\theta}(x,h) = \left(\widetilde{\theta}_{-1}(x,h), \widetilde{\theta}_0(x,h), \widetilde{\theta}_{+1}(x,h)\right) = \begin{cases} \widetilde{\theta}_{-1}(x,h) & \text{si } x \in \mathbb{T}_{\alpha(h),-1} \\ \widetilde{\theta}_0(x,h) : \quad \widetilde{A}_0(x,h) = 0 & \text{si } x \in \mathbb{T}_{\alpha(h),0} \\ \widetilde{\theta}_{+1}(x,h) & \text{si } x \in \mathbb{T}_{\alpha(h),+1} \end{cases}$$

soit continue sur $\mathbb{T}_1$ et constant sur $\mathbb{T}^{c}_{\alpha(h),0}$.

Cas discret : biais de bords

Dans cette section, on rappelle la résolution de problèmes de bords dans l'estimation d'une f.m.p. par le noyau triangulaire général $\mathcal{DT}(m;a_1,a_2;h_1,h_2)$, et notée $g(\cdot;m,a_1,a_2,h_1,h_2)$. En considérant que $m \in \mathbb{Z}$ est son mode, a_1, $a_2 \in \mathbb{N}$ sont respectivement ses bras gauche et droit et h_1, $h_2 \in \mathbb{R}_+$ ses ordres à gauche et à droite, le support de g est

$$\mathbb{T}_{m,a_1,a_2} = \mathbb{T}^{*}_{a_1,m} \cup \mathbb{T}_{m,a_2} = \mathbb{T}_{a_1,m} \cup \mathbb{T}^{*}_{m,a_2} = \mathbb{T}^{*}_{a_1,m} \cup \{m\} \cup \mathbb{T}^{*}_{m,a_2}.$$

où

$$\begin{aligned} &\mathbb{T}_{a_1,m} = \{m-k\ ;\ k=0,1,\ldots,a_1\} \quad , \quad \mathbb{T}^{*}_{a_1,m} := \mathbb{T}_{a_1,m} \setminus \{m\}, \\ &\mathbb{T}_{m,a_2} = \{m+k\ ;\ k=0,1,2,\ldots,a_2\} \quad \text{et} \quad \mathbb{T}^{*}_{m,a_2} := \mathbb{T}_{m,a_2} \setminus \{m\}. \end{aligned}$$

L'expression explicite de $g(\cdot;m,a_1,a_2,h_1,h_2)$ est donnée par

$$g(y;m,a_1,a_2,h_1,h_2) = \frac{1}{D(a_1,a_2,h_1,h_2)} \left\{ \left[1 - \left(\frac{m-y}{a_1+1} \right)^{h_1} \right] \mathbf{1}_{\mathbb{T}^{*}_{a_1,m}}(y) + \left[1 - \left(\frac{y-m}{a_2+1} \right)^{h_2} \right] \mathbf{1}_{\mathbb{T}_{m,a_2}}(y) \right\},$$

avec

$$D(a_1,a_2,h_1,h_2) = (a_1+a_2+1) - (a_1+1)^{-h_1} \sum_{k=1}^{a_1} k^{h_1} - (a_2+1)^{-h_2} \sum_{k=1}^{a_2} k^{h_2}.$$

Il a été signalé dans Kokonendji & Zocchi (2010) que lorsque $h_1 = h_2 = h$ on a la loi triangulaire discret symétrique d'ordre h notée $\mathcal{DT}(m;a_1,a_2;h,h) = \mathcal{DT}(m;a_1,a_2;h)$. La Figure 1.4 illustre les différentes formes qu'on peut obtenir en modifiant les paramètres de $\mathcal{DT}(m;a_1,a_2;h_1,h_2)$.

Considérons maintenant $X_1,\ldots,X_n$ une suite de variables aléatoires discrètes i.i.d. de f.m.p. inconnue f et de support $\mathbb{T}_1 = \mathbb{N}$ ou $\mathbb{T}_1 = \{0,1,\ldots,N\}$ avec N un entier naturel non nul. Dans ces cas, l'estimateur à noyau triangulaire discret général produit des effets de bords lorsque $\bigcup_{x\in\mathbb{N}} \mathbb{T}_{x,a_1,a_2}$ est différent de $\mathbb{T}$ (i.e. $\bigcup_{x\in\mathbb{N}} \mathbb{T}_{x,a_1,a_2} \supsetneq \mathbb{T}_1$). Ainsi, pour un support $\mathbb{T}_1 = \mathbb{N}$ avec $a_1 \neq 0$ où $\bigcup_{x\in\mathbb{N}} \mathbb{T}_{x,a_1,a_2} = \{-a_1,\ldots,-1\} \cup \mathbb{N}$ la solution consiste à considérer un nouveau bras gauche a_0 obtenu à partir de a_1 :

$$a_0 = \begin{cases} x & \text{si} \quad x \in \{0,1,\ldots,a_1-1\} \\ a_1 & \text{si} \quad x \in \{a_1,a_1+1,\ldots,N,\ldots\}. \end{cases} \tag{1.51}$$

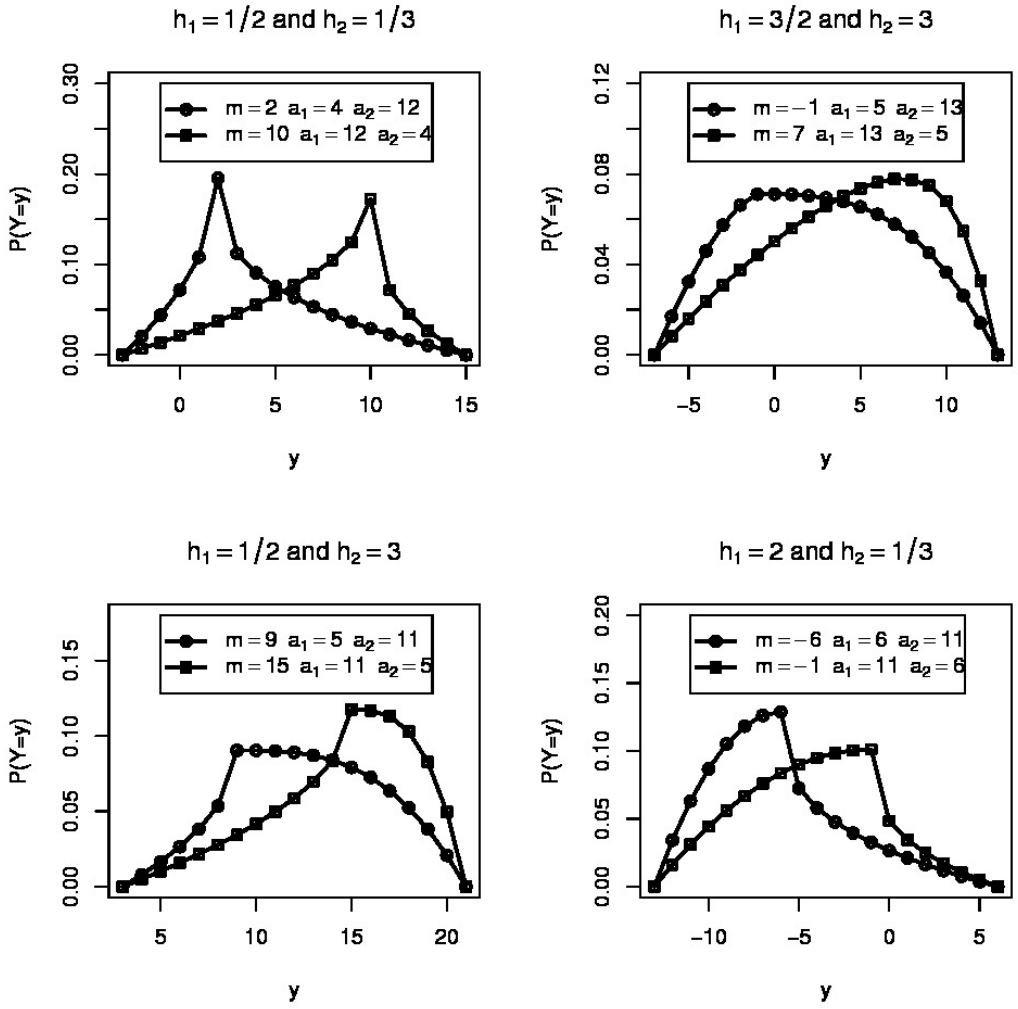


FIGURE 1.4 – Différentes formes du noyau triangulaire général discret (Kokonendji & Zocchi, 2010).

De même, losrque $\mathbb{T}_1 = \{0, 1, \ldots, N\}$, les effets de bords se traduisent par $\bigcup_{x \in \{0,1,\ldots,N\}} \mathbb{T}_{x,a_1,a_2} = \{-a_1, \ldots, -1\} \cup \{0, 1, \ldots, N\} \cup \{N+1, \ldots, N+a_2\}$. La solution consiste alors à appliquer (1.51) à droite de $\{0, 1, \ldots, N\}$ et à introduire un nouveau bras a_N obtenu à partir de a_2 :

$$a_N = \begin{cases} a_2 & \text{si} \quad x \in \{0, 1, \ldots, N - a_2\} \\ N - x & \text{si} \quad x \in \{N - a_2 + 1, \ldots, N - 1, N\}. \end{cases} \tag{1.52}$$

La combinaison de (1.51) et (1.52) fournit un noyau associé modifié et asymétrique qui est plus approprié pour l'estimation de toute f.m.p. compact (ou fonctions discrètes).

1.2 Estimation par noyaux de fonctions de régression multiple

Tout au long de cette section, on s'intéresse à la relation qui existe entre une variable réponse Y et un vecteur de variables explicatives $\mathbf{x}$ de dimension d ($d \geq 1$). Cette relation peut s'exprimer par

$$Y = m(\mathbf{x}) + \epsilon, \tag{1.53}$$

où m est la fonction de régression de $\mathbb{T}_d \subseteq \mathbb{R}^d$ sur $\mathbb{R}$ et ϵ le terme résiduel de moyenne nulle et variance finie. Soit une séquence de vecteurs aléatoires i.i.d. $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ sur $\mathbb{T}_d \times \mathbb{R}$ ($\subseteq \mathbb{R}^{d+1}$) avec $m(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ de (1.53).

Définition 1.2.1 *Soit* $\mathbf{H}$ *une matrice des fenêtres et* $\mathcal{K}$ *une fonction noyau classique vérifiant la Définition 1.1.1. L'estimateur* $\widehat{m}_n(\cdot)$ *de Nadaraya (1969) et Watson (1969) utilisant le noyau classique multivarié est défini en un point* $\mathbf{x} \in \mathbb{T}_d$ *par :*

$$\widehat{m}_n(\mathbf{x}; \mathcal{K}, \mathbf{H}) = \sum_{i=1}^{n} \frac{Y_i \mathcal{K}\left\{\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)\right\}}{\sum_{i=1}^{n} \mathcal{K}\left\{\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)\right\}} = \widehat{m}_n(\mathbf{x}; \mathbf{H}), \quad \forall \mathbf{x} \in \mathbb{T}_d := \mathbb{R}^d. \tag{1.54}$$

Dans le cas de noyau associé multiple (i.e. produit des noyaux associés univariés), l'estimateur de Nadaraya-Watson (1.54) s'écrit :

$$\widetilde{m}_n(\mathbf{x}) = \sum_{i=1}^{n} \frac{Y_i \prod_{j=1}^{d} K^{[j]}_{x_j, h_{jj}}(X_{ij})}{\sum_{i=1}^{n} \prod_{j=1}^{d} K^{[j]}_{x_j, h_{jj}}(X_{ij})}, \quad \forall \mathbf{x} = (x_1, \ldots, x_d)^\top \in \mathbb{T}_d := \times_{j=1}^{d} \mathbb{T}_1^{[j]}. \tag{1.55}$$

La version de (1.55) pour le noyau classique multiple s'obtient en considérant

$$K^{[j]}_{x_j, h_{jj}}(\cdot) = \frac{1}{h_{jj}} \mathcal{K}^{[j]}\left(\frac{\cdot - x_j}{h_{jj}}\right),$$

où $\mathcal{K}^{[j]}$ est le noyau classique univarié.

Choix de fenêtres

Méthode de validation croisée

Dans le cadre de la régression multiple par les noyaux classiques, la sélection de la matrice des fenêtres est faite par validation croisée. On pourra se référer par exemple à Hardle & Marron (1982) pour le cas univarié. Cette validation croisée vaut aussi pour les cas multiple (1.55) ainsi qu'univarié. Dans la pratique, pour un noyau classique multivarié, la matrice de lissage optimale s'obtient par

$$\widehat{\mathbf{H}}_{cv} = \arg \min_{\mathbf{H} \in \mathcal{H}} \mathrm{LSCV}_R(\mathbf{H})$$

avec

$$\text{LSCV}_R(\mathbf{H}) = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \widehat{m}_{-i}(\mathbf{X}_i;\mathbf{H})\right\}^2, \tag{1.56}$$

où $\widehat{m}_{-i}(\mathbf{X}_i;\mathbf{H})$ est obtenu comme $\widehat{m}_n$ de (1.54) en excluant l'observation $\mathbf{X}_i$ et, $\mathcal{H}$ est l'espace des matrices des fenêtres $\mathbf{H}$. On pourra se référer à Kokonendji *et al.* (2009) pour les cas discrets univariés et à Li & Racine (2007) pour le cas multiple composé d'univariés gaussiens et DiracDU.

Un exemple de lissage de fonctions de régressions univariés par noyaux associés continus est proposé en Figure 5.3 avec les coefficients de détermination correspondant en Table 5.4.

1.3 Contributions

Au vu des Sections 1.1 et 1.2 précédentes et des thèses soutenues en univarié de Senga Kiéssé (2008), Libengué (2013) et Zougab (2013), les contributions de ce travail ont été de plusieurs ordres. En multivarié, il s'agit d'introduire et d'étudier le noyau associé multivarié dans le cadre d'estimation de fonctions de densité, de l'analyse discriminante et enfin pour la régression multiple. La vulgarisation de la méthode des noyaux associés en univarié était aussi visé. En effet, on décortique ces résultats de la manière qui suit.

Estimation de densités par noyaux associés multivariés

L'objet de ce travail est de proposer la méthode des noyaux associés multivariés pour l'estimation des densités. Il généralise les travaux de Bouerzmarni & Rombouts (2010) pour les noyaux multiple bêta et gamma et donc de matrices de lissages diagonales. Il est bien connu, de Duong (2004) entre autre, que les matrices de lissages pleines permettent de tirer tout le potentiel de l'estimation à noyau de densités multivariés. Ce travail est aussi une extension au multivarié des noyaux associés de Libengué (2013) pour le cas continu. Le support $\mathbb{T}_d$ des observations sera ici

$$\mathbb{T}_d = \mathbb{R}^{d_\infty} \times [z, \infty[^{d_z} \times [u, w]^{d_{uw}}$$

pour tous réels $u < w$ et z avec des valeurs positives de d_∞, d_z et d_{uw} dans $\{0, 1, \ldots, d\}$ et telle que $d = d_\infty + d_z + d_{uw}$.

D'abord, une définition des noyaux associés multivariés à matrice de lissage pleine est présentée. Une méthode dite *mode dispersion multivarié* est introduite et permet de prendre en compte toute les formes particulières de la matrice de lissage telles que diagonale, Scott et pleine. La matrice des fenêtres de Scott est $\mathbf{H} = h\mathbf{H}_0$ avec $h > 0$ et $\mathbf{H}_0$ une matrice de lissage fixé. On peut se référer à Scott (1992) pour les noyaux classiques. Un algorithme de réduction du biais de l'estimateur à noyau associé multivarié est alors présenté. Un exemple utilisant le type de noyau bêta bivarié avec structure de corrélation de Sarmanov (1966) est donné en Section 2.3. Cette distribution de Sarmanov a été aussi proposée indépendamment par Cohen (1984) dans la littérature physique,

et utilisé par Hernández-Bastida & Fernández-Sanchez (2012) pour l'évaluation de primes. On peut se référer aussi à Lee (1996) pour d'autres types de distributions multivariés utilisant cette technique d'introduction de la structure de corrélation. De même, on pourra consulter Kotz (2000) et Kundu *et al.* (2010) pour d'autres densités multivariés et aussi Johnson *et al.* (1997) pour les cas discrets.

Des simulations et applications pour la version non corrigée de l'estimateur à noyau bêta bivarié avec structure de corrélation sont menées avec un choix de matrice des fenêtres par validation croisée. Les simulations ne sont pas faites ici pour la version modifié compte tenu du temps de calcul non négligeable. Les résultats confirment que le noyau beta bivarié avec structure de corrélation est comparable au cas sans corrélation pour des formes simples de densités à estimer et meilleur pour des cas plus complexes. Le cas des fenêtres de Scott est le plus rapide et permet de se faire une idée de la forme de la densité à estimer.

Comme en (5.6) pour le cas univarié, on montrera, en théorie et en pratique, que la masse totale de l'estimateur à noyau associé multivarié oscille autour de l'unité. Tous ces résultats seront détaillées au Chapitre 2 ainsi qu'en Section 6.1. La Section 6.2 présente une sélection de matrices des fenêtres optimale automatique par la méthode bayésienne adaptative pour l'estimateur à noyau gamma multiple.

Classification supervisée par noyaux associés multiples

Il s'agit, dans ce Chapitre 3, d'utiliser les noyaux associés dans le contexte de l'analyse discriminante. Ces noyaux associés sont connus pour respecter la structure topologique des supports des variables d'intérets et comble ainsi un manque de la méthode à noyaux classiques. Leurs effets ont été montrés dans Kokonendji & Somé (2015) ou Chapitre 2 pour les densités multivariés et par Somé & Kokonendji (2015) ou Chapitre 4 dans le cadre de régression multiple à variables continues et discrètes.

Les matrices de lissages pleines et diagonales ayant les mêmes performances dans certaines situations dans ces deux travaux, on ne s'intéressera donc qu'aux noyaux associés multiples et appropriés pour les supports de variables à étudier. Ces variables sont continues (positives) et discrètes (comptages et catégorielles). La méthode d'analyse discriminante à noyaux associés multivariés est alors présnté, comme en (1.31) pour le noyau classique multivarié, et on mesure l'efficacité de la procédure par le taux d'erreur de classification donnée en (1.32). Les choix des matrices de lissage optimales se font par validation croisée pour des données de même nature. On note que, pour les jeux de données contenant des variables discrètes et continues, l'algorithme de la validation croisée précédente ne converge pas. Une version profilée, en dimension 2, prenant en compte directement le taux d'erreur de classification a été introduite et utilisée.

Les simulations et applications montrent l'efficaté et la pertinence de l'utilisation des noyaux associés en analyse discriminante. Aussi, bien qu' efficace, la nouvelle validation profilée mériterait d'être améliorée avant de l'utiliser pour trois variables mixtes et plus.

Régression multiple par noyaux associés

Dans ce travail, on s'intéresse à la régression multiple en présence de variables catégorielles, de comptages et de variables continues (partiellement ou totalement bornées). Les méthodes jusqu'à présent utilisées tiennent peu ou pas compte de la structure topologique du support $\mathbb{T}_d$ des données multivariées. On peut citer les travaux de Li & Racine (2007) qui ont proposé l'utilisation de noyaux gaussiens univariés pour toute variable continue et DiracDU (voir Table 1.3) pour les variables catégorielles. Le cas des données de comptages n'a été pris en compte qu'en univarié avec Senga Kiessé (2008).

Pour régler ce problème et tenir compte aussi de la structure de corrélation existant entre les variables explicatives, on a proposé l'utilisation des noyaux associés. Avec l'estimateur de Nadaraya (1969) et Watson (1969) et le choix de la matrice de lissage optimale par validation croisée, plusieurs comportements des noyaux associés (avec ou sans structure de corrélation) doivent être signalées. Les simulations et applications ont montré qu'il y a un effet variable du choix du noyau associé multivarié sur la qualité de l'estimation de fonctions de régression. Ainsi, pour des noyaux associés continus avec ou sans structure de corrélation et des données fortement corrélées ou pas, cet effet est négligeable. Dans le cas discret ou mixte par contre, il y'a clairement un effet du choix de noyau associé multiple utilisé. Finalement, plus que la performance du noyau associé utilisé, c'est d'abord le choix approprié suivant le support $\mathbb{T}_d$ des observations qui importe le plus. Le détail de ces résultats est présentée en Section 4.3.1 du Chapitre 4.

Vulgarisation de la méthode des noyaux associés univariés

Enfin, les méthodes à noyaux classiques ont déjà bonne presse auprès des practiciens pour l'estimation de différentes fonctionnelles. On peut citer entre autre les packages `ks` de Duong (2007), `KernSmooth` de Wand & Ripley (2015) pour l'estimation de fonctions de densités multivariés, et `regpro` et `np` pour les fonctions de régression (semi et non-paramétrique) et les quantiles. Le package `ks` de Duong (2007) possède les méthodes de sélection de matrice des fenêtres optimales les plus rapides (efficaces). En effet, il utilise des matrices pilotes dans les différentes procédures à savoir *plug in* et validation croisée. Ce procédé n'est pas encore applicable au noyaux associés de Libengué (2013) pour le cas continu et de Senga Kiessé (2008) pour le cas discret. Ils sont connus pour bien respecter le support de la fonctionnelle à estimer et constituent donc une alternative au noyau gaussien lequel est idéal pour des supports continus et non bornés. Notons que les méthodes bayésiennes constitue une alternative, tout aussi automatique (rapide), aux techniques de Duong (2004). on pourra se référer à Zougab *et al.* (2012, 2013, 2014a) pour des approches bayésiennes locale et globale pour le noyau binomial.

Ainsi, le package `Ake`, dédié à l'estimation de fonctions de densité, de masse de probabilité et de régression, a été créé sous le logiciel `R` (`R` Development Core Team, 2015). Ce package est disponible sur le site "Comprehensible `R` Archive Network" (CRAN) via le lien : `http://cran.r-project.org/web/packages/Ake/`. Le package permet d'estimer certaines fonctionnelles par la méthode des noyaux associés. Il s'agit ici essentiellement de fonctions de densités, de masse de probabilité et de régression (continues et discrètes). Les noyaux associés les plus intéressants et adaptés aux sup-

ports (totalement et partiellement) bornés ou discret (comptage et catégoriel) de ces fonctionnelles sont proposés et permettent de

- calculer les constantes de normalisations dans le cas des fonctions de densités ou de masse de probabilité,
- calculer et donner les représentations des différentes estimations de fonctionnelles.

D'autre part, des méthodes de sélection de fenêtres optimales existantes (validation croisée et méthodes bayésiennes) ont été implémentées. Une approche bayésienne, dite adaptative, pour l'estimateur de densité à noyau gamma a été aussi proposée.

`Ake` est l'un des premiers packages `R` dédié à l'estimation de fonctions de densité, de masse de probabilité et de régression par les noyaux associés. Son implémentation permet de réaliser simplement des lissages par noyaux associés univariés et va permettre une large utilisation et diffusion de cette approche. Le package ainsi que son manuel d'utilsation sont données au Chapitre 5 et à la Section A5 de l'Annexe.

Chapitre 2

On multivariate associated kernels for smoothing general density functions

Abstract. Multivariate associated kernel estimators, which depend on both target point and bandwidth matrix, are appropriate for partially or totally bounded distributions and generalize the classical ones as Gaussian. Previous studies on multivariate associated kernels have been restricted to product of univariate associated kernels, also considered having diagonal bandwidth matrices. However, it is shown in classical cases that for certain forms of target density such as multimodal, the use of full bandwidth matrices offers the potential for significantly improved density estimation. In this paper, general associated kernel estimators with correlation structure are introduced. Properties of these estimators are presented ; in particular, the boundary bias is investigated. Then, the generalized bivariate beta kernels are handled with more details. The associated kernel with a correlation structure is built with a variant of the mode-dispersion method and two families of bandwidth matrices are discussed under the criterion of cross-validation. Several simulation studies are done. In the particular situation of bivariate beta kernels, it is therefore pointed out the very good performance of associated kernel estimators with correlation structure compared to the diagonal case. Finally, an illustration on real dataset of paired rates in a framework of political elections is presented.

2.1 Introduction

Nonparametric estimation of unknown densities on partially or totally bounded supports, with or without correlation in its multivariate components, is a recurrent practical problem. Because of symmetry, the multivariate classical or symmetric kernels, not depending on any parameter, are not appropriate for these densities. In fact, these estimators give weights outside the support causing a bias in boundary regions. In order to reduce the boundary problem with multivariate symmetric kernels as Gaussian, Sain (2002) and recently Zougab *et al.* (2014) have proposed adaptive full bandwidth matrix selection ; but the bias does not disappear completely. Chen (1999, 2000) is one of the pioneers who has proposed, in univariate continuous case, some asymmetric

kernels (i.e. beta and gamma) whose supports coincide with those of the densities to be estimated. Also recently, Libengué(2013) investigated several families of these univariate continuous kernels that he called univariate associated kernels ; see also Kokonendji *et al.* (2007), Kokonendji & Senga Kiéssé (2011), Zougab *et al.* (2012, 2013) for univariate discrete situations. This procedure cancels of course the boundary bias ; however, it creates a quantity in the bias of the estimator which needs reduction ; see, for instance, Malec & Schienle (2014), Hirukawa & Sakudo (2014) and Igarashi & Kakizawa (2015).

Several approaches on multivariate kernel estimation have been proposed for various processings. Garcìa-Portugués *et al.*(2013) used product of kernels for estimating the different nature of both directional and linear components of a random vector. A classical kernel density estimation on the rotation group appropriate for crystallographic texture analysis has been investigated by Hielscher (2013). Symmetric kernel smoothers with univariate local bandwidth have been studied by Garcìa-Portugués *et al.*(2013) for semiparametric mixed effect models. Girard *et al.* (2013) presented frontier estimation with classical kernel regression on high order moments. In discrete case, Aitchison & Aitken (1976) provided kernel estimators for binary data while Racine & Li (2004) proposed the product of them with classical continuous one for smoothing regression on both categorical and continuous data. In the same spirit of Racine & Li (2004), Bouerzmarni & Rombouts (2010) considered some products of different univariate associated kernels in continuous case ; i.e. the bandwidth matrix obtained is diagonal. In the classical kernels case, Chacon & Duong (2011) and Chacon *et al.* (2011) have shown the importance of full bandwidth matrices for certain target densities. See also Hazelton & Marshall (2009) for a support with arbitrary shape.

The main goal of this work is to introduce the multivariate associated kernels with the most general bandwidth matrix. In other words, the support of the suggested associated kernels coincides to the support of the densities to be estimated ; also, the full bandwidth matrices take into account different correlation structures in the sample. Note that, a full bandwidth matrix significantly improves some complex target densities (e.g. multimodal) ; see Sain (2002). In high dimensions, the computational choice of this full bandwidth matrix needs some special techniques. We can refer to Chacon & Duong (2010, 2011) and Chacon & Duong (2011) for classical (symmetric) kernels. For illustrations in the present paper, we then focus on a bivariate case as beta kernel with correlation structure introduced by Sarmanov (1966) ; see also Lee (1996). A motivation to investigate the smoothing of these densities on $[0,1] \times [0,1]$ comes from a joint distribution of two comparable proportions. Many datasets in $[0,1] \times [0,1]$ can be found in statistical problems, for example, for comparing two rates. We shall examine the theoretical bias reduction and practical performances of the full bandwidth selection and two others bandwidth matrix parametrization using least squares or unbiased cross validation ; see, e.g., Wand & Jones (1993).

The rest of the paper is organized as follows. Section 2.2 gives a complete definition of multivariate associated kernels which includes both the product and the classical symmetric ones. A method to construct any multivariate associated kernel from a parametric probability density function (pdf) is then provided. Some pointwise properties of the corresponding estimator are investigated, in particular the convergence in the sense of mean integrated squared error (MISE) and an algorithm of the bias reduction.

Section 2.3 provides a particular study of a bivariate beta kernel with a correlation structure introduced by Sarmanov (1966). Also, some algorithms for the choice of the optimal bandwidth matrix by unbiased cross validation method are presented. This is followed, in Section 2.4, by simulation studies and a real data analysis of electoral behaviour of a population with regard to a candidate. Especially, the role of forms of bandwidth matrices is explored in details. Section 2.5 concludes with summary and final remarks, while the proof of a proposition is deferred to the appendix of Section 2.6.

2.2 Multivariate associated kernel estimators

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent and identically distributed (iid) random vectors with an unknown density function f on $\mathbb{T}_d$, a subset of $\mathbb{R}^d$ ($d \geq 1$). As frequently observed in practice, the subset $\mathbb{T}_d$ might be unbounded, partially bounded or totally bounded as :

$$\mathbb{T}_d = \mathbb{R}^{d_\infty} \times [z, \infty)^{d_z} \times [u, w]^{d_{uw}} \tag{2.1}$$

for given reals $u < w$ and z with nonnegative values of d_∞, d_z and d_{uw} in $\{0, 1, \ldots, d\}$ such that $d = d_\infty + d_z + d_{uw}$. A multivariate associated kernel estimator $\widehat{f_n}$ of f is simply defined by

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i), \quad \forall \mathbf{x} \in \mathbb{T}_d \subseteq \mathbb{R}^d, \tag{2.2}$$

where $\mathbf{H}$ is a $d \times d$ bandwidth matrix (i.e. symmetric and positive definite) such that $\mathbf{H} \equiv \mathbf{H}_n \to \mathbf{0}_d$ (the $d \times d$ null matrix) as $n \to \infty$, and $K_{\mathbf{x},\mathbf{H}}(\cdot)$ is the so-called associated kernel, parametrized by $\mathbf{x}$ and $\mathbf{H}$, and precisely defined as follows.

Définition 2.2.1 *Let $\mathbb{T}_d \left(\subseteq \mathbb{R}^d\right)$ 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 parametrized pdf $K_{\mathbf{x},\mathbf{H}}(\cdot)$ of support $\mathbb{S}_{\mathbf{x},\mathbf{H}} \left(\subseteq \mathbb{R}^d\right)$ is called "multivariate (or general) associated kernel" if the following conditions are satisfied :*

$$\mathbf{x} \in \mathbb{S}_{\mathbf{x},\mathbf{H}}, \tag{2.3}$$

$$\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x} + \mathbf{a}(\mathbf{x}, \mathbf{H}), \tag{2.4}$$

$$\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{B}(\mathbf{x}, \mathbf{H}), \tag{2.5}$$

where $\mathcal{Z}_{\mathbf{x},\mathbf{H}}$ denotes the random vector with pdf $K_{\mathbf{x},\mathbf{H}}$ and both $\mathbf{a}(\mathbf{x}, \mathbf{H}) = (a_1(\mathbf{x}, \mathbf{H}), \ldots, a_d(\mathbf{x}, \mathbf{H}))^\top$ and $\mathbf{B}(\mathbf{x}, \mathbf{H}) = \left(b_{ij}(\mathbf{x}, \mathbf{H})\right)_{i,j=1,\ldots,d}$ tend, respectively, to the null vector $\mathbf{0}$ and the null matrix $\mathbf{0}_d$ as $\mathbf{H}$ goes to $\mathbf{0}_d$.

Remarque 2.2.2 *(i) The function $K_{\mathbf{x},\mathbf{H}}(\cdot)$ is not necessary symmetric and is intrinsically linked to $\mathbf{x}$ and $\mathbf{H}$.*

(ii) The support $\mathbb{S}_{\mathbf{x},\mathbf{H}}$ is not necessary symmetric around of $\mathbf{x}$; it can depend or not on $\mathbf{x}$ and $\mathbf{H}$.

(iii) The condition (2.3) can be viewed as $\cup_{\mathbf{x} \in \mathbb{T}_d} \mathbb{S}_{\mathbf{x},\mathbf{H}} \supseteq \mathbb{T}_d$ and it implies that the associated kernel takes into account the support $\mathbb{T}_d$ of the density f, to be estimated.

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

(v) Both conditions (2.4) and (2.5) indicate that the associated kernel is more and more concentrated around of $\mathbf{x}$ *as* $\mathbf{H}$ *goes to* $\mathbf{0}_d$*. This highlights the peculiarity of associated kernel which can change its shape according to the target position.*

(vi) The form of orientation of the kernel is controlled by the parametrization of bandwidth matrix $\mathbf{H}$*; i.e. a full bandwidth matrix allows any orientation of the kernel and therefore any correlation structure.*

The following two examples provide well-known and also interesting particular cases of multivariate associated kernel estimators. The first can be seen as an interpretation of associated kernels through symmetric kernels. The second deals on associated kernels without correlation structure.

Example 2.2.3 *(Classical kernels) The kernel estimator* $\widehat{f_n}$ *of the density* f*, appropriate for unbounded supports in particular* $\mathbb{R}^d$*, is usually defined by :*

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_i), \quad \forall \mathbf{x}\in\mathbb{T}_d=\mathbb{R}^d, \tag{2.6}$$

where $\mathbf{x}$ *is a target,* $\mathbf{H}$ *a bandwidth matrix and* $K_{\mathbf{H}}(\mathbf{y}) = (1/\det\mathbf{H})K\left(\mathbf{H}^{-1}\mathbf{y}\right)$ *or sometimes* $K_{\mathbf{H}}(\mathbf{y}) = (1/\det\mathbf{H})^{1/2}K\left(\mathbf{H}^{-1/2}\mathbf{y}\right)$ *for all* $\mathbf{y}\in\mathbb{R}^d$*. The function* K *is the multivariate kernel assumed to be spherically symmetric and it does not depend on any parameter in particular* $\mathbf{x}$ *and* $\mathbf{H}$*. The kernel function has also mean vector and covariance matrix respectively equal to zero and* $\mathbf{\Sigma}$*; in general, the covariance matrix is the identity matrix :* $\mathbf{\Sigma}=\mathbf{I}$*. This function* K *is here called classical kernel.*

The following result connects a classical kernel to its corresponding symmetric or classical (multivariate) associated kernel.

Proposition 2.2.4 *Let* $\mathbb{R}^d=\mathbb{T}_d$ *be the support of the density to be estimated. Let* K *be a classical kernel with support* $\mathbb{S}_d\subseteq\mathbb{R}^d$*, mean vector* $\mathbf{0}$ *and covariance matrix* $\mathbf{\Sigma}$*. Given a target vector* $\mathbf{x}\in\mathbb{R}^d$ *and a bandwidth matrix* $\mathbf{H}$*, then the classical kernel induces the so-called classical (multivariate) associated kernel : (i)*

$$K_{\mathbf{x},\mathbf{H}}(\cdot) = \frac{1}{\det\mathbf{H}}K\left\{\mathbf{H}^{-1}(\mathbf{x}-\cdot)\right\} \tag{2.7}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}}=\mathbf{x}-\mathbf{H}\mathbb{S}_d$ *with* $\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right)=\mathbf{x}$ *(i.e.* $\mathbf{a}(\mathbf{x},\mathbf{H})=\mathbf{0}$*) and* $\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right)=\mathbf{H}\mathbf{\Sigma}\mathbf{H}$*; (ii)*

$$K_{\mathbf{x},\mathbf{H}}(\cdot) = \frac{1}{(\det\mathbf{H})^{1/2}}K\left\{\mathbf{H}^{-1/2}(\mathbf{x}-\cdot)\right\}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}}=\mathbf{x}-\mathbf{H}^{1/2}\mathbb{S}_d$ *with* $\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right)=\mathbf{x}$ *(i.e.* $\mathbf{a}(\mathbf{x},\mathbf{H})=\mathbf{0}$*) and* $\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right)=\mathbf{H}^{1/2}\mathbf{\Sigma}\mathbf{H}^{1/2}$*.*

Proof. We only proof (i) because (ii) is similar. From (2.2) and (2.6) with $K_{\mathbf{H}}(\mathbf{y}) = (1/\det\mathbf{H})K\left(\mathbf{H}^{-1}\mathbf{y}\right)$, we easily deduce the expression (2.7). From (2.7) and Definition

2.2.1, for a fixed $\mathbf{x} \in \mathbb{T}_d = \mathbb{R}^d$ and for all $\mathbf{t} \in \mathbb{T}_d = \mathbb{R}^d$, there exists $\mathbf{u} \in \mathbb{S}_d$ such that $\mathbf{u} = \mathbf{H}^{-1}(\mathbf{x} - \mathbf{t})$ and therefore $\mathbf{t} = \mathbf{x} - \mathbf{H}\mathbf{u}$. This implies, from (2.3), that $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbf{x} - \mathbf{H}\mathbb{S}_d$. The last two results are simply derived from calculating the covariance matrix and the mean vector of $\mathcal{Z}_{\mathbf{x},\mathbf{H}}$ (the random vector of pdf $K_{\mathbf{x},\mathbf{H}}$) by making the previous substitution $\mathbf{u} = \mathbf{H}^{-1}(\mathbf{x} - \mathbf{t})$.■

It is known that the choice of classical kernels is not important ; nevertheless, the best classical kernel is the Epanechnikov (1969) one in the sense of MISE with bounded support $\mathbb{S}_d$. The most popular is the Gaussian kernel with $\mathbb{S}_d = \mathbb{R}^d = \mathbb{T}_d$, $\mathbf{\Sigma} = \mathbf{I}$ and therefore $\mathrm{Cov}(\mathcal{Z}_{\mathbf{x},\mathbf{H}}) = \mathbf{H}^2$; see Chacon & Duong (2010) and Zougab *et al.* (2014). An interpretation of any classical multivariate associated kernel $K_{\mathbf{x},\mathbf{H}}(\cdot)$ can be presented as follows : through the symmetry property of the classical kernel, the mean $\mathbb{E}(\mathcal{Z}_{\mathbf{x},\mathbf{H}}) = \mathbf{x}$ coincides with the mode which is the target $\mathbf{x}$; separately and in contrario to the general case (2.5), the dispersion measure around of the target $\mathbf{x}$, $\mathrm{Cov}(\mathcal{Z}_{\mathbf{x},\mathbf{H}}) = \mathbf{H}\mathbf{\Sigma}\mathbf{H}$ which does not here depend on $\mathbf{x}$, serves essentially to the smoothing parameters or to the bandwidth matrix. This is the basic concept of general associated kernels and it is a different approach with respect to the convolution point of view. Note that the bandwidth matrix is similar to the dispersion matrix, which is symmetric and positive definite ; see for instance Jørgensen (1997). For univariate dispersion parameter, we can refer to Jørgensen (1997) and Jørgensen & Kokonendji (2011) for different uses.

Example 2.2.5 *(Multiple kernels) The product kernel estimator introduced by Bouerzmarni & Rombouts (2010) can be defined as a product of univariate associated kernel estimators of Libengué (2013). We here call it "multiple associated kernel estimator" $\widehat{f_n}$ of the density f :*

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} K^{[j]}_{x_j,h_{jj}}(X_{ij}), \quad \forall x_j \in \mathbb{T}_1^{[j]} \subseteq \mathbb{R}, \tag{2.8}$$

where $\mathbb{T}_1^{[j]}$ is the support of univariate margin of f for $j = 1,\ldots,d$, $\mathbf{x} = (x_1,\ldots,x_d)^\top \in \times_{j=1}^{d}\mathbb{T}_1^{[j]}$, $\mathbf{X}_i = (X_{i1},\ldots,X_{id})^\top$ for $i = 1,\ldots,n$, and $h_{11},\ldots,h_{dd}$ are the univariate bandwidth parameters. The function $K^{[j]}_{x_j,h_{jj}}$ is the jth univariate associated kernel on the support $\mathbb{S}_{x_j,h_{jj}} \subseteq \mathbb{R}$. In principle, this estimator is more appropriate for bounded or partially bounded distributions without correlation in its components. A particular multiple associated kernel estimator is obtained by using univariate classical kernels.

In the following proposition, we point out that all multiple associated kernels are multivariate associated kernels without correlation structure in the bandwidth matrix.

Proposition 2.2.6 *Let $\times_{j=1}^{d}\mathbb{T}_1^{[j]} = \mathbb{T}_d$ be the support of the density f to be estimated with $\mathbb{T}_1^{[j]}(\subseteq \mathbb{R})$ the supports of univariate margins of f. Let $\mathbf{x} = (x_1,\ldots,x_d)^\top \in \times_{j=1}^{d}\mathbb{T}_1^{[j]}$ and $\mathbf{H} = \mathbf{Diag}(h_{11},\ldots,h_{dd})$ with $h_{jj} > 0$. Let $K^{[j]}_{x_j,h_{jj}}$ be a univariate associated kernel (see Definition 2.2.1 for $d = 1$) with its corresponding random variable $\mathcal{Z}^{[j]}_{x_j,h_{jj}}$ on $\mathbb{S}_{x_j,h_{jj}}(\subseteq \mathbb{R})$ for all $j = 1,\ldots,d$. Then, the multiple associated kernel is also a multivariate associated kernel :*

$$K_{\mathbf{x},\mathbf{H}}(\cdot) = \prod_{j=1}^{d} K^{[j]}_{x_j,h_{jj}}(\cdot) \tag{2.9}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \times_{j=1}^{d} \mathbb{S}_{x_j,h_{jj}}$ *with* $\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = (x_1 + a_1(x_1,h_{11}),\ldots,x_d + a_d(x_d,h_{dd}))^\top$ *and* $\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{Diag}\left(b_{jj}(x_j,h_{jj})\right)_{j=1,\ldots,d}$. *In other words, the random variables* $\mathcal{Z}_{x_j,h_{jj}}^{[j]}$ *are independent components of the random vector* $\mathcal{Z}_{\mathbf{x},\mathbf{H}}$.

Proof. From (2.2) and (2.8), the expression (2.9) is easily deduced. The remainder results are obtained directly by calculating the mean vector and covariance matrix of $(\mathcal{Z}_{x_1,h_{11}}^{[1]},\ldots,\mathcal{Z}_{x_d,h_{dd}}^{[d]})^\top = \mathcal{Z}_{\mathbf{x},\mathbf{H}}$ which is the random vector of the pdf (2.9).■

The multiple associated kernels have been illustrated in bivariate case by Bouerzmarni & Rombouts (2010). For simulation studies, the authors used two independent univariate beta kernels and also two independent univariate gamma kernels. It is easy to generalize the investigation from two to more independent univariate associated kernels.

If we have an associated kernel, an estimator can be easily deduced as in (2.2). Otherwise, a construction of associated kernels is possible by using an appropriate pdf. The pdf used must have at least the same numbers of parameters than the number of components in the couple $(\mathbf{x},\mathbf{H})$ as parameters of the expected associated kernel. The rest of this section is devoted to a construction of the multivariate associated kernels and then to some properties of the corresponding estimators.

2.2.1 Construction of general associated kernels

In order to build a multivariate associated kernel $K_{\mathbf{x},\mathbf{H}}(\cdot)$, we have to evaluate the dimensions of $\mathbf{x}$ and $\mathbf{H}$. We always have d components for the target vector $\mathbf{x} \in \mathbb{T}_d$ which is completely separate from the bandwidth matrix $\mathbf{H}$ in the classical multivariate associated kernel ; but, in general, $\mathbf{x}$ is intrinsically linked to $\mathbf{H}$.

$\mathbf{H}$	General	Multiple	Classical
Full	$d(d+3)/2$	$2d$	$d(d+3)/2$
Scott	$d+1$	$d+1$	$d+1$
Diagonal	$2d$	$2d$	$2d$

TABLE 2.1 – Numbers k_d of parameters according to the form of bandwidth matrices for general, multiple and classical asociated kernels.

Table 2.1 gives the exact or minimal numbers k_d ($> d$) of parameters in $(\mathbf{x},\mathbf{H})$ according to different forms of $\mathbf{H}$. The bandwidth matrix $\mathbf{H} = (h_{ij})_{i,j=1,\ldots,d}$ is said *full* (i.e. with complete structure of correlations) and admits $d(d+1)/2$ parameters. It is said *diagonal* if $\mathbf{H} = \mathbf{Diag}(h_{11},\ldots,h_{dd})$, i.e. without correlation, and has only d parameters. We denote by the *Scott* bandwidth matrix the form $\mathbf{H} = h\mathbf{H}_0$ with only one parameter $h > 0$ and fixed $\mathbf{H}_0 = (h_{ij,0})_{i,j=1,\ldots,d}$; see Scott (1992, page 154). In practice, the matrix $\mathbf{H}_0$ can be fixed empirically from the data. Although k_d is the same for classical and general associated kernels in Table 2.1, the differences arise because of separation or not between $\mathbf{x}$ and $\mathbf{H}$ and, also, the presented k_d are exact numbers for classical and minimal numbers for

both general and multiple associated kernels. It becomes clear that any pdf, with at least k_d parameters and having a unique mode and a dispersion matrix, can lead to an associated kernel. Now, we introduce the notion of *type of kernel* which is necessary for the construction from a given pdf.

Définition 2.2.7 *A type of kernel K_θ is a parametrized pdf with support $\mathbb{S}_\theta \subseteq \mathbb{R}^d$, $\theta \in \Theta \subseteq \mathbb{R}^{k_d}$, such that K_θ is squared integrable, unimodal with mode $\mathbf{m} \in \mathbb{R}^d$ and admitting a $d \times d$ dispersion matrix $\mathbf{D}$ (which is symmetric and positive definite) ; i.e. $\theta = \theta(\mathbf{m}, \mathbf{D})$ a $k_d \times 1$ vector with k_d given in Table 2.1 and the d first coordinates of θ corresponds to those of $\mathbf{m}$.*

Let us denote by $\mathbf{D}_{\mathbf{x_0}} = \int_{\mathbb{R}^d} (\mathbf{x} - \mathbf{x_0})(\mathbf{x} - \mathbf{x_0})^\top K_\theta(d\mathbf{x})$ the dispersion matrix of K_θ around the fixed vector $\mathbf{x_0}$. Here is a series of facts to have in mind.

Lemme 2.2.8 *Let K_θ be a type of kernel on $\mathbb{S}_\theta \subseteq \mathbb{R}^d$. The three following assertions are satisfied :*

(i) the mode vector $\mathbf{m}$ of K_θ always belongs in $\mathbb{S}_\theta$;

(ii) $K_\theta(\mathbf{m}) \geq K_\theta(\boldsymbol{\mu})$ where $\boldsymbol{\mu}$ is the mean vector of K_θ ;

(iii) if $\mathbf{D}_\mathbf{m}$ tends to the null matrix, then $\mathbf{D}_\mu$ also goes to the null matrix.

Proof. (i) and (ii) are trivial. As for (iii), it is easy to check that $\mathbf{D}_\mathbf{m} = \mathbf{D}_\mu + (\boldsymbol{\mu} - \mathbf{m})(\boldsymbol{\mu} - \mathbf{m})^\top$. Thus, $\mathbf{D}_\mathbf{m}$ tends to the null matrix means K_θ goes to the Dirac probability in the sense of distribution ; then, $\boldsymbol{\mu} - \mathbf{m}$ goes to the null vector and therefore $\mathbf{D}_\mu$ also goes to the null matrix.■

Without loss of generality, we only present a construction of general (or multivariate) associated kernels excluding both multiple and classical ones. In fact, Libengué (2013) built some univariate associated kernels that can be used in multiple associated kernels. For this, he proposed a "mode-dispersion principle" saying that it must put the mode on the target and the dispersion parameter on the bandwidth. In the same spirit, we here propose a construction of general associated kernels using the *multivariate mode-dispersion* method given in (2.11) below.

Indeed, since $\theta = \theta(\mathbf{m}, \mathbf{D})$ is a $k_d \times 1$ vector, we must vectorize the couple $(\mathbf{x}, \mathbf{H})$ where $\mathbf{x} \in \mathbb{T}_d$ is the target and $\mathbf{H} = (h_{ij})_{i,j=1,\ldots,d}$ is the bandwidth matrix. According to the symmetry of $\mathbf{H}$, we use the so-called half vectorization of $(\mathbf{x}, \mathbf{H})$. That is defined as $\operatorname{vech}(\mathbf{x}, \mathbf{H}) = (\mathbf{x}, \operatorname{vech} \mathbf{H})$, where

$$\operatorname{vech} \mathbf{H} = \left(h_{11}, \ldots, h_{1d}, h_{22}, \ldots, h_{2d}, \ldots, h_{(d-1)(d-1)}, h_{(d-1)d}, h_{dd}\right)^\top \tag{2.10}$$

is a $[d(d+1)/2] \times 1$ vector obtained by stacking the columns of the lower triangular of $\mathbf{H}$; see, e.g., Henderson & Searle (1979). Making general associated kernels from a type of kernel K_θ on $\mathbb{S}_\theta$ with $\theta = \theta(\mathbf{m}, \mathbf{D})$ by multivariate mode-dispersion method requires solving the system of the equations

$$(\theta(\mathbf{m}, \mathbf{D}))^\top = (\mathbf{x}, \operatorname{vech} \mathbf{H})^\top. \tag{2.11}$$

The solution of (2.11), if there exists, is a $k_d \times 1$ vector denoted by $\theta(\mathbf{x}, \mathbf{H}) := \theta\left(\mathbf{m}(\mathbf{x}, \mathbf{H}), \mathbf{D}(\mathbf{x}, \mathbf{H})\right)$. For $d = 1$, the system (2.11) provides the result in Libengué (2013). A light version will

be presented for bivariate case ($d = 2$) in Section 2.3.2. For classical associated kernels, the system (2.11) means to solve directly $(\mathbf{m}, \mathbf{D}) = (\mathbf{x}, \mathbf{H})$. More generally, the solution of (2.11) depends on the flexibility of the type of kernel K_θ and leads to the corresponding associated kernel denoted $K_{\theta(\mathbf{x},\mathbf{H})}$. The constructed associated kernel satisfies Definition 2.2.1 of multivariate associated kernel :

Proposition 2.2.9 *The associated kernel function $K_{\theta(\mathbf{x},\mathbf{H})}(\cdot)$, obtained from (2.11) and having support $\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}$, is such that :*

$$\mathbf{x} \in \mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}, \tag{2.12}$$

$$\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right) - \mathbf{x} = \mathbf{a}_\theta(\mathbf{x}, \mathbf{H}), \tag{2.13}$$

$$\mathrm{Cov}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right) = \mathbf{B}_\theta(\mathbf{x}, \mathbf{H}), \tag{2.14}$$

where $\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}$ is the random vector with pdf $K_{\theta(\mathbf{x},\mathbf{H})}$ and both $\mathbf{a}_\theta(\mathbf{x}, \mathbf{H}) = (a_{\theta 1}(\mathbf{x}, \mathbf{H}), \ldots, a_{\theta d}(\mathbf{x}, \mathbf{H}))^\top$ and $\mathbf{B}_\theta(\mathbf{x}, \mathbf{H}) = \left(b_{\theta ij}(\mathbf{x}, \mathbf{H})\right)_{i,j=1,\ldots,d}$ tend, respectively, to the null vector $\mathbf{0}$ and the null matrix $\mathbf{0}_d$ as $\mathbf{H}$ goes to $\mathbf{0}_d$.

Proof. The multivariate mode-dispersion method (2.11) implies $\theta(\mathbf{x}, \mathbf{H}) = \theta\left(\mathbf{m}(\mathbf{x}, \mathbf{H}), \mathbf{D}(\mathbf{x}, \mathbf{H})\right)$ which leads to $\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})} := \mathbb{S}_{\theta(\mathbf{m}(\mathbf{x},\mathbf{H}),\mathbf{D}(\mathbf{x},\mathbf{H}))}$. Since K_θ is unimodal of mode $\mathbf{m} \in \mathbb{S}_\theta$ (see Part (i) of Lemma 2.2.8), we obviously have (2.12), and then (2.3) is checked, because $\mathbf{m}$ is identified to $\mathbf{x}$ in (2.11). Let $\mathcal{Z}_{\theta(\mathbf{m},\mathbf{D})}$ be the random vector with unimodal pdf as the type of kernel $K_{\theta(\mathbf{m},\mathbf{D})}$. From Part (ii) of Lemma 2.2.8 we can write

$$\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{m},\mathbf{D})}\right) = \mathbf{m} + \boldsymbol{\tau}(\mathbf{m}, \mathbf{D}),$$

where $\boldsymbol{\tau}(\mathbf{m}, \mathbf{D})$ is the difference between the mean vector $\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{m},\mathbf{D})}\right)$ and the mode vector $\mathbf{m}$ of $\mathcal{Z}_{\theta(\mathbf{m},\mathbf{D})}$. Thus, from the mode-dispersion method (2.11), we have $\mathbf{m} = \mathbf{x}$ and $\boldsymbol{\tau}(\mathbf{m}, \mathbf{D}) = \boldsymbol{\tau}(\mathbf{m}(\mathbf{x}, \mathbf{H}), \mathbf{D}(\mathbf{x}, \mathbf{H}))$; taking $\mathbf{a}_\theta(\mathbf{x}, \mathbf{H}) = \boldsymbol{\tau}(\mathbf{m}(\mathbf{x}, \mathbf{H}), \mathbf{D}(\mathbf{x}, \mathbf{H}))$, Part (iii) of Lemma 2.2.8 leads to the second result (2.13), and therefore (2.4) is verified. Also, since $K_{\theta(\mathbf{m},\mathbf{D})}$ admits a moment of second order, the covariance matrix of $K_{\theta(\mathbf{m},\mathbf{D})}$ exists and that can be written as $\mathrm{Cov}\left(\mathcal{Z}_{\theta(\mathbf{m},\mathbf{D})}\right) = \mathbf{B}_\theta(\mathbf{m}, \mathbf{D})$; solving (2.11) and then taking $\mathbf{B}_\theta(\mathbf{x}, \mathbf{H}) := \mathbf{B}_\theta\left(\mathbf{m}(\mathbf{x}, \mathbf{H}), \mathbf{D}(\mathbf{x}, \mathbf{H})\right)$, the last result (2.14) holds in the sense of (2.5) using again Part (iii) of Lemma 2.2.8.■

In practice, both characteristics $\mathbf{a}_\theta(\mathbf{x}, \mathbf{H})$ and $\mathbf{B}_\theta(\mathbf{x}, \mathbf{H})$ are derived from the calculation of the mean vector and covariance matrix of $\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}$ in terms of the mode vector $\mathbf{m}(\mathbf{x}, \mathbf{H})$ and the dispersion matrix $\mathbf{D}(\mathbf{x}, \mathbf{H})$. In a general way, a given $K_{\mathbf{x},\mathbf{H}}$ or the constructed associated kernel $K_{\theta(\mathbf{x},\mathbf{H})}$ in Proposition 2.2.9 (that we will call standard version) creates a quantity in the bias of the kernel density estimation. In order to eliminate this quantity in the larger part of the support $\mathbb{T}_d$ of the density to be estimated, we will also study a modified version of the standard one. The following two subsections investigate some properties of these estimators.

2.2.2 Standard version of the estimator

Here, we give some properties of the estimator $\widehat{f_n}$ of f through a given associated kernel presented in Definition 2.2.1 or the constructed associated kernel in Proposi-

tion 2.2.9 ; i.e. $K_{\theta(\mathbf{x},\mathbf{H})}(\cdot) \equiv K_{\mathbf{x},\mathbf{H}}(\cdot)$. For a given bandwidth matrix $\mathbf{H}$, similarly to (2.2) we consider

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{X}_i), \quad \forall \mathbf{x} \in \mathbb{T}_d. \tag{2.15}$$

Proposition 2.2.10 *For given* $\mathbf{x} \in \mathbb{T}_d$,

$$\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} = \mathbb{E}\left\{f\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right\} \tag{2.16}$$

and $\widehat{f_n}(\mathbf{x}) \geq 0$. *Furthermore, one has*

$$\int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x})d\mathbf{x} = \Lambda, \tag{2.17}$$

where the total mass $\Lambda = \Lambda(n; \mathbf{H}, K_\theta)$ *is a positive real and, it is also a finite constant if* $\int_{\mathbb{T}_d} K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{t})d\mathbf{x} < \infty$ *for all* $\mathbf{t} \in \mathbb{T}_d$.

Proof. The first result (2.16) is straightforwardly obtained from (2.15) as follows :

$$\mathbb{E}\left\{\widehat{f_n}(\mathbf{x})\right\} = \int_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}\cap\mathbb{T}_d} K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{t})f(\mathbf{t})d\mathbf{t} = \mathbb{E}\left\{f\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right\}.$$

Also, the estimates $\widehat{f_n}(\mathbf{x}) \geq 0$ and the total mass $\Lambda > 0$ stem immediately from the fact that $K_{\theta(\mathbf{x},\mathbf{H})}(\cdot)$ is a pdf. Finally, the values of $\mathbf{X}_i$ belonging to the set $\mathbb{T}_d$, we have (2.17) as

$$\int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x})d\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{T}_d} K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{X}_i)d\mathbf{x} < \infty,$$

since the integration vector of $\int_{\mathbb{T}_d} K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{X}_i)d\mathbf{x}$ is on the target $\mathbf{x}$ which is a parameter of $K_{\theta(\mathbf{x},\mathbf{H})}(\cdot)$.■

From the above Proposition 2.2.10, the total mass Λ of $\widehat{f_n}$ by a non-classical associated kernel generally fails to be equal to 1 ; an illustration for $d = 2$ is given in Table 2.3 below. Hence, non-classical associated kernel estimators $\widehat{f_n}$ are improper density estimates or as kind of "balloon estimators" ; see Sain (2002) and Zougab *et al.* (2014). The fact that $\Lambda = \Lambda(n; \mathbf{H}, K_\theta)$ is close to 1 can find a statistical explanation in both Examples 1 and 2 of Romano and Thombs (1996), or by showing $\mathbb{E}\left(\int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x})d\mathbf{x}\right) = 1$ which does not depend on n : $\mathbb{E}\left(\int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x})d\mathbf{x}\right) = \mathbb{E}\left\{\phi(\mathbf{X}_1\mathbb{1}_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}\cap\mathbb{T}_d}; \mathbf{H}, K_\theta)\right\}$ with $\phi(\mathbf{t}; \mathbf{H}, K_\theta) = \int_{\mathbb{T}_d} K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{t})d\mathbf{x} < \infty$ for all $\mathbf{t} \in \mathbb{T}_d$. Without loss of generality, we study $\mathbf{x} \mapsto \widehat{f_n}(\mathbf{x})$ up to normalizing constant which is used at the end of the density estimation process.

Proposition 2.2.11 *Let* $\mathbf{x} \in \mathbb{T}_d$ *be a target and a bandwidth matrix* $\mathbf{H} \equiv \mathbf{H}_n \to \mathbf{0}_d$ *as* $n \to \infty$. *Assume* f *in the class* $\mathscr{C}^2(\mathbb{T}_d)$, *then*

$$\begin{aligned}\text{Bias}\left\{\widehat{f_n}(\mathbf{x})\right\} &= \mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H})\nabla f(\mathbf{x}) + \frac{1}{2}\text{trace}\left[\left\{\mathbf{a}_\theta(\mathbf{x},\mathbf{H})\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H}) + \mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right\}\nabla^2 f(\mathbf{x})\right] \\ &\quad + o\left\{\text{trace}\left(\mathbf{H}^2\right)\right\}. \end{aligned} \tag{2.18}$$

Furthermore, if f is bounded on $\mathbb{T}_d$ then there exists the largest positive real number $r_2 = r_2(K_\theta)$ such that $\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 \lesssim c_2(\mathbf{x})(\det \mathbf{H})^{-r_2}$, $0 \leq c_2(\mathbf{x}) \leq \infty$ and

$$\operatorname{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} = \frac{1}{n}\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x}) + \mathrm{o}\left(n^{-1}\,(\det \mathbf{H})^{-r_2}\right), \tag{2.19}$$

with $\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 := \int_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}} K^2_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{u})d\mathbf{u}$ and where "$\lesssim$" stands for "$\leq$ and then approximation as $n \to \infty$".

Proof. See Appendix in Section 2.6.

The univariate case ($d = 1$) of Proposition 2.2.11 can be found in Libengué (2013) ; see Chen (1999, 2000) for both beta and gamma kernels with $r_2 = 1/2$ and, also, Hirukawa and Sakudo (2014) for other values of r_2. In the situation of multiple associated kernels (2.9), in contrario to (2.18) the general representation (2.19) is simply expressed in terms of univariate associated kernels as follows :

Corollaire 2.2.12 *Under (2.9) with $r_{2,j} = r_2\left(K^{[j]}_{\theta_j}\right)$ and where $K^{[j]}_{\theta_j} \equiv K^{[j]}$ refers to the jth univariate type of kernel, then (2.19) is equivalent to*

$$\operatorname{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} = \frac{1}{n}\left(\prod_{j=1}^{d}\left\|K^{[j]}_{\theta_j(x_j,h_{jj})}\right\|_2^2\right) f(\mathbf{x}) + \mathrm{o}\left(n^{-1}\ \prod_{j=1}^{d} h_{jj}^{-r_{2,j}}\right).$$

Proof. Easy.■

Now, we recall natural measures for assessing the similarity of the multivariate associated kernel estimator $\widehat{f_n}$ according to the true pdf f, to be estimated. Since the pointwise measure is the mean squared error (MSE) and expressed by

$$\operatorname{MSE}(\mathbf{x}) = \operatorname{Bias}^2\left\{\widehat{f_n}(\mathbf{x})\right\} + \operatorname{Var}\left\{\widehat{f_n}(\mathbf{x})\right\}, \tag{2.20}$$

the integrated form of MSE on $\mathbb{T}_d$ is $\operatorname{MISE}\left(\widehat{f_n}\right) = \int_{\mathbb{T}_d} \operatorname{MSE}(\mathbf{x})\,d\mathbf{x}$ and its approximate expression satisfies

$$\begin{aligned}\operatorname{AMISE}\left(\widehat{f_n}\right) \;=\; & \int_{\mathbb{T}_d}\left(\left[\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H})\nabla f(\mathbf{x}) + \frac{1}{2}\operatorname{trace}\left\{\left(\mathbf{a}_\theta(\mathbf{x},\mathbf{H})\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H}) + \mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right)\nabla^2 f(\mathbf{x})\right\}\right]^2\right. \\ & \left. + \frac{1}{n}\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x})\right) d\mathbf{x}. \end{aligned} \tag{2.21}$$

In the general case of associated kernels with correlation structure, an optimal bandwidth matrix that minimizes the AMISE (2.21) still remains challenging problem, even in the bivariate case. However, the particular case of diagonal bandwidth matrix is $\mathbf{H}_{\text{opt,diag}} = O\left(n^{-2/(d+4)}\right)$ for multiple gamma kernels ; see Bouerzmarni and Rombouts (2010) for further details. So, the next result only gives the optimal bandwidth matrix for the Scott form which still has a correlation structure.

In fact, let us consider $\mathbf{H} = h\mathbf{H}_0$ be a Scott bandwidth matrix with fixed matrix $\mathbf{H}_0$ and positive $h \equiv h_n \to 0$ as $n \to \infty$. From Definition 2.2.1, Proposition 2.2.4 and Proposition 2.2.9, it follows that there exists a $d \times 1$ vector $\mathbf{c}^*(\mathbf{x}, \mathbf{H}_0) = (c_1^*(\mathbf{x}, \mathbf{H}_0), \ldots, c_d^*(\mathbf{x}, \mathbf{H}_0))^\top$ and a $d \times d$ matrix $\mathbf{C}^{**}(\mathbf{x}, \mathbf{H}_0) = (c_{ij}^{**}(\mathbf{x}, \mathbf{H}_0))_{i,j=1,\ldots,d}$ of finite constants connected respectively to $\mathbf{a}_\theta(\mathbf{x}, \mathbf{H}) = (a_{\theta 1}(\mathbf{x}, \mathbf{H}), \ldots, a_{\theta d}(\mathbf{x}, \mathbf{H}))^\top$ and $\mathbf{B}_\theta(\mathbf{x}, \mathbf{H}) = (b_{\theta ij}(\mathbf{x}, \mathbf{H}))_{i,j=1,\ldots,d}$ such that for all $i, j = 1, \ldots, d$, $a_{\theta i}(\mathbf{x}, \mathbf{H}) \leq h c_i^*(\mathbf{x}, \mathbf{H}_0)$ and $b_{\theta ij}(\mathbf{x}, \mathbf{H}) \leq h^2 c_{ij}^{**}(\mathbf{x}, \mathbf{H}_0)$. Using Proposition 2.2.11 and assuming f such that its all first and second partial derivatives are bounded, one has :

$$\text{Bias}\left\{\widehat{f_n}(\mathbf{x})\right\} \leq h s_1(\mathbf{x}) \text{ and } \text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} \leq n^{-1} h^{-dr_2} s_2(\mathbf{x}),$$

where $s_1(\mathbf{x}) \geq \mathbf{c}^{*\top}(\mathbf{x}, \mathbf{H}_0) \nabla f(\mathbf{x}) + (h/2) \operatorname{trace}\left[\{\mathbf{c}^{*\top}(\mathbf{x}, \mathbf{H}_0)\mathbf{c}^*(\mathbf{x}, \mathbf{H}_0) + \mathbf{C}^{**}(\mathbf{x}, \mathbf{H}_0)\} \nabla^2 f(\mathbf{x})\right]$ and $s_2(\mathbf{x})$ are positive scalars. From (2.20), it follows that $\text{MSE}(\mathbf{x}) \leq h^2 s_1^2(\mathbf{x}) + n^{-1} h^{-dr_2} s_2(\mathbf{x})$. By integration of $\text{MSE}(\mathbf{x})$, one obtains

$$\text{AMISE}(\widehat{f_{n,h\mathbf{H}_0,K_\theta,f}}) \leq h^2 t_1 + n^{-1} h^{-dr_2} t_2, \tag{2.22}$$

with t_1 and t_2 the anti-derivatives of respectively $s_1^2(\mathbf{x})$ and $s_2(\mathbf{x})$ on $\mathbb{T}_d$. Taking the derivative of the second member of the inequality (2.22) equal to 0 leads to the following proposition.

Proposition 2.2.13 *Under assumption of Proposition 2.2.11, let* $\mathbf{H} = h\mathbf{H}_0$ *be a Scott bandwidth matrix with fixed matrix* $\mathbf{H}_0$ *and positive* $h \equiv h_n \to 0$ *as* $n \to \infty$ *and such that the right member of (2.22) satisfies*

$$0 < h^2 t_1 + n^{-1} h^{-dr_2} t_2 < \infty, \tag{2.23}$$

where $r_2 = r_2(K_\theta)$ *given in (2.19). Then, the optimal bandwidth matrix* $\mathbf{H}_{\text{opt,Scott}}$ *minimizing the* AMISE *in (2.21) is*

$$\mathbf{H}_{\text{opt,Scott}} = C n^{-1/(dr_2+2)} \mathbf{H}_0, \tag{2.24}$$

where C is a positive constant.

Note that we can use the Scott bandwidth matrix $\mathbf{H} = h\mathbf{H}_0$ if (2.23) holds. However, in practice, we cannot check (2.23) because f is unknown. But, if the quantity $h^2 t_1 + n^{-1} h^{-dr_2} t_2$ of (2.23) becomes 0 (resp. ∞) then one observes an undersmoothing (resp. oversmoothing). Finally, the practical choice of $\mathbf{H}_0$ in (2.24) can be the sample covariance matrix.

Unlike to classical associated kernels for $\mathbb{T}_d = \mathbb{R}^d$ (Example 2.2.3), the choice of non-classical associated kernels is very important for the support $\mathbb{T}_d \left(\subseteq \mathbb{R}^d\right)$ of the pdf f to be estimated ; see Parts (i)-(iv) of Remark 2.2.2. Also, from Parts (v)-(vi) of Remark 2.2.2, different positions of the target $\mathbf{x} \in \mathbb{T}_d$ and correlation structure need a suitable general associated kernel. Nevertheless, the selection of bandwidth matrix remains crucial when the general associated kernel is chosen ; see, e.g. Chacon and Duong (2011) and Chacon *et al.* (2011). Here, we consider the multivariate least squares cross validation (LSCV) method to select the bandwidth matrix. From (2.15), the LSCV method is based on the minimization of the integrated squared error (ISE) which can be written as

$$\text{ISE}(\mathbf{H}) = \int_{\mathbb{T}_d} \widehat{f_n}^2(\mathbf{x}) d\mathbf{x} - 2 \int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} + \int_{\mathbb{T}_d} f^2(\mathbf{x}) d\mathbf{x}.$$

Minimizing this ISE($\mathbf{H}$) means to minimize the two first terms. However, we need to estimate the second term since it depends on the unknown pdf f. The LSCV estimator of ISE($\mathbf{H}$) $- \int_{\mathbb{T}_d} f^2(\mathbf{x})d\mathbf{x}$ is

$$\text{LSCV}(\mathbf{H}) = \int_{\mathbb{T}_d} \left\{\widehat{f_n}(\mathbf{x})\right\}^2 d\mathbf{x} - \frac{2}{n}\sum_{i=1}^{n} \widehat{f_{n,-i}}(\mathbf{X}_i),$$

where $\widehat{f_{n,-i}}(\mathbf{X}_i) = (n-1)^{-1}\sum_{j\neq i} K_{\theta(\mathbf{X}_i,\mathbf{H})}(\mathbf{X}_j)$ is being computed as $\widehat{f_n}(\mathbf{X}_i)$ excluding the observation $\mathbf{X}_i$. The bandwidth matrix obtained by the LSCV rule selection is defined as follows :

$$\widehat{\mathbf{H}} = \arg\min_{\mathbf{H}\in\mathcal{H}} \text{LSCV}(\mathbf{H}), \tag{2.25}$$

where $\mathcal{H}$ is the set of all positive definite full bandwidth matrices. The LSCV rule is the same for the Scott and diagonal bandwidth matrices where $\mathcal{S}$ and $\mathcal{D}$ are their respective sets. The difficulty comes from the level of dimension of $\mathbf{H}$ which is, respectively, $d(d+1)/2$, 1 and d for $\mathcal{H}$, $\mathcal{S}$ and $\mathcal{D}$. Thus, for high dimension $d > 2$, the set of the Scott bandwidth matrices might be a good compromise between computational problems and correlation structures in the sample.

2.2.3 Modified version of the estimator

Following Chen (1999, 2000), Libengué (2013), Malec and Schienle (2014), Hirukawa and Sakudo (2014) and Igarashi and Kakizawa (2015) in univariate case, a second version of the estimator (2.15) is sometimes necessary ; see, e.g., Funke and Kawka (2015) for multiple kernels. Indeed, the presence of the non null term $\mathbf{a}_\theta(\mathbf{x},\mathbf{H})$ with the gradient $\nabla f(\mathbf{x})$ in (2.18) increases the pointwise bias of $\widehat{f_n}(\mathbf{x})$. Thus, we propose below an algorithm for eliminating the term of gradient in the largest region of $\mathbb{T}_d$. Since Bouerzmarni and Rombouts (2010) has shown the results for multiple associated kernels, we here investigate the case of general associated kernels with $d > 1$.

The algorithm of bias reduction has two steps. The first step consists to define both inside and boundary regions. The second one deals on the modified associated kernel which leads to the bias reduction in the interior domain.

First step. Partitioning $\mathbb{T}_d$ into two regions of order $\boldsymbol{\alpha}(\mathbf{H}) = (\alpha_1(\mathbf{H}),\ldots,\alpha_d(\mathbf{H}))^\top$ which is a $d\times 1$ vector with $\alpha_1(\mathbf{H})$, ..., $\alpha_d(\mathbf{H}) \in \mathbb{R}$, and where $\boldsymbol{\alpha}(\mathbf{H})$ tends to the null vector $\mathbf{0}$ as $\mathbf{H}$ goes to the null matrix $\mathbf{0}_d$:

a. *interior region* is the largest one inside the interior of $\mathbb{T}_d$ in order to contain at least 95 percent of observations, and it is denoted by $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}$;
b. *boundary regions* representing the complementary of $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}$ in $\mathbb{T}_d$, and it is denoted by $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B}$ which could be empty ; recall that $\mathbb{T}_d = \mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I} \cup \mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B}$ and $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I} \cap \mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B} = \emptyset$.

Since $\mathbb{T}_d \left(\subseteq \mathbb{R}^d\right)$ might have each one of its d convex components as unbounded, partially bounded or totally bounded interval as in (2.1), there is only one $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}$ but

$$N_d = 1^{d_\infty} 2^{d_z} 3^{d_{uw}} - 1 \tag{2.26}$$

boundary subregions of $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B}$. In the below Section 2.3.3 an illustration is provided for $d = 2$ with $\mathbb{T}_2 = [0,1] \times [0,1]$ and, therefore, the number of boundary subregions (2.26) is $N_2 = 3^2 - 1 = 8$.

Second step. Changing the general associated kernel $K_{\theta(\mathbf{x},\mathbf{H})}$ into its modified version $K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}$; that leads to replace the couple $(\mathbf{a}_\theta(\mathbf{x},\mathbf{H}), \mathbf{B}_\theta(\mathbf{x},\mathbf{H}))$ into $\left(\mathbf{a}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H}), \mathbf{B}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H})\right)$ with $\mathbf{a}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H}) = \mathbf{a}_{\widetilde{\theta}_B}(\mathbf{x},\mathbf{H})\mathbb{1}_{\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B}}(\mathbf{x})$ because $\mathbf{a}_{\widetilde{\theta}_I}(\mathbf{x},\mathbf{H})\mathbb{1}_{\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}}(\mathbf{x}) = \mathbf{0}$ in the interior, and $\mathbf{B}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H}) = \mathbf{B}_{\widetilde{\theta}_I}(\mathbf{x},\mathbf{H})\mathbb{1}_{\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}}(\mathbf{x}) + \mathbf{B}_{\widetilde{\theta}_B}(\mathbf{x},\mathbf{H})\mathbb{1}_{\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B}}(\mathbf{x})$. This modified associated kernel is such that, for any fixed bandwidth matrix $\mathbf{H}$,

$$\widetilde{\theta}(\mathbf{x},\mathbf{H}) = \begin{cases} \widetilde{\theta}_I(\mathbf{x},\mathbf{H}) : \quad \mathbf{a}_{\widetilde{\theta}_I}(\mathbf{x},\mathbf{H}) = \mathbf{0} & \text{if } \mathbf{x} \in \mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I} \\ \widetilde{\theta}_B(\mathbf{x},\mathbf{H}) & \text{if } \mathbf{x} \in \mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B} \end{cases} \tag{2.27}$$

must be continuous on $\mathbb{T}_d$ and constant on $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B}$.

Proposition 2.2.14 *The function $K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}$ on its support $\mathbb{S}_{\widetilde{\theta}(\mathbf{x},\mathbf{H})} = \mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}$ and obtained from (2.27) is also a general associated kernel.*

Proof. Since $K_{\theta(\mathbf{x},\mathbf{H})}$ is a general associated kernel for all $\mathbf{x} \in \mathbb{T}_d = \mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I} \cup \mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B}$, one gets the first condition (2.3) of Definition 2.2.1 because we have $\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})} = \mathbb{S}_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}$. According to Proposition 2.2.9 it follows that, for a given random variable $\mathcal{Z}_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}$ with pdf $K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}$, we obtain the last two conditions (2.4) and (2.5) of Definition 2.2.1 as, respectively,

$$\mathbb{E}\left(\mathcal{Z}_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}\right) = \mathbf{x} + \mathbf{a}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H}) \text{ and } \mathrm{Cov}\left(\mathcal{Z}_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}\right) = \mathbf{B}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H}).$$

Both quantities $\mathbf{a}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H})$ and $\mathbf{B}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H})$ tend, respectively, to the null vector $\mathbf{0}$ and the null matrix $\mathbf{0}_d$ as $\mathbf{H}$ goes to $\mathbf{0}_d$. In particular, from (2.27) we have easily $\mathbf{a}_{\widetilde{\theta}_I}(\mathbf{x},\mathbf{H}) = \mathbf{0}$ in the interior.■

Similar to both (2.2) and (2.15), the *modified associated kernel estimator* $\widetilde{f_n}$ using $K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}$ is then defined by

$$\widetilde{f_n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}(\mathbf{X}_i). \tag{2.28}$$

The following result gives only in the interior $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}$ of $\mathbb{T}_d$ the pointwise expressions of the bias and the variance of $\widetilde{f_n}$. Of course, the corresponding expressions in the boundary regions $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),B}$ are tedious to write with respect to the N_d (2.26) boundary situations from (2.27).

Proposition 2.2.15 *Let $\widehat{f_n}$ and $\widetilde{f_n}$ be the multivariate associated kernel estimators of f defined in (2.15) and (2.28) respectively. Then, for $\mathbf{x} \in \mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}$ as in (2.27) :*

$$\mathrm{Bias}\left\{\widetilde{f_n}(\mathbf{x})\right\} = \frac{1}{2}\,\mathrm{trace}\left(\mathbf{B}_{\widetilde{\theta}_I}(\mathbf{x},\mathbf{H})\nabla^2 f(\mathbf{x})\right) + \mathrm{o}\left\{\mathrm{trace}\left(\mathbf{H}^2\right)\right\} \tag{2.29}$$

and

$$\mathrm{Var}\left\{\widetilde{f_n}(\mathbf{x})\right\} \simeq \mathrm{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} \text{ as } n \to \infty.$$

Proof. We have the first result (2.29) by replacing in (2.18) $\widehat{f_n}$, $\mathbf{a}_\theta$ and $\mathbf{B}_\theta$ by $\widetilde{f_n}$, $\mathbf{a}_{\widetilde{\theta}_I}$ and $\mathbf{B}_{\widetilde{\theta}_I}$, respectively. For the last result, considering (2.19) it is sufficient to show that

$$\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 \simeq \left\|K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}\right\|_2^2 \text{ as } \mathbf{H} \to \mathbf{0}_d.$$

Since $K_{\theta(\mathbf{x},\mathbf{H})}$ and $K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}$ are general associated kernels of the same type K_θ with $r_2 = r_2(K_\theta)$ and $\widetilde{r}_2 = \widetilde{r}_2(K_{\widetilde{\theta}})$, then there exists a common largest positive real number $r_2^* = r_2^*(K_\theta)$ such that

$$\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 \lesssim c_2(\mathbf{x})(\det \mathbf{H})^{-r_2^*} \text{ and } \left\|K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}\right\|_2^2 \lesssim \widetilde{c}_2(\mathbf{x})(\det \mathbf{H})^{-r_2^*}$$

with $0 < c_2(\mathbf{x}), \widetilde{c}_2(\mathbf{x}) < \infty$. Taking $c(\mathbf{x}) = \sup\{c_2(\mathbf{x}), \widetilde{c}_2(\mathbf{x})\}$, we have $\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 \lesssim c(\mathbf{x})(\det \mathbf{H})^{-2r_2^*}$ and $\left\|K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}\right\|_2^2 \lesssim c(\mathbf{x})(\det \mathbf{H})^{-2r_2^*}$. Since $c(\mathbf{x})/n(\det \mathbf{H})^{2r_2^*} = o(n^{-1}(\det \mathbf{H})^{-2r_2^*})$ then $\left\|K_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}\right\|_2^2 \simeq \left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2$.■

Thus, we define the asymptotic expression of the MISE of $\widetilde{f_n}$ in the interior $\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}$ as follows :

$$\text{AMISE}_{\widetilde{\theta}_I}\left(\widetilde{f_n}\right) = \int_{\mathbb{T}_d^{\boldsymbol{\alpha}(\mathbf{H}),I}} \left(\left\{\frac{1}{2}\,\text{trace}\left(\mathbf{B}_{\widetilde{\theta}_I}(\mathbf{x},\mathbf{H})\nabla^2 f(\mathbf{x})\right)\right\}^2 + \frac{1}{n}\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x})\right) d\mathbf{x}.$$

All the results in this section can be easily deduced for both cases of diagonal and Scott bandwidth matrices. The following section provides some detailed results for $d = 2$ with a bivariate beta kernel having correlation structure on $\mathbb{T}_2 = [0,1] \times [0,1]$. The corresponding associated kernel is built from a technique due to Sarmanov (1966) and using two independent univariate beta pdfs. See Balakrishnan and Lai (2009), Kundi *et al.* (2010) for some other examples of bivariate types of kernels and Kotz *et al.* (2000) in multivariate case.

2.3 Bivariate beta kernel with correlation structure

This section presents the generalizable procedure of a bivariate beta kernel estimator from a bivariate beta pdf with correlation structure to its corresponding associated kernel. The standard associated kernel is built by a variant of the mode-dispersion method deduced from (2.11). Then, we provide properties of both versions of the corresponding estimators.

2.3.1 Type of bivariate beta kernel

In order to control better the effects of correlation, we here consider a flexible type of bivariate beta kernel for which the correlation structure is introduced by Sarmanov (1966) ; see also Lee (1996). Let us take two independent univariate beta distributions with pdfs

$$g_j(t) = \frac{1}{\mathscr{B}(p_j, q_j)} t^{p_j - 1}(1-t)^{q_j - 1}\mathbb{1}_{[0,1]}(t), \quad j = 1, 2, \tag{2.30}$$

where $\mathscr{B}(p_j, q_j) = \int_0^1 t^{p_j-1}(1-t)^{q_j-1}dt$ is the usual beta function with $p_j > 0$ and $q_j > 0$. Their means and variances are, respectively,

$$\mu_j = \frac{p_j}{p_j + q_j} = \mu_j(p_j, q_j) \quad \text{and} \quad \sigma_j^2 = \frac{p_j q_j}{(p_j + q_j)^2(p_j + q_j + 1)} = \sigma_j^2(p_j, q_j). \tag{2.31}$$

Also, g_j are unimodal for $p_j \geq 1$, $q_j \geq 1$ and $(p_j, q_j) \neq (1, 1)$, with mode and dispersion parameters :

$$m_j(p_j, q_j) = \frac{p_j - 1}{p_j + q_j - 2} \quad \text{and} \quad d_j = \frac{1}{p_j + q_j - 2} = d_j(p_j, q_j). \tag{2.32}$$

The corresponding pdf (or type of kernel) of the bivariate beta-Sarmanov with correlation and from g_j of (2.30) is then denoted by g_θ (= BS_θ) and defined as :

$$g_\theta(\mathbf{v}) = g_1(v_1)g_2(v_2)\left[1 + \rho \times \frac{v_1 - \mu_1(p_1, q_1)}{\sigma_1(p_1, q_1)} \times \frac{v_2 - \mu_2(p_2, q_2)}{\sigma_2(p_2, q_2)}\right] \mathbb{1}_{[0,1]\times[0,1]}(\mathbf{v}), \tag{2.33}$$

with $\mathbf{v} = (v_1, v_2)^\top$ and $\theta := \theta(p_1, q_1, p_2, q_2, \rho) \in \Theta \subseteq \mathbb{R}^5$. Depending on p_j and q_j, the correlation parameter $\rho = \rho(p_1, q_1, p_2, q_2)$ belongs to the following interval

$$[-\varepsilon, \varepsilon'] \subset [-1, 1], \tag{2.34}$$

with nonnegative reals

$$\varepsilon = \left(\max_{v_1, v_2}\left\{\frac{v_1 - \mu_1(p_1, q_1)}{\sigma_1(p_1, q_1)} \times \frac{v_2 - \mu_2(p_2, q_2)}{\sigma_2(p_2, q_2)}\right\}\right)^{-1}$$

and

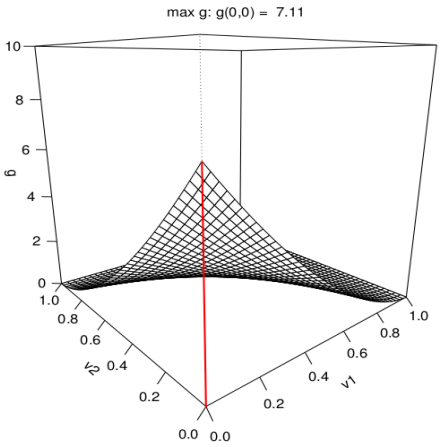
$$\varepsilon' = \left|\left(\min_{v_1, v_2}\left\{\frac{v_1 - \mu_1(p_1, q_1)}{\sigma_1(p_1, q_1)} \times \frac{v_2 - \mu_2(p_2, q_2)}{\sigma_2(p_2, q_2)}\right\}\right)^{-1}\right|.$$

Thus, the mean vector and covariance matrix of g_θ are, respectively,

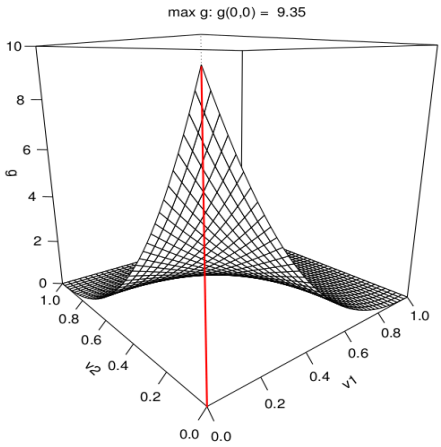
$$\boldsymbol{\mu} = (\mu_1, \mu_2)^\top \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}.$$

The unimodality of g_θ in (2.33) also occurs for $p_j \geq 1$, $q_j \geq 1$ and $(p_j, q_j) \neq (1, 1)$ with $j = 1, 2$. However, the corresponding mode vector $\mathbf{m}_\theta = \mathbf{m}(p_1, p_2, q_1, q_2, \rho) =: \mathbf{m}_\rho$ of g_θ does not have an explicit expression ; but, we numerically verified that this mode vector $\mathbf{m}_\rho$ is slightly shifted with respect to the modal vector (2.32) of the two independent margins that we denote by $\mathbf{m}_0 = (m_1(p_1, q_1), m_2(p_2, q_2))^\top$ for $\rho = 0$.

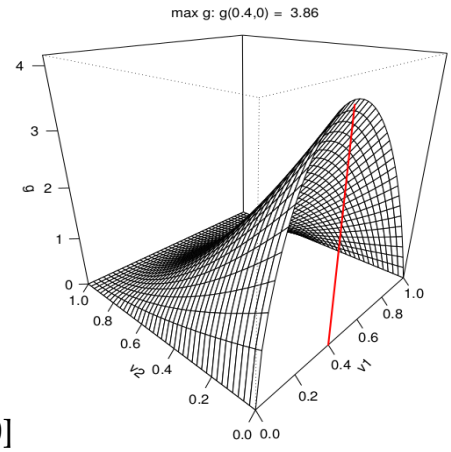
Figure 2.1 illustrates some different effects of both null and positive correlations (2.34) on the unimodality of (2.33). The negative correlations will show the opposite effects in terms of positions according to the modal vector $\mathbf{m}_0$ for $\rho = 0$. In other words, the correlation parameter ρ enables the pdf g_θ to reach points which are inaccessible with the null correlation. The parameter values of both univariate beta (2.30) used for Figure 2.1 produce the following intervals (2.34) of correlation :



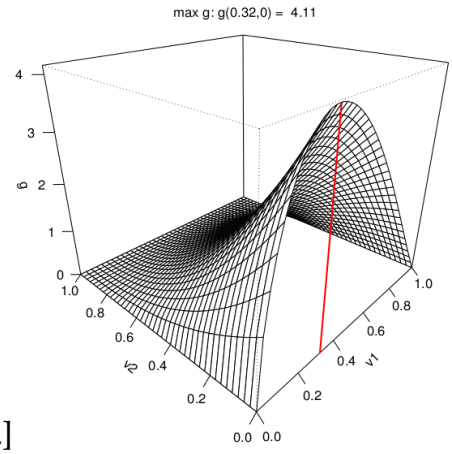
[(a0) : $\rho = 0$]



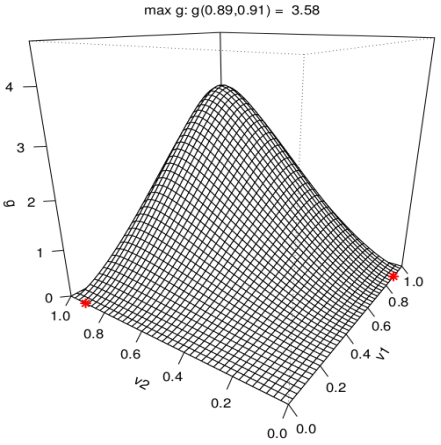
[(a1) : $\rho = 0.18$]



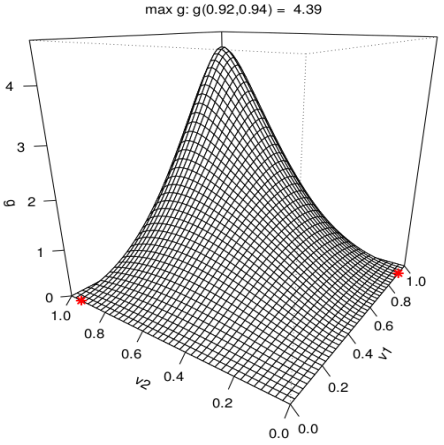
[(b0) : $\rho = 0$]



[(b1) : $\rho = 0.12$]



[(c0) : $\rho = 0$]



[(c1) : $\rho = 0.19$]

FIGURE 2.1 – Some shapes of the bivariate beta-Sarmanov type (2.33) with different effects of correlations on the unimodality ((a) : $p_1 = p_2 = 1, q_1 = q_2 = 8/3$; (b) : $p_1 = 5/3, p_2 = 1, q_1 = 2, q_2 = 8/3$; (c) : $p_1 = 149/60, p_2 = 151/60, q_1 = 71/60, q_2 = 23/20$).

– $\rho \in [-0.100, 0.210]$ if $p_1 = p_2 = 1, q_1 = q_2 = 8/3$ for an angle ;
– $\rho \in [-0.120, 0.143]$ if $p_1 = 5/3, p_2 = 1, q_1 = 2, q_2 = 8/3$ for an edge ;
– $\rho \in [-0.008, 0.214]$ if $p_1 = 149/60, p_2 = 151/60, q_1 = 71/60, q_2 = 23/20$ for the interior.

Concerning the dispersion matrix $\mathbf{D}_\rho$ of the bivariate beta-Sarmanov type we consider

$$\mathbf{D}_\rho = \begin{pmatrix} d_1 & (d_1 d_2)^{1/2}\rho \\ (d_1 d_2)^{1/2}\rho & d_2 \end{pmatrix}, \tag{2.35}$$

where d_1 and d_2 are the dispersion parameters (2.32) of margins and ρ the correlation parameter. This dispersion matrix is the analogue of the covariance one. Since we do not have a closed expression of the modal vector $\mathbf{m}_\rho$, we cannot use the bivariate mode-dispersion method (2.11) for a construction of the bivariate beta-Sarmanov kernel.

2.3.2 Bivariate beta-Sarmanov kernel

From the previous section, the standard version of the bivariate beta-Sarmanov kernel is here constructed by using the modal vector $\mathbf{m}_0 = (m_1(p_1, q_1), m_2(p_2, q_2))^\top$ of $\rho = 0$ instead of $\mathbf{m}_\rho$ as a variant of the mode-dispersion method (2.11). This choice will be compensated in the bandwidth matrix $\mathbf{H}$ connected to the complete dispersion matrix $\mathbf{D}_\rho$ with correlation structure (2.35).

Indeed, solving $\Big(\theta(\mathbf{m}_0, \mathbf{D}_\rho)\Big)^\top = (\mathbf{x}, \operatorname{vech}\mathbf{H})^\top$ in the sense of $\mathbf{m}_0 = \mathbf{x}$ and $\mathbf{D}_\rho = \mathbf{H}$ leads to the new reparametrization of g_θ of (2.33) from $\theta = \theta(p_1, q_1, p_2, q_2, \rho) \subseteq \mathbb{R}^5$ into

$$\theta(\mathbf{x}, \mathbf{H}) = \left(\frac{x_1}{h_{11}} + 1, \frac{1 - x_1}{h_{11}} + 1, \frac{x_2}{h_{22}} + 1, \frac{1 - x_2}{h_{22}} + 1, \frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top, \ \forall \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}. \tag{2.36}$$

Rewriting (2.31) in terms of (2.36), the means $\mu_j(p_j, q_j)$ and variances $\sigma_j^2(p_j, q_j)$ of the univariate beta pdfs become

$$\widetilde{\mu}_j = \frac{x_j + h_{jj}}{1 + 2h_{jj}} = \widetilde{\mu}_j(x_j, h_{jj}) \quad \text{and} \quad \widetilde{\sigma}_j^2 = \frac{(x_j + h_{jj})(1 + h_{jj} - x_j)}{(1 + 2h_{jj})^2(1 + 3h_{jj})} h_{jj} = \widetilde{\sigma}_j^2(x_j, h_{jj}).$$

Finally, the bivariate beta-Sarmanov kernel is defined as $BS_{\theta(\mathbf{x},\mathbf{H})} := g_{\theta(\mathbf{x},\mathbf{H})}$ such that

$$\begin{aligned} BS_{\theta(\mathbf{x},\mathbf{H})}(v_1, v_2) \ = \ & \left(\frac{v_1^{x_1/h_{11}}(1 - v_1)^{(1-x_1)/h_{11}}}{\mathscr{B}(1 + x_1/h_{11}, 1 + (1 - x_1)/h_{11})}\right)\left(\frac{v_2^{x_2/h_{22}}(1 - v_2)^{(1-x_2)/h_{22}}}{\mathscr{B}(1 + x_2/h_{22}, 1 + (1 - x_2)/h_{22})}\right) \\ & \times \left(1 + h_{12} \times \frac{v_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2}\widetilde{\sigma}_1(x_1, h_{11})} \times \frac{v_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2}\widetilde{\sigma}_2(x_2, h_{22})}\right) \mathbb{1}_{[0,1]^2}(v_1, v_2), \end{aligned} \tag{2.37}$$

with the constraints

$$h_{12} \in [-\beta, \beta'] \cap \left(-\sqrt{h_{11}h_{22}}, \sqrt{h_{11}h_{22}}\right), \tag{2.38}$$

$$\beta = \left(\max_{v_1, v_2} \left\{\frac{v_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2}\widetilde{\sigma}_1(x_1, h_{11})} \times \frac{v_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2}\widetilde{\sigma}_2(x_2, h_{22})}\right\}\right)^{-1}$$

and

$$\beta' = \left| \left(\min_{v_1, v_2} \left\{ \frac{v_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2} \widetilde{\sigma}_1(x_1, h_{11})} \times \frac{v_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2} \widetilde{\sigma}_2(x_2, h_{22})} \right\} \right)^{-1} \right|.$$

The first interval $[-\beta, \beta']$ of (2.38) is the equivalent $[-\varepsilon, \varepsilon']$ in (2.34) and the second one $(-\sqrt{h_{11}h_{22}}\,,\, \sqrt{h_{11}h_{22}})$ of (2.38) is due to the constraints from the bandwidth matrix **H**, which is symmetric and positive definite. In practice, one often has : $[-\beta, \beta'] \subset (-\sqrt{h_{11}h_{22}},\ \sqrt{h_{11}h_{22}})$. Of course, the beta-Sarmanov kernel $BS_{\theta(\mathbf{x},\mathbf{H})}$ satisfies Definition 2.2.1 of associated kernel :

$$\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})} = [0,1] \times [0,1]\,, \quad \mathbf{a}_\theta(\mathbf{x}, \mathbf{H}) = (a_{\theta 1}, a_{\theta 2})^\top \ \text{ with } a_{\theta j} = \frac{(1 - 2x_j)h_{jj}}{1 + 2h_{jj}} = a_{\theta j}(x_j, h_{jj}),$$

$$\mathbf{B}_\theta(\mathbf{x}, \mathbf{H}) = (b_{\theta ij})_{i,j=1,2} \text{ with } b_{\theta jj} = \widetilde{\sigma}_j^2(x_j, h_{jj}) \text{ and } b_{\theta 12} = \frac{h_{12}}{\sqrt{h_{11}h_{22}}} \widetilde{\sigma}_1(x_1, h_{11}) \widetilde{\sigma}_2(x_2, h_{22}). \tag{2.39}$$

Table 2.2 shows some effects of the correlation parameter h_{12} in $BS_{\theta(\mathbf{x},\mathbf{H})}$ on the modal vector and maximum values, which are obtained by using corresponding values of $(p_1, q_1, p_2, q_2, \rho)$ for Figure 2.1.

Position	$\mathbf{x}$	Interval of h_{12}	$\mathbf{H}$	Maximum value of BS_θ
Angle	(0, 0)	[-0.040, 0.128]	**Diag** (0.600, 0.600)	$BS_{\theta(\mathbf{x},\mathbf{H})}(0,0) = 7.11$
			$\begin{pmatrix} 0.600 & 0.128 \\ 0.128 & 0.600 \end{pmatrix}$	$BS_{\theta(\mathbf{x},\mathbf{H})}(0,0) = 9.77$
Edge	(0.40, 0.00)	[-0.050, 0.104]	**Diag** (0.600, 0.600)	$BS_{\theta(\mathbf{x},\mathbf{H})}(0.40, 0.00) = 3.86$
			$\begin{pmatrix} 0.600 & 0.104 \\ 0.104 & 0.600 \end{pmatrix}$	$BS_{\theta(\mathbf{x},\mathbf{H})}(0.32, 0.00) = 4.11$
Interior	(0.89, 0.91)	[-0.060, 0.128]	**Diag** (0.612, 0.600)	$BS_{\theta(\mathbf{x},\mathbf{H})}(0.89, 0.91) = 3.58$
			$\begin{pmatrix} 0.612 & 0.123 \\ 0.123 & 0.600 \end{pmatrix}$	$BS_{\theta(\mathbf{x},\mathbf{H})}(0.92, 0.94) = 4.46$

TABLE 2.2 – The corresponding values of Figure 2.1 for the bivariate beta-Sarmanov kernel (2.37).

2.3.3 Bivariate beta-Sarmanov kernel estimators

Standard version of the estimator

In this particular case, the beta-Sarmanov kernel estimator

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} BS_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{X}_i), \quad \forall \mathbf{x} \in [0,1] \times [0,1] \tag{2.40}$$

$\Lambda(n, \mathbf{H}, BS_\theta)$	sample 1	sample 2	sample 3	sample 4
h_{12} =-0.0003	1.034019	0.9954256	1.002369	1.025671
$h_{12} = 0$	1.034078	0.9955653	1.002418	1.025731
$h_{12} = 0.0004$	1.034157	0.9957517	1.002483	1.025811

TABLE 2.3 – Some values of $\Lambda(n; \mathbf{H}, BS_\theta)$ for $h_{11} = 0.10$, $h_{22} = 0.07$ and $n = 1000$.

also satisfies Proposition 2.2.10. Table 2.3 allows to observe the effect of correlation (2.38) on the total mass $\Lambda(n; \mathbf{H}, BS_\theta) \neq 1$ by using four samples of simulated data. Fixing $\mathbf{x} = (x_1, x_2)^\top$ in $[0,1] \times [0,1]$, the pointwise bias is written as

$$\begin{aligned}\text{Bias}\left\{\widehat{f_n}(\mathbf{x})\right\} &= a_{\theta 1}\frac{\partial f}{\partial x_1}(\mathbf{x}) + a_{\theta 2}\frac{\partial f}{\partial x_2}(\mathbf{x}) + \frac{1}{2}\left\{(a_{\theta 1}^2 + b_{\theta 11})\frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{x}) + 2(a_{\theta 1}a_{\theta 2} + b_{\theta 12})\frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x})\right. \\ &\quad \left. + (a_{\theta 2}^2 + b_{\theta 22})\frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{x})\right\} + \mathrm{o}(h_{11}^2 + 2h_{12}^2 + h_{22}^2);\end{aligned}$$

and, the pointwise variance is

$$\text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} = \frac{1}{n}\left\|BS_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x}) + \mathrm{o}\left(n^{-1}\,(\det\mathbf{H})^{-r_2}\right),$$

with

$$\begin{aligned}\left\|BS_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 &= \frac{\mathscr{B}\left(1 + 2x_1/h_{11}, 1 + (1-x_1)/h_{11}\right)\mathscr{B}\left(1 + 2x_2/h_{22}, 1 + (1-x_2)/h_{22}\right)}{\left\{\mathscr{B}\left(1 + x_1/h_{11}, 1 + (1-x_1)/h_{11}\right)\mathscr{B}\left(1 + x_2/h_{22}, 1 + (1-x_2)/h_{22}\right)\right\}^2} \\ &\quad \times \left\{1 + \frac{(2x_1-1)(2x_2-1)}{2(1+h_{11})(1+h_{22})}\left(\frac{(x_1+h_{11})^{-1}(1+3h_{11})(x_2+h_{22})^{-1}(1+3h_{22})}{(1-x_1+h_{11})(1-x_2+h_{22})}\right)^{1/2} h_{12}\right. \\ &\quad + \left(\frac{2x_1+h_{11}}{h_{11}(1+2h_{11})^{-1}} + \frac{2(x_1+h_{11})(1+h_{11})}{h_{11}^2(2+3h_{11})^{-1}}\right)\left(\frac{2x_2+h_{22}}{h_{22}(1+2h_{22})^{-1}} + \frac{2(x_2+h_{22})(1+h_{22})}{h_{22}^2(2+3h_{22})^{-1}}\right) \\ &\quad \left.\times\left(\frac{(1+h_{11})^{-1}(1+3h_{11})(1+h_{22})^{-1}(1+3h_{22})}{4(1-x_1+h_{11})(2+3h_{11})(1-x_2+h_{22})(2+3h_{22})}\right)h_{12}^2\right\}.\end{aligned}$$

Using (2.39) the AMISE (2.21) becomes here

$$\begin{aligned}\text{AMISE}\left(\widehat{f_n}\right) &= \int_{[0,1]\times[0,1]}\left(\left[a_{\theta 1}\frac{\partial f}{\partial x_1}(\mathbf{x}) + a_{\theta 2}\frac{\partial f}{\partial x_2}(\mathbf{x}) + \frac{1}{2}\left\{(a_{\theta 1}^2 + b_{\theta 11})\frac{\partial^2 f}{\partial x_1^2}(\mathbf{x})\right.\right.\right. \\ &\quad \left.\left.\left. + 2(a_{\theta 1}a_{\theta 2} + b_{\theta 12})\frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) + (a_{\theta 2}^2 + b_{\theta 22})\frac{\partial^2 f}{\partial x_2^2}(\mathbf{x})\right\}\right]^2 + \frac{1}{n}\left\|BS_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x})\right)d\mathbf{x}.\end{aligned}$$

The bandwidth matrix is selected by the LSCV method (2.25) on the set $\mathcal{D}$ of diagonal bandwidth matrices. Concerning full and Scott cases, this LSCV method is used under $\mathcal{H}_1$ and $\mathcal{S}_1$, respectively, subsets of $\mathcal{H}$ and $\mathcal{S}$ verifying the constraint of the beta-Sarmanov kernel (2.38). Their algorithms are described below and used for numerical studies in Section 2.4.

Algorithms of LSCV method (2.25) for three forms of bandwidth matrices in two dimensions ($d = 2$)

A1. Full bandwidth matrices.
   1. Choose two intervals H_{11} and H_{22} related to h_{11} and h_{22}, respectively.
   2. For $\delta = 1, \ldots, \ell(H_{11})$ and $\gamma = 1, \ldots, \ell(H_{22})$,
      (a) Compute the interval $H_{12}[\delta, \gamma]$ related to h_{12} from constraints (2.38) ;
      (b) For $\lambda = 1, \ldots, \ell(H_{12}[\delta, \gamma])$,
      Compose the full bandwidth matrix $\mathbf{H}(\delta, \gamma, \lambda) := \left(h_{ij}(\delta, \gamma, \lambda)\right)_{i,j=1,2}$ with $h_{11}(\delta, \gamma, \lambda) = H_{11}(\delta)$, $h_{22}(\delta, \gamma, \lambda) = H_{22}(\gamma)$ and $h_{12}(\delta, \gamma, \lambda) = H_{12}[\delta, \gamma](\lambda)$.
   3. Apply LSCV method on the set $\mathcal{H}_1$ of all full bandwidth matrices $\mathbf{H}(\delta, \gamma, \lambda)$.

A2. Scott bandwidth matrices.
   1. Choose an interval H related to h and a fixed bandwidth matrix $\mathbf{H}_0 = \left(h_{0ij}\right)_{i,j=1,2}$.
   2. For $\zeta = 1, \ldots, \ell(H)$,
      (a) Compute the interval $H_{012}[\zeta]$ related to h_{012} from constraints (2.38) ;
      (b) For $\kappa = 1, \ldots, \ell(H_{012}[\zeta])$,
      Compose the given bandwidth matrix $\mathbf{H}_0(\zeta, \kappa) := \left(h_{0ij}(\zeta, \kappa)\right)_{i,j=1,2}$ with $h_{011}(\zeta, \kappa) = h_{011}$, $h_{022}(\zeta, \kappa) = h_{022}$ and $h_{012}(\zeta, \kappa) = H_{012}[\zeta](\kappa)$;
      (c) Compose then the Scott bandwidth matrix $\mathbf{H}(\zeta, \kappa) := H(\zeta) \times \mathbf{H}_0(\zeta, \kappa)$.
   3. Apply LSCV method on the set $\mathcal{S}_1$ of all Scott bandwidth matrices $\mathbf{H}(\zeta, \kappa)$.

A3. Diagonal bandwidth matrices.
   1. Choose two intervals H_{11} and H_{22} related to h_{11} and h_{22}, respectively.
   2. For $\delta = 1, \ldots, \ell(H_{11})$ and $\gamma = 1, \ldots, \ell(H_{22})$,
      Compose the diagonal bandwidth matrix $\mathbf{H}(\delta, \gamma) := \mathbf{Diag}\left(H_{11}(\delta), H_{22}(\gamma)\right)$.
   3. Apply LSCV method on the set $\mathcal{D}$ of all diagonal bandwidth matrices $\mathbf{H}(\delta, \gamma)$.

Let us conclude these algorithms by the following precisions. For a given interval I, the notation $\ell(I)$ is the total number of subdivisions of I and $I(\eta)$ denotes the real value at the subdivision η of I. Also, for practical uses of (A1) and (A3), both intervals H_{11} and H_{22} are generally chosen to be $(0, 1)$. In the case of the Scott bandwidth matrix (A2), we retain the interval $H = (0, 2)$ and the fixed bandwidth matrix $\mathbf{H}_0 = \widehat{\mathbf{\Sigma}}$, where $\widehat{\mathbf{\Sigma}}$ is the sample covariance matrix. See Figure 2.3 for graphical illustrations.

Modified version of the estimator

Being large in the standard version, the pointwise bias of the beta-Sarmanov kernel estimator (2.40) must be reduced. Following the algorithm of Section 2.2.3 and without numerical illustration in this paper, the first step divides $[0, 1] \times [0, 1]$ in nine subregions of order $\boldsymbol{\alpha}(\mathbf{H}) = (\alpha_1(h_{11}), \alpha_2(h_{22}))^\top$ with $\alpha_j(h_{jj}) > 0$ for $j = 1, 2$:

a. only one interior subregion denoted as $\mathbb{T}_2^{\boldsymbol{\alpha}(\mathbf{H}),I} = (\alpha_1(h_{11}), 1 - \alpha_1(h_{11})) \times (\alpha_2(h_{22}), 1 - \alpha_2(h_{22}))$;

b. eight boundary subregions divided in two parts as

(i) four angle subregions denoted by

$$\begin{aligned}\mathbb{T}_2^{\boldsymbol{\alpha}(\mathbf{H}),A} &= [0,\alpha_1(h_{11})]\times[0,\alpha_2(h_{22})]\cup[1-\alpha_1(h_{11}),1]\times[0,\alpha_2(h_{22})]\\ &\quad\cup[0,\alpha_1(h_{11})]\times[1-\alpha_2(h_{22}),1]\cup[1-\alpha_2(h_{22}),1]\times[1-\alpha_2(h_{22}),1],\end{aligned}$$

(ii) four edge subregions denoted by

$$\begin{aligned}\mathbb{T}_2^{\boldsymbol{\alpha}(\mathbf{H}),E} &= (\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times[1-\alpha_2(h_{22}),1]\cup[0,\alpha_1(h_{11})]\times(\alpha_2(h_{22}),1-\alpha_2(h_{22}))\\ &\quad\cup(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times[0,\alpha_2(h_{22})]\\ &\quad\cup[1-\alpha_1(h_{11}),1]\times(\alpha_2(h_{22}),1-\alpha_2(h_{22})).\end{aligned}$$

As for the second step, we consider the three functions $\psi_1,\psi_2:[0,1]\to[0,1]$ and $\psi_3:\mathcal{H}\to\mathbb{R}$ such that

$$\psi_j(z_j)=\alpha_j(h_{jj})\{z_j-\alpha_j(h_{jj})+1\},\ \ \forall z_j\in[0,1],\ \ j=1,2 \text{ and } \psi_3(\mathbf{H})=\frac{h_{12}}{\sqrt{h_{11}h_{22}}}. \tag{2.41}$$

Each axis of $[0,1]\times[0,1]$ has one interior subregion $(\alpha_j(h_{jj}),1-\alpha_j(h_{jj}))$ and two boundary regions $[0,\alpha_j(h_{jj})]$ and $[1-\alpha_1(h_{jj}),1]$ for $j=1,2$. Thus, from (2.27) with $d=1$, one gets the new parametrization of each margin beta kernel used in $BS_{\theta(\mathbf{x},\mathbf{H})}$ of (2.37) with $\mathbf{x}=(x_1,x_2)^\top$: for $j=1,2$,

$$\begin{cases}\left(\frac{\psi_j(x_j)}{h_{jj}},\frac{x_j}{h_{jj}}\right) & \text{if } x_j\in[0,\alpha_j(h_{jj})],\\ \left(\frac{x_j}{h_{jj}},\frac{1-x_j}{h_{jj}}\right) & \text{if } x_j\in(\alpha_j(h_{jj}),1-\alpha_j(h_{jj})),\\ \left(\frac{1-x_j}{h_{jj}},\frac{\psi_j(1-x_j)}{h_{jj}}\right) & \text{if } x_j\in[1-\alpha_i(h_{jj}),1].\end{cases} \tag{2.42}$$

Therefore, using ψ_3 of (2.41) and by combination of (4.2.2) for each of the nine subregions of $[0,1]\times[0,1]$, then $\widetilde{\theta}$ is expressed by

$$\widetilde{\theta}(\mathbf{x},\mathbf{H})=\begin{cases}\left(\frac{\psi_1(x_1)}{h_{11}},\frac{x_1}{h_{11}},\frac{\psi_2(x_2)}{h_{22}},\frac{x_2}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times[0,\alpha_2(h_{22})]\\ \left(\frac{\psi_1(x_1)}{h_{11}},\frac{x_1}{h_{11}},\frac{1-x_2}{h_{22}},\frac{\psi_2(1-x_2)}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times[1-\alpha_2(h_{22}),1]\\ \left(\frac{1-x_1}{h_{11}},\frac{\psi_1(1-x_1)}{h_{11}},\frac{\psi_2(x_2)}{h_{22}},\frac{x_2}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times[0,\alpha_2(h_{11})]\\ \left(\frac{1-x_1}{h_{11}},\frac{\psi_1(1-x_1)}{h_{11}},\frac{1-x_2}{h_{22}},\frac{\psi_2(1-x_2)}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times[1-\alpha_2(h_{22}),1]\\ \left(\frac{x_1}{h_{11}},\frac{1-x_1}{h_{11}},\frac{x_2}{h_{22}},\frac{1-x_2}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)=\widetilde{\theta}_I(\mathbf{x},\mathbf{H}) & \text{if } \mathbf{x}\in(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times(\alpha_2(h_{22}),1-\alpha_2(h_{22}))\\ \left(\frac{\psi_1(x_1)}{h_{11}},\frac{x_1}{h_{11}},\frac{x_2}{h_{22}},\frac{1-x_2}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times(\alpha_2(h_{22}),1-\alpha_2(h_{22}))\\ \left(\frac{x_1}{h_{11}},\frac{1-x_1}{h_{11}},\frac{\psi_2(x_2)}{h_{22}},\frac{1-x_2}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top & \text{if } \mathbf{x}\in(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times[0,\alpha_2(h_{22})]\\ \left(\frac{x_1}{h_{11}},\frac{1-x_1}{h_{11}},\frac{1-x_2}{h_{22}},\frac{\psi_2(1-x_2)}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top & \text{if } \mathbf{x}\in(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times[1-\alpha_2(h_{22}),1]\\ \left(\frac{1-x_1}{h_{11}},\frac{\psi_1(1-x_1)}{h_{11}},\frac{x_2}{h_{22}},\frac{1-x_2}{h_{22}},\frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times(\alpha_2(h_{22}),1-\alpha_2(h_{22})).\end{cases} \tag{2.43}$$

Notice that the last component $h_{12}(h_{11}h_{22})^{-1/2}$ of $\widetilde{\theta}$ does not change in all subregions of the support. This is because of the choice of $\mathbf{m}_0$ related to the construction of the beta-Sarmanov kernel (2.37). According to Proposition 2.2.14 the modified beta-Sarmanov

kernel obtained from (2.43) and denoted by $BS_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}$ is an associated kernel with $\mathbb{S}_{\widetilde{\theta}(\mathbf{x},\mathbf{H})} = [0,1]\times[0,1]$,

$$\mathbf{a}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H}) = \begin{cases} \left(\frac{(1-x_1)\psi_1(x_1)+x_1^2}{x_1+\psi_1(x_1)}, \frac{(1-x_2)\psi_2(x_2)+x_2^2}{x_2+\psi_2(x_2)}\right)^\top & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times[0,\alpha_2(h_{22})] \\ \left(\frac{(1-x_1)\psi_1(x_1)+x_1^2}{x_1+\psi_1(x_1)}, 0\right)^\top & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times(\alpha_2(h_{22}),1-\alpha_2(h_{22})) \\ \left(\frac{(1-x_1)\psi_1(x_1)+x_1^2}{x_1+\psi_1(x_1)}, \frac{(1-x_2)\{x_2-\psi_2(1-x_2)\}}{1-x_2+\psi_2(1-x_2)}\right)^\top & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times[1-\alpha_2(h_{22}),1] \\ \left(0, \frac{(1-x_2)\psi_2(x_2)+x_2^2}{x_2+\psi_2(x_2)}\right)^\top & \text{if } \mathbf{x}\in(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times[0,\alpha_2(h_{22})] \\ (0,0)^\top & \text{if } \mathbf{x}\in(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times(\alpha_2(h_{22}),1-\alpha_2(h_{22})) \\ \left(0, \frac{(1-x_2)\{x_2-\psi_2(1-x_2)\}}{1-x_2+\psi_2(1-x_2)}\right)^\top & \text{if } \mathbf{x}\in(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times[1-\alpha_2(h_{22}),1] \\ \left(\frac{(1-x_1)\{x_1-\psi_1(1-x_1)\}}{1-x_1+\psi_1(1-x_1)}, \frac{(1-x_2)\psi_2(x_2)+x_2^2}{x_2+\psi_2(x_2)}\right)^\top & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times[0,\alpha_2(h_{11})] \\ \left(\frac{(1-x_1)\{x_1-\psi_1(1-x_1)\}}{1-x_1+\psi_1(1-x_1)}, 0\right)^\top & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times(\alpha_2(h_{22}),1-\alpha_2(h_{22})) \\ \left(\frac{(1-x_1)\{x_1-\psi_1(1-x_1)\}}{1-x_1+\psi_1(1-x_1)}, \frac{(1-x_2)\{x_2-\psi_2(1-x_2)\}}{1-x_2+\psi_2(1-x_2)}\right)^\top & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times[1-\alpha_2(h_{22}),1] \end{cases} \tag{2.44}$$

and, $\mathbf{B}_{\widetilde{\theta}}(\mathbf{x},\mathbf{H}) = (b_{\widetilde{\theta}ij})_{i,j=1,2}$ such that $b_{\widetilde{\theta}12} = \left(h_{12}/(h_{11}h_{22})^{1/2}\right) b_{\widetilde{\theta}11} b_{\widetilde{\theta}22}$ and $\left(b_{\widetilde{\theta}11}, b_{\widetilde{\theta}22}\right)$ is detailed as

$$\begin{cases} \left(\frac{h_{11}x_1\psi_1(x_1)\{x_1+\psi_1(x_1)\}^{-2}}{\{x_1+\psi_1(x_1)+h_{11}\}}, \frac{h_{22}x_2\psi_2(x_2)\{x_2+\psi_2(x_2)\}^{-2}}{\{x_2+\psi_2(x_2)+h_{22}\}}\right) & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times[0,\alpha_2(h_{22})] \\ \left(\frac{h_{11}x_1\psi_1(x_1)\{x_1+\psi_1(x_1)\}^{-2}}{\{x_1+\psi_1(x_1)+h_{11}\}}, \frac{h_{22}x_2(1-x_2)}{1+h_{22}}\right) & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times(\alpha_2(h_{22}),1-\alpha_2(h_{22})) \\ \left(\frac{h_{11}x_1\psi_1(x_1)\{x_1+\psi_1(x_1)\}^{-2}}{\{x_1+\psi_1(x_1)+h_{11}\}}, \frac{h_{22}(1-x_2)\{1-x_2+\psi_2(1-x_2)\}^{-2}}{\{\psi_2(1-x_2)\}^{-1}\{1-x_2+\psi_2(1-x_2)+h_{22}\}}\right) & \text{if } \mathbf{x}\in[0,\alpha_1(h_{11})]\times[1-\alpha_2(h_{22}),1] \\ \left(\frac{h_{11}x_1(1-x_1)}{1+h_{11}}, \frac{h_{22}x_2\psi_2(x_2)\{x_2+\psi_2(x_2)\}^{-2}}{\{x_2+\psi_2(x_2)+h_{22}\}}\right) & \text{if } \mathbf{x}\in(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times[0,\alpha_2(h_{22})] \\ \left(\frac{h_{11}x_1(1-x_i)}{1+h_{11}}, \frac{h_{22}x_2(1-x_2)}{1+h_{22}}\right) & \text{if } \mathbf{x}\in\mathbb{T}_2^{\alpha(\mathbf{H}),I} \\ \left(\frac{h_{11}x_1(1-x_1)}{1+h_{11}}, \frac{h_{22}(1-x_2)\{1-x_2+\psi_2(1-x_2)\}^{-2}}{\{\psi_2(1-x_2)\}^{-1}\{1-x_2+\psi_2(1-x_2)+h_{22}\}}\right) & \text{if } \mathbf{x}\in(\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times[1-\alpha_2(h_{22}),1] \\ \left(\frac{h_{11}(1-x_1)\{1-x_1+\psi_1(1-x_1)\}^{-2}}{\{\psi_1(1-x_1)\}^{-1}\{1-x_1+\psi_1(1-x_1)+h_{11}\}}, \frac{h_{22}x_2\psi_2(x_2)\{x_2+\psi_2(x_2)\}^{-2}}{\{x_2+\psi_2(x_2)+h_{22}\}}\right) & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times[0,\alpha_2(h_{11})] \\ \left(\frac{h_{11}(1-x_1)\{1-x_1+\psi_1(1-x_1)\}^{-2}}{\{\psi_1(1-x_1)\}^{-1}\{1-x_1+\psi_1(1-x_1)+h_{11}\}}, \frac{h_{22}x_2(1-x_2)}{1+h_{22}}\right) & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times(\alpha_2(h_{22}),1-\alpha_2(h_{22})) \\ \left(\frac{h_{11}(1-x_1)\{1-x_1+\psi_1(1-x_1)\}^{-2}}{\{\psi_1(1-x_1)\}^{-1}\{1-x_1+\psi_1(1-x_1)+h_{11}\}}, \frac{h_{22}(1-x_2)\{1-x_2+\psi_2(1-x_2)\}^{-2}}{\{\psi_2(1-x_2)\}^{-1}\{1-x_2+\psi_2(1-x_2)+h_{22}\}}\right) & \text{if } \mathbf{x}\in[1-\alpha_1(h_{11}),1]\times[1-\alpha_2(h_{22}),1]. \end{cases}$$

The corresponding modified beta-Sarmanov kernel estimator

$$\widetilde{f_n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} BS_{\widetilde{\theta}(\mathbf{x},\mathbf{H})}(X_i), \quad \forall \mathbf{x}\in[0,1]\times[0,1] \tag{2.45}$$

has, for all $\mathbf{x}\in\mathbb{T}_2^{\boldsymbol{\alpha}(\mathbf{H}),I} = (\alpha_1(h_{11}),1-\alpha_1(h_{11}))\times(\alpha_2(h_{22}),1-\alpha_2(h_{22}))$,

$$\text{Bias}\left\{\widetilde{f_n}(\mathbf{x})\right\} = \frac{1}{2}\left\{b_{\widetilde{\theta}_I 11}\frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) + 2b_{\widetilde{\theta}_I 12}\frac{\partial^2 f}{\partial x_2\partial x_2}(\mathbf{x}) + b_{\widetilde{\theta}_I 22}\frac{\partial^2 f}{\partial x_2^2}(\mathbf{x})\right\} + \text{o}(h_{11}^2+2h_{12}^2+h_{22}^2)$$

and

$$\text{Var}\left\{\widetilde{f_n}(\mathbf{x})\right\} \simeq \text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} \text{ as } n\to\infty.$$

Thus, the asymptotic expression of the MISE of $\widetilde{f_n}$ on $\mathbb{T}_2^{\boldsymbol{\alpha}(\mathbf{H}),I}$ is given by

$$\begin{aligned}\mathrm{AMISE}_{\widetilde{\theta}_I}\left(\widetilde{f_n}\right) &= \int_{\mathbb{T}_2^{\boldsymbol{\alpha}(\mathbf{H}),I}}\left(\left[\frac{1}{2}\left\{b_{\widetilde{\theta}_I 11}\frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) + 2b_{\widetilde{\theta}_I 12}\frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{x}) + b_{\widetilde{\theta}_I 22}\frac{\partial^2 f}{\partial x_2^2}(\mathbf{x})\right\}\right]^2\right.\\ &\quad \left.+\frac{1}{n}\left\|BS_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x})\right)d\mathbf{x}.\end{aligned}$$

The modified beta-Sarmanov kernel also depend on the choice of scalars $\alpha_j(h_{jj})$ for $j = 1, 2$. The user can set the values of $\alpha_j(h_{jj})$ according to his practical objective. For example in univariate case, Chen (1999, 2000) took $\alpha_j(h_{jj}) = 2h_{jj}$. From (2.36) to (2.45) and when $h_{12} = 0$, we have the same formulas for multiple associated kernels of Bouerzmarni and Rombouts (2010). Also, similar results can be obtained for the Scott bandwidth matrices. However, numerical illustrations are so long and tedious tasks ; see, e.g., Hirukawa and Sakudo (2014) for $d = 1$.

2.4 Simulation studies and real data analysis

In this section, we compare the performance of the three forms of bandwidth matrices of Table 2.1. The optimal bandwidth matrix is chosen by LSCV method (2.25) using the algorithms A1, A2 and A3 given at the end of Section 2.3.3 and their indications. All computations were done on the computational resource [1] of Laboratoire de Mathématiques de Besançon by using the R software ; see R Development Core Team (2012). The comparisons will be done using the standard version of the beta-Sarmanov kernel estimator (2.40) through simulations studies and an illustration on real dataset.

2.4.1 Simulation studies

We consider six target densities with supports included in $[0,1]\times[0,1]$ and labeled A, B, C, D, E and F respectively. They have different correlation structure and some local modes. The plots for these densities are given in Figure 2.2.

- Density A is the bivariate beta density without correlation $\rho = 0$ such that $(p_1, q_1) = (3,3)$ and $(p_2, q_2) = (5,5)$ as parameters values in univariate beta density (2.30), respectively ;
- density B is the bivariate Dirichlet density

$$f(v_1, v_2) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} v_1^{\alpha_1-1} v_2^{\alpha_2-1}(1-v_1-v_2)^{\alpha_3-1}, \quad v_1, v_2 \geq 0, v_1+v_2 \leq 1,$$

  where $\Gamma(\cdot)$ is the classical gamma function, with parameters values $\alpha_1 = \alpha_2 = 2$, $\alpha_3 = 7$ and, therefore, the moderate value of $\rho = -(\alpha_1\alpha_2)^{1/2}(\alpha_1+\alpha_3)^{-1/2}(\alpha_2+\alpha_3)^{-1/2} = -0.2222$;
- density C is the bivariate beta density without correlation $\rho = 0$ such that $(p_1, q_1) = (3,2)$ and $(p_2, q_2) = (2,5)$ in (2.30), respectively ;

1. Dell Poweredge R900, Processor Xeon X7350, 2.93 GHz, 32 Go RAM

– density D is the bivariate Dirichlet as the density B but with $\alpha_1 = \alpha_2 = 10$, $\alpha_3 = 3$ and then $\rho = -0.7692$.
– density E is the bivariate density without correlation $\rho = 0$ defined as follows :

$$f(v_1, v_2) = [(3/7)g_1(v_1) + (4/7)g_2(v_1)] \times g_3(v_2)$$

such that g_1, g_2 and g_3 are univariate beta densities (2.30) with parameters values $(p_1, q_1) = (2, 7)$ and $(p_2, q_2) = (7, 2)$ and $(p_3, q_3) = (6, 6)$.
– density F is the bivariate density without correlation $\rho = 0$:

$$f(v_1, v_2) = [(8/11)g_1(v_1) + (3/11)g_2(v_1)] \times [(5/7)g_3(v_2) + (2/7)g_4(v_2)]$$

such that g_1, g_2, g_3 and g_4 are univariate beta densities (2.30) with parameters values $(p_1, q_1) = (3.5, 7)$ and $(p_2, q_2) = (7, 3.5)$, $(p_3, q_3) = (7, 2)$ and $(p_4, q_4) = (2, 7)$.

Table 2.4 presents the execution times needed for computing the LSCV method for each of the three types of bandwidth matrix with respect to only one replication of sample sizes $n = 100$ for each of the target densities A, B, C, D, E and F of Figure 2.2. For $n = 100$ the computational times of the LSCV method for full bandwidth matrix are longer than both the Scott and diagonal bandwidth matrices, which have the almost identical Central Processing Unit (CPU) times. The constraints (2.38) of the correlation of Sarmanov induce that execution times of the Scott bandwidth matrix are relatively a bit longer than for the diagonal ones. Let us note that for these bandwidth matrix types, the CPU times can be considerably reduced by parallelism processing, in particular for for the full LSCV method. These constraints (2.38) reflect the difficulty of finding the appropriate bandwidth matrix with correlation structure by LSCV method (2.25). Hence, we are able to infer that the Scott and diagonal LSCV method do not impose excessive computation burdens ; and, the Scott procedure takes into account the structure of (null and moderate) correlations in the sample.

n	**H**	A	B	C	D	E	F
	Full	2.8310	2.7570	2.8793	2.7931	27940	2.8362
100	Scott	0.2173	0.2251	0.2207	0.2202	0.2212	0.2204
	Diagonal	0.1436	0.1456	0.1526	0.1536	1.1531	1.1446

Table 2.4 – Typical CPU times (in hours) for one replication of LSCV method (2.25) by using the algorithms A1, A2 and A3 given at the end of Section 2.3.3.

We now examine the efficiency of various bandwidth matrices in Table 2.1 via

$$\widehat{ISE} = \frac{1}{N}\sum\nolimits_{m=1}^{N} \int_{[0,1]\times[0,1]} \left\{\widehat{f_n}(\mathbf{x}) - f(\mathbf{x})\right\}^2 d\mathbf{x},$$

where $N = 100$ is the number of replications. Table 2.5 shows some expected values of $\widehat{ISE}$ for the three forms of bandwidth matrices with respect to the densities A, B, C, D, E and F of Figure 2.2, and according only to the sample size $n = 100$ because of excess of computational times (see Table 2.4). Globally, the full and Scott bandwidth matrices

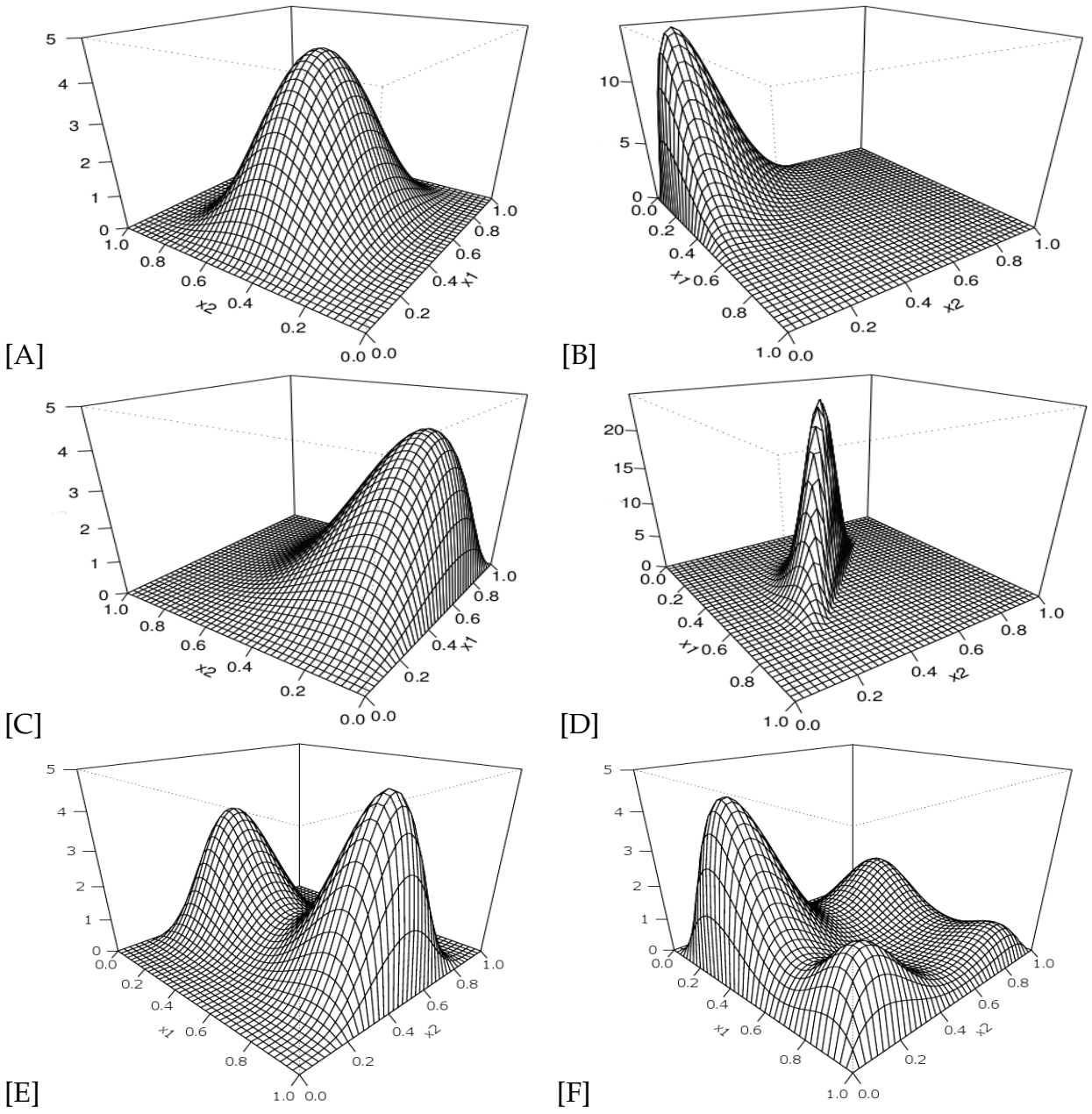


FIGURE 2.2 – Six plots of density functions defined at the begining of Section 2.4.1 and considered for simulations.

with correlation structure perform better in terms of the quality of smoothing than the diagonal one without correlation. Even if the correlation is almost non-existent in the sample (e.g. models A and C of Table 2.5), we attend the good behavior of the full and Scott bandwidth matrices. Also, these bandwidth matrices with correlation structure suit for multimodal target densities (e.g. E and F of Table 2.5). For moderate correlation (e.g. models B of Table 2.5) we can recommend the light version of bandwidth matrices with correlation structure which is the Scott one. As for strong correlation in the sample (e.g. models D of Table 2.5) it is not preferable to use the Scott bandwidth matrix.

Models	Full	Scott	Diagonal
A	0.2554(0.2253)	0.1887(0.0890)	0.2963(0.2121)
B	0.8571(0.2422)	0.5999(0.2582)	0.9743(0.5973)
C	0.1985(0.0737)	0.2150(0.0828)	0.2384(0.1740)
D	1.1481(0.0937)	9.1188(0.7466)	1.6420(0.5735)
E	0.2786(0.0785)	0.2498(0.0985)	0.5056(0.1526)
F	0.6025(0.1122)	0.6791(0.1475)	0.8025(0.1254)

TABLE 2.5 – Expected values (and their standard deviation) of $\widehat{ISE}$ with $N = 100$ replications of sample sizes $n = 100$ using three types of bandwidth matrices $\widehat{\mathbf{H}}$ (2.25) for each of four models of Figure 2.2.

Finally, Tables 2.4 and 2.5 indicate that the choice of the Scott bandwidth matrix using LSCV method is a good alternative to the full one in the purpose of preserving a correlation structure in the bandwidth matrix for an associated kernel estimator.

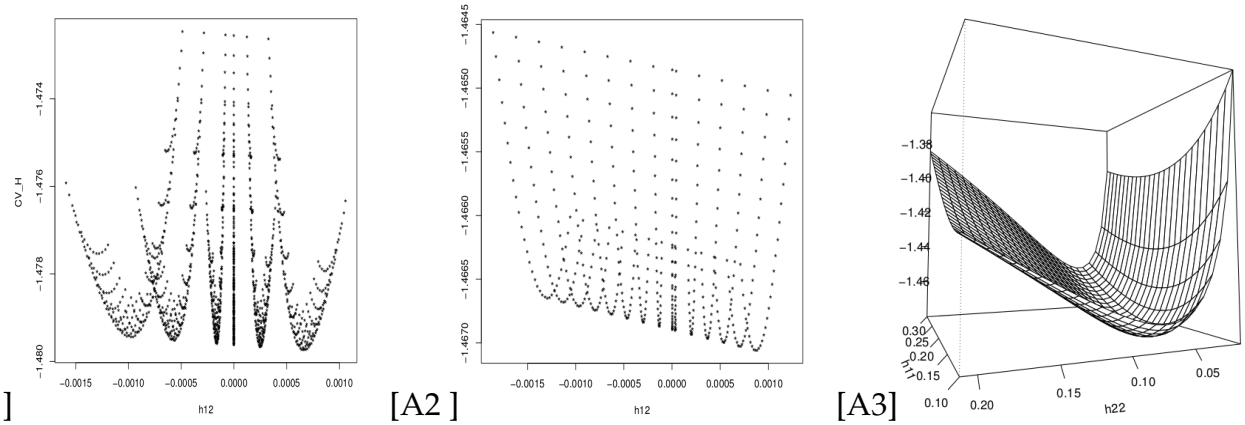
2.4.2 Real data analysis

We applied the standard version of the beta-Sarmanov estimator (2.40) on paired rates data set according to the three types of bandwidth matrices in Table 2.1. The dataset of sample size $n = 80$ in Graph (o) of Figure 2.4 has been provided by Francial G. Libengué (2013) during his last stay in Burkina Faso. It represents the popular ratings, for the two first consecutive ballots of the same electoral mandate of five years, of a political figure in different departments. Note that, for both elections, the prominent politician was finally elected in the second round. We are here interested to the opposite behavior of peoples during first rounds of both elections since the second round is governed by political alliances, which do not constitute a reference for the own popularity of a candidate.

Indeed, for many African countries, political elections are generally fought on tribal ethnic origins and partisan interests. The data displays opposed viewpoints between the results of the first round of the first election (x_1) and the first round of the second one (x_2). In fact, the first election saw the candidate program mostly adopted by its clan (tribe and allied tribes) and rejected by the others. A few years later, facing the social discontent, he assumed a new political program which ends to another consultation

of the people in a time x_2. This reversal made him loose the support of his supporters but he received the backing up of formers opponents. It is thus noticeable that its own popularity is not much different between the first round of both elections ; the first gave an average $\overline{x}_1 = 0.4915$ and the second $\overline{x}_2 = 0.4126$. However, there is a significant negative correlation in the dataset : $\widehat{\rho} = -0.6949$. Unlike overall trends $\overline{x}_1$, $\overline{x}_2$ and $\widehat{\rho}$, the empirical distribution Graph (o) of Figure 2.4 gives more details of the electoral situation department by department. Hence, we need a nonparametric smoothing of this joint distribution by using associated kernels.

In order to smooth the joint empirical distribution of these paired data, we apply the beta-Sarmanov kernel estimator in its standard version (2.40). Figure 2.3 shows the results of the LSCV algorithms A1, A2 and A3 with the ratings dataset ; see (2.25) and the end of Section 2.3.3. The computation time of the LSCV is in the same trend as in Table 2.4 for $n = 100$. To simplify the presentations in Figure 2.3 for both the Scott and full bandwidth matrices, we only plot $h_{12} \mapsto LSCV(\mathbf{H})$ for some values of h_{11} and h_{22}. In all cases we observe that there is a global minimum. The obtained optimal bandwidth



[A1] [A2] [A3]

FIGURE 2.3 – Some plots of $\mathbf{H} \mapsto LSCV(\mathbf{H})$ from algorithms at the end of Section 2.3.3 for real data in Graph (o) of Figure 2.4 : (A1) full $\widehat{\mathbf{H}}$, (A2) Scott $\widehat{\mathbf{H}}$ and (A3) diagonal $\widehat{\mathbf{H}}$.

matrices are

$$\widehat{\mathbf{H}}_{\text{full}} = \begin{pmatrix} 0.160000 & 0.000655 \\ 0.000655 & 0.057500 \end{pmatrix},$$

$$\widehat{\mathbf{H}}_{\text{Scott}} = 1.405 \times \begin{pmatrix} 0.076039 & 0.000622 \\ 0.000622 & 0.073994 \end{pmatrix} = \begin{pmatrix} 0.106836 & 0.000874 \\ 0.000874 & 0.103962 \end{pmatrix}$$

and $\widehat{\mathbf{H}}_{\text{diagonal}} = \mathbf{Diag}(0.160000, 0.057500)$. The resulting estimates are displayed in Figure 2.4.

From simulation studies of previous section, the full bandwidth matrix provides the reference smoothing which is appropriated for correlated data ; see Graph (a) of Figure 2.4. In record time, the Scott bandwidth matrix $\widehat{\mathbf{H}}_{\text{Scott}}$ delivers similar smoothing in Graph (b) of Figure 2.4 as the full and diagonal ones (see respectively Graphs (a) and (c) of Figure 2.4). In conclusion, we found anywhere the shape of a "carpet flying" in balance, smoothing thus the joint empirical distribution of the electoral situation of

the candidate. This balance situation makes him to win in the second round of both elections.

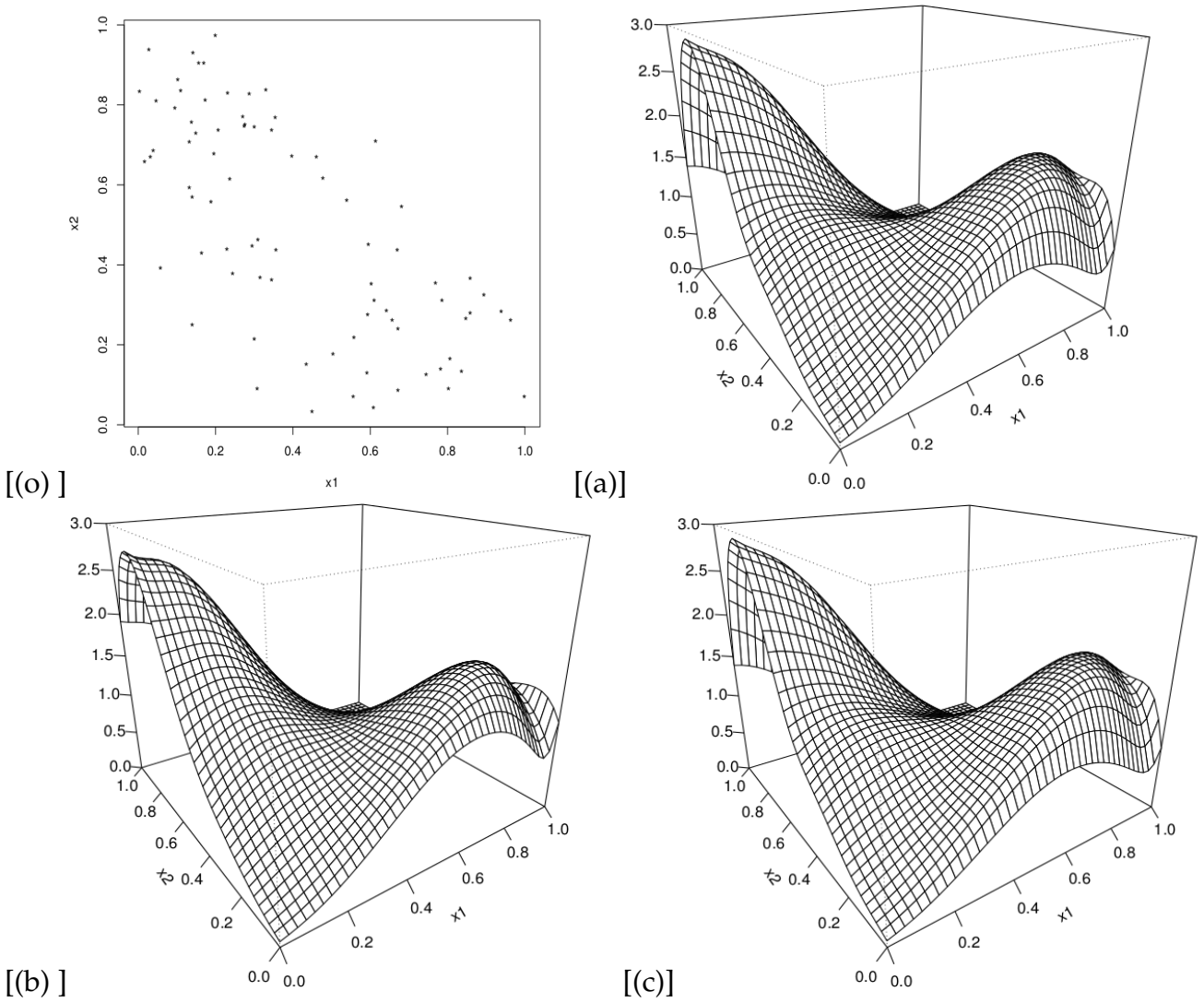


FIGURE 2.4 – Graphical representations (o) of the real dataset with $n = 80$, $\overline{x}_1 = 0.4915$, $\overline{x}_2 = 0.4126$, $\widehat{\sigma}_1 = 0.2757$, $\widehat{\sigma}_2 = 0.2720$, $\widehat{\rho} = -0.6949$ and its corresponding smoothings according to the choice of bandwidth matrix $\widehat{\mathbf{H}}$ by LSCV : (a) full, (b) Scott and (c) diagonal.

2.5 Summary and final remarks

We have presented general associated kernels (with or without correlation structure) that varying their shape according to the target point along the support. Excluding the classical associated kernels, the local adaptability of these associated kernels (depending on the target point $\mathbf{x}$ and the bandwidth matrix $\mathbf{H}$) means that they are free of boundary bias but have a slightly bias different. Furthermore, the forms of bandwidth

matrices used in the case with correlation structure have both theoretical and practical significances. Under the criterion of cross-validation, we therefore recommend the Scott bandwidth matrix which is more workable than the full one. A method of construction, called *multivariate mode dispersion* method, for these kernels are introduced. Also, we have proposed an algorithm of bias-reductions of their corresponding associated kernel estimators. An extension to discrete multivariate associated kernels is obviously possible. Similarly, a work is in progress on nonparametric multiple regression composed by continuous and discrete univariate associated kernels (e.g. Proposition 2.2.6).

Constructed by the correlation structure of Sarmaanov (1966) and by a variant of mode dispersion method, the bivariate beta-Sarmanov kernel estimator is completely study with the optimal bandwidth matrix chosen by cross-validation. This technique can be extended in multivariate case for different type of kernels which are continuous and also discrete. In fact, from two or more univariate independent pdfs or probability mass functions, the correlation structure of Sarmanov (1966) and a variant of mode dispersion method can always allow to build a multivariate type of kernel with correlation structure ; and, therefore, produced the corresponding multivariate Sarmanov kernel. In terms of execution times from using the cross-validation method, we advise to use the Scott bandwidth matrix because of its flexibility and efficiency for no very strong correlation in the dataset.

Simulation experiments and analysis of a real dataset provide insight into the behavior of the bandwidth matrix for small and moderate sample sizes. Tables 2.4 and 2.5 and Figure 2.4 can be conceptually summarized as follows. As expected, the full bandwidth matrix is frequently better than the others. An alternative with correlation structure has been proposed for the cross-validation technique : the Scott bandwidth matrix which has a comparable gain in execution times than the diagonal one ; also, it produces better results than the diagonal in most cases. So, we recommend the Scott bandwidth matrix for multivariate use with the cross-validation technique. Further research in this direction are in progress, especially on the choice of optimal bandwidth matrix by using Bayesian approaches ; see, e.g., Zougab *et al.* (2014).

2.6 Appendix

This section is dedicated to the proof of Proposition 2.2.11. Indeed, using successively (2.16) and Taylor's formula around $\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)$ and then $\mathbf{x}$, and also the invariance

under cyclic permutations of the operator trace, the result (2.18) is shown by

$$\begin{aligned}
\text{Bias}\left\{\widehat{f_n}(\mathbf{x})\right\} &= \mathbb{E}\left\{f\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right\} - f(\mathbf{x}) \\
&= f\left(\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right) + \frac{1}{2}\mathbb{E}\left[\text{trace}\left\{\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)^{\top}\nabla^2 f\left(\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\right. \\
&\quad \left.\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\right\}\Big] - f(\mathbf{x}) + \text{o}\left[\mathbb{E}\left\{\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)^{\top}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\right\}\right] \\
&= f\left(\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right) - f(\mathbf{x}) + \frac{1}{2}\,\text{trace}\left[\nabla^2 f\left(\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\mathbb{E}\left\{\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)^{\top}\right. \\
&\quad \left.\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\right\}\Big] + \text{o}\left[\text{trace}\left(\mathbb{E}\left\{\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(x,\mathbf{H})}\right)\right)^{\top}\right\}\right)\right] \\
&= f\left(\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right) - f(\mathbf{x}) + \frac{1}{2}\,\text{trace}\left[\nabla^2 f\left(\mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\text{Cov}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right] + \text{o}\left[\text{trace}\left\{\text{Cov}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right\}\right] \\
&= f\left(\mathbf{x} + \mathbf{a}_\theta(\mathbf{x},\mathbf{H})\right) - f(\mathbf{x}) + \frac{1}{2}\,\text{trace}\left[\nabla^2 f\left(\mathbf{x} + \mathbf{a}_\theta(\mathbf{x},\mathbf{H})\right)\mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right] + \text{o}\left[\text{trace}\left\{\mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right\}\right] \\
&= \mathbf{a}_\theta^{\top}(\mathbf{x},\mathbf{H})\nabla f(\mathbf{x}) + \frac{1}{2}\,\text{trace}\left[\left\{\mathbf{a}_\theta(\mathbf{x},\mathbf{H})\mathbf{a}_\theta^{\top}(\mathbf{x},\mathbf{H}) + \mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right\}\nabla^2 f(\mathbf{x})\right] + \text{o}\left\{\text{trace}\left(\mathbf{H}^2\right)\right\}.
\end{aligned}$$

In fact, the rest $\text{o}\{\text{trace}(\mathbf{H}^2)\}$ comes from $\text{trace}(\mathbf{B}_\theta(\mathbf{x},\mathbf{H})) = O(\text{trace}\,\mathbf{H}^2)$ deduced from Proposition 2.2.4 of classical associated kernels and

$$\begin{aligned}
&\text{o}\left[\mathbb{E}\left\{\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)^{\top}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\right\}\right] \;= \\
&\mathbb{E}\left\{\text{o}_p\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)^{\top}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})} - \mathbb{E}\left(\mathcal{Z}_{\theta(\mathbf{x},\mathbf{H})}\right)\right)\right\},
\end{aligned}$$

where $\text{o}_p(\cdot)$ is the probability rate of convergence.

Concerning the variance (2.19) one first has

$$\begin{aligned}
\text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} &= \frac{1}{n}\mathbb{E}\left\{K^2_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{X}_1)\right\} - \frac{1}{n}\left[\mathbb{E}\left\{K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{X}_1)\right\}\right]^2 \\
&= \frac{1}{n}\int_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}\cap\mathbb{T}_d} K^2_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{u})f(\mathbf{u})d\mathbf{u} - \frac{1}{n}\left[\mathbb{E}\left\{K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{X}_1)\right\}\right]^2 \\
&= I_1 - I_2.
\end{aligned}$$

From (2.16) and (2.18), one has the following behavior of the second term

$$I_2 := (1/n)\left[\mathbb{E}\left\{K_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{X}_1)\right\}\right]^2 \simeq (1/n)f^2(\mathbf{x}) \simeq O(1/n)$$

since f is bounded for all $\mathbf{x} \in \mathbb{T}_d$. By using Taylor's expansion around of $\mathbf{x}$, the first term $I_1 := (1/n)\int_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}\cap\mathbb{T}_d} K^2_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{u})f(\mathbf{u})d\mathbf{u}$ becomes

$$I_1 = \frac{1}{n}f(\mathbf{x})\int_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}\cap\mathbb{T}_d} K^2_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{u})d\mathbf{u} + R(\mathbf{x},\mathbf{H}),$$

with

$$\begin{aligned}
R(\mathbf{x},\mathbf{H}) &= \frac{1}{n}\int_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}\cap\mathbb{T}_d} K^2_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{u})\left[(\mathbf{u}-\mathbf{x})^{\top}\nabla f(\mathbf{x}) + \frac{1}{2}(\mathbf{u}-\mathbf{x})^{\top}\nabla^2 f(\mathbf{x})(\mathbf{u}-\mathbf{x})\right. \\
&\quad \left. + \text{o}\left\{(\mathbf{u}-\mathbf{x})^{\top}(\mathbf{u}-\mathbf{x})\right\}\right]d\mathbf{u}.
\end{aligned}$$

A similar argument from Chen (1999, Lemma) with f bounded on $\mathbb{T}_d$ shows the existence of r_2 and then the condition $\|K_{\theta(\mathbf{x},\mathbf{H})}\|_2^2 \lesssim c_2(\mathbf{x})(\det \mathbf{H})^{-r_2}$ leads successively to

$$\begin{aligned} 0 \;\leq\; R(\mathbf{x},\mathbf{H}) \;\lesssim\; & \frac{1}{n\,(\det \mathbf{H})^{r_2}} \int_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}\cap\mathbb{T}_d} c_2(\mathbf{x}) \left\{ (\mathbf{u}-\mathbf{x})^\top \nabla f(\mathbf{x}) + \frac{1}{2}(\mathbf{u}-\mathbf{x})^\top \nabla^2 f(\mathbf{x})\,(\mathbf{u}-\mathbf{x}) \right\} d\mathbf{u} \\ & \simeq \mathrm{o}\left\{ n^{-1}(\det \mathbf{H})^{-r_2} \right\}. \blacksquare \end{aligned}$$

Acknowledgements

We sincerely thank Francial G. Libengué for useful discussions and for the dataset of illustration.

Chapitre 3

Associated kernel discriminant analysis for multivariate mixed data

Abstract. Associated kernels have been introduced to improve the classical (symmetric) continuous kernels for smoothing any functional on several kinds of supports such as bounded continuous and discrete sets. In this paper, an associated kernel for discriminant analysis with multivariate mixed variables is proposed. These variables are of three types : continuous, categorical and count. The method consists of using a product of adapted univariate associated kernels and an estimate of the misclassification rate. A new profile version cross-validation procedure of bandwidth matrices selection is introduced for multivariate mixed data, while a classical cross-validation is used for homogeneous data sets having the same reference measures. Simulations and validation results show the relevance of the proposed method. The method has been validated on real coronary heart disease data in comparison to the classical kernel discriminant analysis.

3.1 Introduction

Discriminant analysis is a classical method in many scientific domains, such as bankruptcy prediction, face recognition, marketing and medicine. The dataset can be a mix of discrete (categorical, count) and continuous (rates, positive) data, such as the biomedical study in Rousseauw *et al.* (1983). Henceforth, the set $\mathbb{T}_d$ ($\subseteq \mathbb{R}^d$) denotes the support of the d-variate mixed data

$$\mathbb{T}_d = \mathbb{T}_{k_1}^{[1]} \times \cdots \times \mathbb{T}_{k_L}^{[L]} \text{ with } \sum_{\ell=1}^{L} k_\ell = d \tag{3.1}$$

and $\boldsymbol{\nu} = \nu_1 \times \ldots \times \nu_L$ represents the reference measure on $\mathbb{T}_d$, where ν_ℓ ($1 \leq \ell \leq L$) is the measure related to the corresponding support $\mathbb{T}_{k_\ell}^{[\ell]}$ with fixed k_ℓ-dimension (that has a given correlation structure).

In a discriminant analysis problem, a decision rule is used to classify a d-dimensionnal observation $\mathbf{x}$ belonging to $\mathbb{T}_d$ in one of the J classes. Specifically, the optimal Bayes rule

assigns an observation to the class with the largest posterior probability. It can be described as follows

$$\text{Allocate } \mathbf{x} \text{ to group } j_0 \text{ where } j_0 = \arg \max_{j\in\{1,\dots,J\}} \pi_j f_j(\mathbf{x}), \tag{3.2}$$

where π_j is the prior probability and $f_j(\mathbf{x})$ is the probability density function (pdf) of the jth class. These pdfs are usually unknown in practice, and can be estimated from the *training data set* using either parametric or nonparametric approaches. In parametric approaches, the underlying populations distributions are assumed to be known except for some unknown parameters such as mean vector and dispersion matrix ; see, e.g., Nath *et al.* (1992) and Simonoff (1996). The classic parametric approaches are the most used linear and quadratic discriminant techniques. However, they suffer from the restrictive assumption of normality. In nonparametric approaches, this assumption is relaxed and therefore more complex cases can be investigated. Kernel density estimation is a well-known method for constructing nonparametric estimations of population densities, namely in discriminant analysis ; see for instance Duong (2007) and Ghosh and Chaudhury (2004). Other methods have also been performed for nonparametric density estimation such as splines in Wahba (1990), Gu (1993) and Koo *et al.* (2009) and wavelets in Antoniadis (1997) and Shi *et al.* (2006).

More precisely, let $\mathcal{X}_j = \{\mathbf{X}_{j1}, \dots, \mathbf{X}_{jn_j}\}$ be a d-dimensional observations drawn from an unknown density function f_j, for $j = 1, \dots, J$, where the sample sizes n_j are known and non-random. Moreover, the observations $\mathbf{X}_{j1}, \dots, \mathbf{X}_{jn_j}$ are independent and identically distributed (iid) and are they belong to the jth population on the support $\mathbb{T}_d (\subseteq \mathbb{R}^d)$. The classical kernel estimator $\widehat{f_j}$ of f_j in (3.2) which uses continuous symmetric kernels is of the form :

$$\widehat{f_j}(\mathbf{x}; K, \mathbf{H}_j) = \frac{1}{n_j \det \mathbf{H}_j} \sum_{i=1}^{n_j} K\left\{\mathbf{H}_j^{-1}(\mathbf{x} - \mathbf{X}_{ji})\right\} = \widehat{f_j}(\mathbf{x}; \mathbf{H}_j), \quad \forall \mathbf{x} \in \mathbb{T}_d := \mathbf{R}^d, \tag{3.3}$$

where $\mathbf{H}_j$ is the jth symmetric and positive definite bandwidth matrix of dimension $d\times d$ and the function $K(\cdot)$ is the multivariate kernel assumed to be a spherically symmetric pdf. Since the choice of the kernel K does not change the result in classical case, a common notation will be used $\widehat{f_j}(\mathbf{x}; \mathbf{H}_j)$ for the jth density estimation with classical kernel. Thus, the Kernel Discriminant Rule (KDR) is obtained from the Bayes rule (3.2) by replacing f_j by $\widehat{f_j}$ and π_j is usually replaced by the sample proportion $\widehat{\pi}_j = n_j/n$ with $\sum_{j=1}^{J} n_j = n$; that is

$$\text{KDR : Allocate } \mathbf{x} \text{ to group } \widehat{j_0} \text{ where } \widehat{j_0} = \arg \max_{j\in\{1,\dots,J\}} \widehat{\pi}_j \widehat{f_j}(\mathbf{x}; \mathbf{H}_j). \tag{3.4}$$

This now raises the question of the choice of the most appropriate multivariate kernel according to the mix of discrete and continuous variables. In fact, The multivariate classical kernel (e.g. Gaussian) suits only for estimating the densities f_j with unbounded supports (i.e. $\mathbb{R}^d$) ; see Scott (1992), Duong (2004) and Zougab *et al.* (2014). Moreover, the multivariate Gaussian kernel is well-known in discriminant analysis. In order to estimate different functionals, Racine & Li (2007) proposed mutiple kernels composed of univariate Gaussian kernels for continuous variables and Aitchison & Aitken (1976)

kernels for categorical variables. Some implementations of the multiple kernels using the R software (R Development Core Team, 2015) have been performed in Hayfield & Racine (2007). Besides, it should be noted that the application of Gaussian kernels (i.e. symmetric) gives weights outside variables with bounded or discrete supports. In the univariate continuous case, Chen (1999, 2000) is one of the pioneers who has proposed asymmetric kernels, as beta and gamma, whose their supports coincide with those of the unknown functions to be estimated. Zhang (2010) and Zhang & Karunamuni (2010) studied the performance of these beta and gamma kernel estimators at the boundaries in comparison with those of the classical kernels ; see also Malec & Schienle (2014) and Igarashi & Kakizawa (2015). Libengué (2013) investigated several families of these univariate continuous kernels that he called univariate associated kernels ; see also Kokonendji *et al.* (2007), Kokonendji & Senga Kiessé (2011), Zougab *et al.* (2012) for univariate discrete situations. As for the multivariate case, Bouerzmarni & Rombouts (2009) propose product of univariate gamma and beta kernels. Another multivariate version with correlation structure of these associated kernels has been studied by Kokonendji and Somé (2015) for continuous density functions and by Somé and Kokonendji (2015) for multiple regression.

In this work, a new application of multivariate associated kernels for the discriminant analysis is proposed. The associated kernels are appropriated for both mixed training data $\mathbf{X}_{j1}, \ldots, \mathbf{X}_{jn_j}$ for $j = 1, \ldots, J$ and test data $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ drawn from $f = \sum_{j=1}^{J} \pi_j f_j$. In order to estimate the densities f_j in (3.2), we propose multiple (or product of) associated kernels composed by univariate discrete (e.g. binomial) and continuous (e.g. beta, gamma) associated kernels. The bandwidth matrices selection remains crucial to minimize the misclassification rate. A new profile cross-validation will be introduced for mixed variables, whereas a classic cross-validation will be used for homogeneous data sets. It should be noted that these associated kernels are adapted for this situation of mixing axis, since they fully respect the support of each explanatory variable. Some appropriated type of associated kernels, denoted by $\boldsymbol{\kappa}$, will be used for discriminant analysis by mean of simulations and validations.

The paper is organized as follows. Section 3.2 represents a general definition of multivariate associated kernels including both of the continuous classical symmetric and the multiple composed by univariate discrete and continuous. Then, the corresponding KDR appropriated for both continuous and discrete explanatory variables is given. Also, we present the profile cross-validation suitable for bandwidth matrices selection in the case of mixed variables. In Section 3.3, we investigate only the appropriated associate kernels according to the support of the variables through simulations studies and real data analysis. Finally, concluding remarks are drawn in Section 3.4.

3.2 Associated kernels for discriminant analysis

With the assumptions (3.1), the associated kernel $K_{\mathbf{x},\mathbf{H}}(\cdot)$ which replaces the classical kernel $K(\cdot)$ of (3.3) is a pdf according to some measure $\boldsymbol{\nu}$. This kernel $K_{\mathbf{x},\mathbf{H}}(\cdot)$ can be defined as follows :

Définition 3.2.1 *Let $\mathbb{T}_d \left(\subseteq \mathbb{R}^d\right)$ be the support of the densities f_j, to be estimated, $\mathbf{x} \in \mathbb{T}_d$ a target vector and $\mathbf{H}$ a bandwidth matrix. A parametrized pdf $K_{\mathbf{x},\mathbf{H}}(\cdot)$ of support $\mathbb{S}_{\mathbf{x},\mathbf{H}} \left(\subseteq \mathbb{R}^d\right)$ is called "multivariate (or general) associated kernel" if the following conditions are satisfied :*

$$\mathbf{x} \in \mathbb{S}_{\mathbf{x},\mathbf{H}}, \tag{3.5}$$

$$\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x} + \mathbf{a}(\mathbf{x},\mathbf{H}), \tag{3.6}$$

$$\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{B}(\mathbf{x},\mathbf{H}), \tag{3.7}$$

where $\mathcal{Z}_{\mathbf{x},\mathbf{H}}$ denotes the random vector with pdf $K_{\mathbf{x},\mathbf{H}}$ and both $\mathbf{a}(\mathbf{x},\mathbf{H}) = (a_1(\mathbf{x},\mathbf{H}),\ldots,a_d(\mathbf{x},\mathbf{H}))^\top$ and $\mathbf{B}(\mathbf{x},\mathbf{H}) = \left(b_{ij}(\mathbf{x},\mathbf{H})\right)_{i,j=1,\ldots,d}$ tend, respectively, to the null vector $\mathbf{0}$ and the null matrix $\mathbf{0}_d$ as $\mathbf{H}$ goes to $\mathbf{0}_d$.

From this definition and in comparison with (3.3), the jth associated kernel estimator $\widetilde{f_j}$ of f_j is

$$\widetilde{f_j}(\mathbf{x}) = \frac{1}{n_j}\sum_{i=1}^{n_j} K_{\mathbf{x},\mathbf{H}_j}\left(\mathbf{X}_{ji}\right) = \widetilde{f_j}(\mathbf{x};\boldsymbol{\kappa},\mathbf{H}_j), \tag{3.8}$$

where $\mathbf{H}_j \equiv \mathbf{H}_{jn_j}$ is the bandwidth matrix such that $\mathbf{H}_{jn_j} \to \mathbf{0}$ as $n_j \to \infty$, and $\boldsymbol{\kappa}$ represents the type of the associated kernel $K_{\mathbf{x},\mathbf{H}_j}$, parametrized by $\mathbf{x}$ and $\mathbf{H}_j$. Without loss of generality, and hereafter $\widetilde{f_j}(\mathbf{x};\boldsymbol{\kappa},\mathbf{H}_j) \equiv \widetilde{f_j}(\mathbf{x};\boldsymbol{\kappa})$ will be used in order to point out the effect of $\boldsymbol{\kappa}$, since the bandwidth matrix is here investigated only by cross-validation. Therefore, the classical KDR of (3.4) becomes the associated KDR (AKDR) using (3.8)

$$\text{AKDR : Allocate } \mathbf{x} \text{ to group } \widetilde{j_0} \text{ where } \widetilde{j_0} = \arg\max_{j\in\{1,\ldots,J\}} \widehat{\pi}_j \widetilde{f_j}(\mathbf{x};\boldsymbol{\kappa}). \tag{3.9}$$

The following two examples provide the well-known multivariate associated kernel estimators and they represent particular cases that can be used in (3.1). The first can be seen as an interpretation of classical associated kernels through continuous symmetric kernels. The second deals on non-classical associated kernels without correlation structure.

Given a target vector $\mathbf{x} \in \mathbb{R}^d =: \mathbb{T}_d$ and a bandwidth matrix $\mathbf{H}$, it follows that the classical kernel in (3.3) with null mean vector and covariance matrix $\boldsymbol{\Sigma}$ induces the so-called (multivariate) classical associated kernel :

$$(i)\ \ K_{\mathbf{x},\mathbf{H}}(\cdot) = \frac{1}{\det\mathbf{H}} K\left\{\mathbf{H}^{-1}(\mathbf{x}-\cdot)\right\} \tag{3.10}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbf{x} - \mathbf{H}\mathbb{S}_d$ with $\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x}$ (i.e. $\mathbf{a}(\mathbf{x},\mathbf{H}) = \mathbf{0}$) and $\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}$;

$$(ii)\ \ K_{\mathbf{x},\mathbf{H}}(\cdot) = \frac{1}{(\det\mathbf{H})^{1/2}} K\left\{\mathbf{H}^{-1/2}(\mathbf{x}-\cdot)\right\}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbf{x} - \mathbf{H}^{1/2}\mathbb{S}_d$ with $\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x}$ (i.e. $\mathbf{a}(\mathbf{x},\mathbf{H}) = \mathbf{0}$) and $\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{H}^{1/2}\boldsymbol{\Sigma}\mathbf{H}^{1/2}$.

A second particular case of Definition 3.2.1, appropriate for a mix of both continuous and count explanatory variables without correlation structure is presented as follows.

Let $\mathbf{x} = (x_1, \ldots, x_d)^\top \in \times_{\ell=1}^{d} \mathbb{T}_1^{[\ell]} =: \mathbb{T}_d$ with $k_\ell = 1$ of (3.1) and $\mathbf{H} = \mathbf{Diag}(h_{11}, \ldots, h_{dd})$ with $h_{\ell\ell} > 0$. Let $K_{x_\ell,h_{\ell\ell}}^{[\ell]}$ be a (discrete or continuous) univariate associated kernel (see Definition 3.2.1 for $d = 1$) with its corresponding random variable $\mathcal{Z}_{x_\ell,h_{\ell\ell}}^{[\ell]}$ on $\mathbb{S}_{x_\ell,h_{\ell\ell}} (\subseteq \mathbb{R})$ for all $\ell = 1, \ldots, d$. Then, the multiple associated kernel is also a multivariate associated kernel :

$$K_{\mathbf{x},\mathbf{H}}(\cdot) = \prod_{\ell=1}^{d} K_{x_\ell,h_{\ell\ell}}^{[\ell]}(\cdot) \tag{3.11}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \times_{\ell=1}^{d} \mathbb{S}_{x_\ell,h_{\ell\ell}}$ with $\mathbb{E}(\mathcal{Z}_{\mathbf{x},\mathbf{H}}) = (x_1 + a_1(x_1, h_{11}), \ldots, x_d + a_d(x_d, h_{dd}))^\top$ and $\mathrm{Cov}(\mathcal{Z}_{\mathbf{x},\mathbf{H}}) = \mathbf{Diag}\,(b_{\ell\ell}(x_\ell, h_{\ell\ell}))_{\ell=1,\ldots,d}$. In other words, the random variables $\mathcal{Z}_{x_\ell,h_{\ell\ell}}^{[\ell]}$ are independent components of the random vector $\mathcal{Z}_{\mathbf{x},\mathbf{H}}$.

In the following two sections, some examples of associated kernels are illustrated and then criteria of discriminant analysis are presented.

3.2.1 Some associated kernels

In order to point out the importance associated kernel in discriminant analysis, below some kernels most used in the literature will be listed. These concern seven basic associated kernels for which three of them are univariate discrete, three others are univariate continuous and the last one is a bivariate beta with correlation structure.

- The binomial kernel is defined on the support $\mathbb{S}_x = \{0, 1, \ldots, x+1\}$ with $x \in \mathbb{T}_1 := \mathbb{N} = \{0, 1, \ldots\}$ and then $h \in (0, 1]$:

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

  where $\mathbb{1}\{A\}$ denotes the indicator function of any given event A. Note that $B_{x,h}$ is the probability mass function (pmf) of the binomial distribution $\mathcal{B}(x+1; (x+h)/(x+1))$ with its number of trials $x+1$ and its success probability in each trial $(x+h)/(x+1)$. It is appropriated for count data with small or moderate sample sizes and, also, it does not satisfy (3.7) ; see Kokonendji and Senga Kiessé (2011) and also Zougab *et al.* (2014) for a bandwidth selection by Bayesian method.
- For a given fixed arm $a \in \mathbb{N}$, the discrete triangular kernel is defined on $\mathbb{S}_{x,a} = \{x, x \pm 1, \ldots, x \pm a\}$ with $x \in \mathbb{T}_1 = \mathbb{N}$:

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

  where $h > 0$ and $P(a, h) = (2a+1)(a+1) - 2\sum_{k=0}^{a} k^h$ is the normalizing constant. It is symmetric around the target x, satisfying Definition 3.2.1 and suitable for count variables ; see Kokonendji *et al.* (2007) and also Kokonendji and Zocchi (2010) for an asymmetric version.
- From Aitchison and Aitken (1976), Kokonendji and Senga Kiessé (2011) deduced the following discrete kernel which is here labelled DiracDU as "Dirac Discrete

Uniform". For fixed $c \in \{2, 3, \ldots\}$ the number of categories, one has $\mathbb{S}_c = \{0, 1, \ldots, c-1\}$ and

$$DU_{x,h;c}(u) = (1-h)^{1-\mathbb{1}\{u\in\mathbb{S}_c\setminus\{x\}\}}\left(\frac{h}{c-1}\right)^{\mathbb{1}\{u\in\mathbb{S}_c\setminus\{x\}\}},$$

where $h \in (0, 1]$ and $x \in \mathbb{T}_1$. This DiracDU kernel is symmetric around the target, satisfying Definition 3.2.1 and appropriated for categorical set $\mathbb{T}_1$.

– From the well known Gaussian kernel $K^G(u) = (h\sqrt{2\pi})^{-1}\exp(u^2)\mathbb{1}_{\mathbb{R}}(u)$, the associated kernel version on $\mathbb{S}_{x,h} = \mathbb{R}$ with $x \in \mathbb{T}_1 := \mathbb{R}$ and $h > 0$ is defined as :

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

It is obtained from (3.10) and it is well adapted for continuous variables with unbounded supports.

– The gamma kernel is defined on $\mathbb{S}_{x,h} = [0, \infty) = \mathbb{T}_1$ with $x \in \mathbb{T}_1$ and $h > 0$:

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

where $\Gamma(\cdot)$ is the classical gamma function. It is the pdf of the gamma distribution $\mathcal{G}a(1 + x/h, h)$ with scale parameter $1 + x/h$ and shape parameter h. It satisfies Definition 3.2.1 and it is appropriated for non-negative real set $\mathbb{T}_1$; see Chen (2000).

– The beta kernel is however defined on $\mathbb{S}_{x,h} = [0, 1] = \mathbb{T}_1$ with $x \in \mathbb{T}_1$ and $h > 0$:

$$BE_{x,h}(u) = \frac{u^{x/h}(1-u)^{(1-x)/h}}{\mathscr{B}(1 + x/h, 1 + (1-x)/h)}\mathbb{1}\{u \in [0, 1]\},$$

where $\mathscr{B}(r, s) = \int_0^1 t^{r-1}(1-t)^{s-1}dt$ is the usual beta function with $r > 0$ and $s > 0$. It is the pdf of the beta distribution $\mathcal{B}e(1 + x/h, (1-x)/h)$ with shape parameters $1 + x/h$ and $(1-x)/h$. This pdf satisfies Definition 3.2.1 and is appropriated for rates, proportions and percentages dataset $\mathbb{T}_1$; see

– Finally, the bivariate beta kernel define by

$$\begin{aligned} BS_{\mathbf{x},\mathbf{H}}(u_1, u_2) &= \left(\frac{u_1^{x_1/h_{11}}(1-u_1)^{(1-x_1)/h_{11}}}{\mathscr{B}(1 + x_1/h_{11}, 1 + (1-x_1)/h_{11})}\right)\left(\frac{u_2^{x_2/h_{22}}(1-u_2)^{(1-x_2)/h_{22}}}{\mathscr{B}(1 + x_2/h_{22}, 1 + (1-x_2)/h_{22})}\right) \\ &\times\left(1 + h_{12}\times\frac{u_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2}\widetilde{\sigma}_1(x_1, h_{11})}\times\frac{u_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2}\widetilde{\sigma}_2(x_2, h_{22})}\right)\mathbb{1}\left\{\begin{pmatrix}u_1\\u_2\end{pmatrix}^\top \in [0, 1]^2\right\}, \quad (3.12) \end{aligned}$$

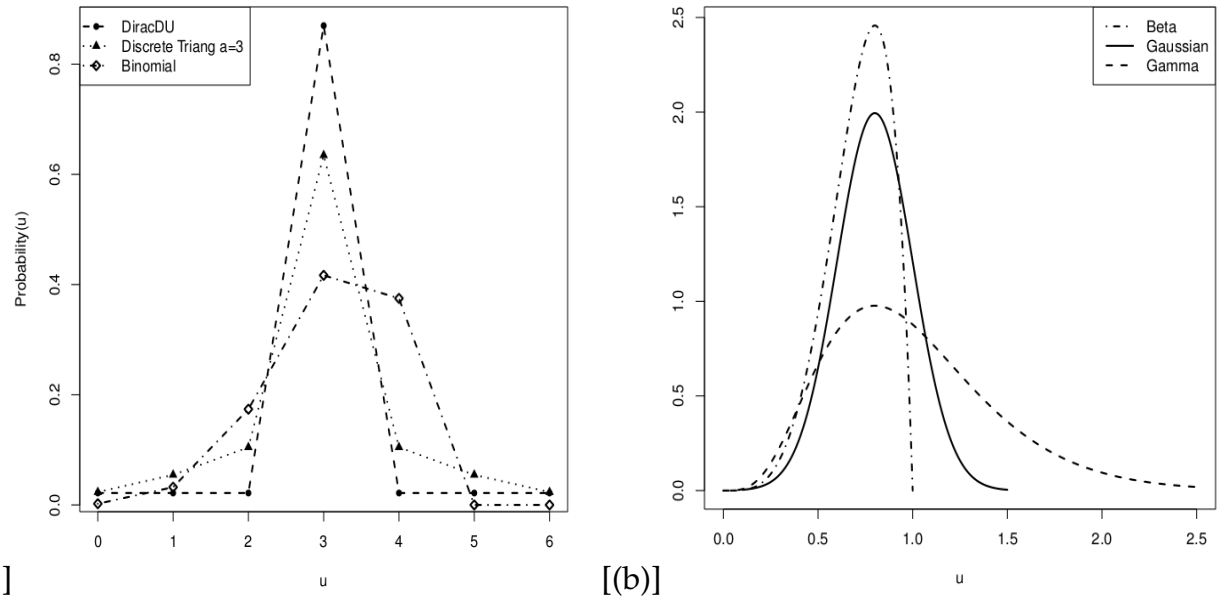
with $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbb{T}_2 = [0, 1]^2$, $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{T}_2$ and $\mathbf{H} = \begin{pmatrix}h_{11} & h_{12}\\ h_{12} & h_{22}\end{pmatrix}$ will be considered. For $j = 1, 2$, the characteristics in (3.12) are given by $h_{jj} > 0$, $\widetilde{\mu}_j(x_j, h_{jj}) = (x_j + h_{jj})/(1 + 2h_{jj})$, $\widetilde{\sigma}_j^2(x_j, h_{jj}) = (x_j + h_{jj})(1 + h_{jj} - x_j)(1 + 2h_{jj})^{-2}(1 + 3h_{jj})^{-1}h_{jj}$, and the constraints

$$h_{12} \in [-\beta, \beta'] \cap \left(-\sqrt{h_{11}h_{22}}\,,\, \sqrt{h_{11}h_{22}}\right)$$

with $\beta = \left(\max_{v_1,v_2} \left\{ \frac{v_1 - \widetilde{\mu}_1(x_1,h_{11})}{h_{11}^{1/2}\widetilde{\sigma}_1(x_1,h_{11})} \times \frac{v_2 - \widetilde{\mu}_2(x_2,h_{22})}{h_{22}^{1/2}\widetilde{\sigma}_2(x_2,h_{22})} \right\} \right)^{-1}$ and

$\beta' = \left| \left(\min_{v_1,v_2} \left\{ \frac{v_1 - \widetilde{\mu}_1(x_1,h_{11})}{h_{11}^{1/2}\widetilde{\sigma}_1(x_1,h_{11})} \times \frac{v_2 - \widetilde{\mu}_2(x_2,h_{22})}{h_{22}^{1/2}\widetilde{\sigma}_2(x_2,h_{22})} \right\} \right)^{-1} \right|$.

It satisfies Definition 3.2.1 and is adapted for bivariate rates. The full bandwidth matrix **H** allows any orientation of the kernel. Therefore, it can reach any point of the space which might be inaccessible with diagonal matrix. This type of kernel is called beta-Sarmanov kernel by Kokonendji and Somé (2015) ; see Sarmanov (1966) for this construction of multivariate densities with some correlation structure from independent components.



[(a)] [(b)]

FIGURE 3.1 – Shapes of univariate (discrete and continuous) associated kernels : (a) DiracDU, discrete triangular $a = 3$ and binomial with same target $x = 4$ and bandwidth $h = 0.13$; (b) Beta, Gaussian and gamma with same $x = 0.8$ and $h = 0.2$.

Figure 3.1 shows some forms of the above-mentioned univariate associated kernels. The plots highlight the importance of the target point and around it in discrete (a) and continuous (b) cases. Furthermore, for a fixed bandwidth h, the Gaussian keeps its same shape along the support ; however, they change according to the target for other non-classical associated kernels. This explains the inappropriateness of the Gaussian kernel for density estimation in any bounded interval ; see Part (b) of Figure 3.1.

3.2.2 Misclassification rate and banwdith matrix selection

The performance of the AKDR method is investigated by the misclassification rate denoted by MR. This error rate is the proportion of the points that are assigned to an

incorrect group based on the discriminant rule (3.9). Then, we have

$$\begin{aligned} 1 - \text{MR} &= \mathbb{P}(\mathbf{Y} \text{ is correctly classified}) \\ &= \mathbb{E}_{\mathbf{Y}}(\mathbb{1}\{\mathbf{Y} \text{ is correctly classified}\}), \end{aligned}$$

where $\mathbb{E}_{\mathbf{Y}}$ is the expectation with respect to $\mathbf{Y}$ or $\sum_{j=1}^{J} \pi_j f_j$. See also Hall and Wand (1988) who proposed to find optimal bandwidth that directly optimise this MR for a two-class problem. In practice, an estimate of MR is used, denoted $\widehat{\text{MR}}$, for test data $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ and also for training data :

$$\widehat{\text{MR}} = 1 - m^{-1} \sum_{k=1}^{m} \mathbb{1}\{\mathbf{Y}_k \text{ is correctly classified using AKDR}\}. \tag{3.13}$$

The optimal bandwidth matrices that give the best $\widehat{\text{MR}}$ is chosen by the least squared cross-validation (LSCV) method. The multivariate cross-validation is a straightforward generalization of the one dimensional case. For the jth group, the LSCV estimator is

$$\text{LSCV}(\mathbf{H}_j) = \int_{\mathbb{T}_d} \left\{\widetilde{f}_j(\mathbf{x})\right\}^2 \boldsymbol{\nu}(d\mathbf{x}) - \frac{2}{n_j} \sum_{i=1}^{n_j} \widetilde{f}_{j,-i}(\mathbf{X}_{ji}), \tag{3.14}$$

where $\widetilde{f}_{j,-i}\left(\mathbf{X}_{ji}\right) = (n_j - 1)^{-1} \sum_{k \neq i}^{n_j} K_{\mathbf{X}_{ji},\mathbf{H}_j}(\mathbf{X}_{jk})$ is being computed as $\widetilde{f}_j(\mathbf{X}_{ji})$ excluding the observation $\mathbf{X}_{ji}$. Practically, for homogeneous (continuous and discrete) data the first term of (3.14) is calculated by successive sums or integrals according to the appropriated measure $\boldsymbol{\nu}$ (Lebesgue or count). In that case, the optimal bandwidth matrix $\widetilde{\mathbf{H}}_j$ obtained by LSCV rule (3.14) with multiple asociated kernels (3.11) is defined as follows :

$$\widetilde{\mathbf{H}}_j = \arg\min_{\mathbf{H}_j \in \mathcal{D}} \text{LSCV}(\mathbf{H}_j), \tag{3.15}$$

where $\mathcal{D}$ is the set of all positive definite diagonal bandwidth matrices. Its algorithm is described below and used for numerical studies in the following section.

Algorithms of LSCV method (3.2.2) for some type of associated kernels and their correponsding bandwidth matrices

A1. Multiple associated kernels (i.e. diagonal bandwidth matrices) for $d \geq 2$.
   1. Choose d intervals $H_{11}, \ldots, H_{dd}$ related to $h_{11}, \ldots, h_{dd}$, respectively.
   2. For $\delta_1 = 1, \ldots, \ell(H_{11}), \ldots, \delta_d = 1, \ldots, \ell(H_{dd})$,
      Compose the diagonal bandwidth matrix $\mathbf{H}_j(\delta_1, \ldots, \delta_d) := \mathbf{Diag}\left(H_{11}(\delta_1), \ldots, H_{dd}(\delta_d)\right)$.
   3. Apply LSCV method on the set $\mathcal{D}$ of all diagonal bandwidth matrices $\mathbf{H}_j(\delta_1, \ldots, \delta_d)$.

However, for mixed case where the Fubini theorem is not applicable in (3.14), one uses the so called profile cross-validation since there is no convergence of the classical cross-validation algorithm below. The method is presented in bivariate case for any group

($j = 1, \ldots, J$). For instance, for $\mathbb{T}_2 = [0,1] \times \mathbb{N}$ and a multiple associated kernel (3.11) beta×binomial, the smoothing parameter $h_{11(j)}$ of the beta kernel is fixed, followed by a minimisation of the cross-validation function on $h_{22(j)} \in (0,1]$ of the binomial kernel :

$$\widetilde{h}_{22(j)[h_{11(j)}]} = \arg \min_{h_{22(j)} \in (0,1]} \mathrm{LSCV}_{h_{11(j)}}(h_{22(j)}),$$

with $\mathrm{LSCV}_{h_{11(j)}}(h_{22(j)}) = \mathrm{LSCV}(h_{11(j)}, h_{22(j)}) := \mathrm{LSCV}(\mathbf{H}_j)$. Furthermore, $h_{11(j)}$ is unknown, hence for each $h_{11(j)}$ we can evaluate $\widetilde{h}_{22(j)[h_{11(j)}]}$ and estimate the corresponding misclassification rate $\widehat{\mathrm{MR}}(h_{11(j)}, \widetilde{h}_{22(j)[h_{11(j)}]})$. The jth optimal bandwidth matrix using the profile cross-validation and denoted $\widehat{\mathbf{H}}_{jp}$ is

$$\widehat{\mathbf{H}}_{jp} = \arg \min_{h_{11(j)}, \widetilde{h}_{22[h_{11(j)}]}} \widehat{\mathrm{MR}}(h_{11(j)}, \widetilde{h}_{22[h_{11(j)}]}). \tag{3.16}$$

This example of profile cross-validation (3.16) and the classical one (3.14) are implemented in the next section. It should be noted that, other bandwidth matrix selection by Bayesian methods is possible ; see, e.g., Ziane *et al.* (2015) for the adaptive case using asymmetric kernel.

3.3 Numerical studies

Before presenting some simulations results and real data analysis, let us start by the algorithm of the associated kernel discriminant analysis.

Algorithm for associated kernel discriminant analysis

1. For each training sample $\mathcal{X}_j = \{\mathbf{X}_{j1}, \ldots, \mathbf{X}_{jn_j}\}$, $j = 1, 2, \ldots, J$, compute a kernel density estimate (3.8) using the optimal bandwidth matrix obtained by cross-validation (3.2.2).
2. Use the prior probabilities once they are known. Otherwise, estimate them using the training sample proportions $\widehat{\pi}_j = n_j/n$.
3. (a) Allocate test data points $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ according to AKDR of (3.9).
   (b) Allocate all points $\mathbf{x}$ from the sample space according to AKDR of (3.9).
4. (a) If we have test data then the estimate of the misclassification rate is $\widehat{\mathrm{MR}}$ of (3.13).
   (b) If the test data are not available, the cross-validation estimate of misclassification rate is

$$\widehat{\mathrm{MR}}_{\mathrm{cv}} = 1 - n^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \mathbb{1}\{\mathbf{X}_{ji} \text{ is correctly classified using } \mathrm{AKDR}_{ji}\} \tag{3.17}$$

where AKDR_{ji} is similar to AKDR except that $\widehat{\pi}_j$ and $\widetilde{f_j}(\mathbf{x};\boldsymbol{\kappa})$ are replaced by their leave-one-out estimates by removing $\mathbf{X}_{ji}$ i.e. $\widehat{\pi}_{j,-i} = (n_j - 1)/n$ and

$$\widetilde{f_{j,-i}}(\mathbf{x};\boldsymbol{\kappa}) = \frac{1}{n_j - 1}\sum_{r\neq i}^{n_j} K_{\mathbf{x},\mathbf{H}_{j,-i}}\left(\mathbf{X}_{jr}\right).$$

That is, we repeat step 3 to classify all $\mathbf{X}_{ji}$ using AKDR_{ji}.

3.3.1 Simulation studies

In this section, the results of a simulation study that was conducted for evaluating the performance of the algorithm of AKDR is presented. Computations have been performed on the supercomputer facilities of the "*Mésocentre de calcul de Franche-Comté*" using the `R` software ; see `R` Development Core Team (2015). This simulation study has two objectives. First, we investigate the ability of multiple associated kernels (3.14) which can scrupulously respect any bounded, count and mixed support $\mathbb{T}_d$ of dataset, and more give good estimate $\widehat{\text{MR}}$ of (3.13). We, therefore, use only appropriate multiple associated kernels for these simulations studies. Second, we evaluate the sentitivity of the proposed method in relation to the sample size n.

Three scenarios denoted A, B and C are considered in dimension $d = 2$. With Scenario A data are generated using a mixture of two bivariates Dirichlet density

$$\begin{aligned} f_A(x_1,x_2) &= \frac{3\Gamma(\alpha_1+\alpha_2+\alpha_3)}{7\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}x_1^{\alpha_1-1}x_2^{\alpha_2-1}(1-x_1-x_2)^{\alpha_3-1}\mathbb{1}_{\{x_1,\ x_2\geq 0,\ x_1+x_2\leq 1\}}(x_1,x_2)\\ &\quad+\frac{4\Gamma(\beta_1+\beta_2+\beta_3)}{7\Gamma(\beta_1)\Gamma(\beta_2)\Gamma(\beta_3)}x_1^{\beta_1-1}x_2^{\beta_2-1}(1-x_1-x_2)^{\beta_3-1}\mathbb{1}_{\{x_1,\ x_2\geq 0,\ x_1+x_2\leq 1\}}(x_1,x_2), \end{aligned}$$

where $\Gamma(\cdot)$ is the classical gamma function, with parameter values $\alpha_1 = \alpha_2 = 5$, $\alpha_3 = 6$, $\beta_1 = \beta_2 = 2$ and $\beta_3 = 10$. With Scenario B data are generated using a mixture of two bivariates Poisson with any correlation structure

$$\begin{aligned} f_B(x_1,x_2) &= \frac{2e^{-(\theta_1+\theta_2+\theta_{12})}}{5}\sum_{i=0}^{min(x_1,x_2)}\frac{\theta_1^{x_1+i}\theta_2^{x_2+i}\theta_{12}^{i}}{(x_1+i)!(x_2+i)!i!}\mathbb{1}_{\mathbb{N}\times\mathbb{N}}(x_1,x_2)\\ &\quad+\frac{3e^{-(\theta_a+\theta_b+\theta_{ab})}}{5}\sum_{i=0}^{min(x_1,x_2)}\frac{\theta_a^{x_1+i}\theta_b^{x_2+i}\theta_{ab}^{i}}{(x_1+i)!(x_2+i)!i!}\mathbb{1}_{\mathbb{N}\times\mathbb{N}}(x_1,x_2), \end{aligned}$$

with parameter values $\theta_1 = 2$, $\theta_2 = 3$, $\theta_{12} = 4$, $\theta_a = 3$, $\theta_b = 4$ and $\theta_{ab} = 5$. Finally, with Scenario C is a bivariate beta without correlation

$$f_C(x_1,x_2) = \left(\frac{3e^{-2}2^{x_1}}{7x_1!}+\frac{4e^{-3}3^{x_2}}{7x_2!}\right)\times\frac{x_2^{p_1-1}(1-x_2)^{q_1-1}}{\mathscr{B}(p_1,q_1)}\ \mathbb{1}_{\mathbb{N}}(x_1)\mathbb{1}_{[0,1]}(x_2),$$

with $(p_1,q_1) = (2,7)$. The use of these bivariate distributions is motivated by the aim of investigating unbounded continuous, count and mixed situations and, if possible, correlation structure in the data. For each scenario we generate $N_{sim} = 100$ dataset of different sizes.

The choice of the multiple beta kernel is motivated by the cumbersome procedures of the beta-Sarmanov kernel ; see Kokonendji and Somé (2015) for further details. Also, the authors show that this beta-Sarmanov kernel with correlation structure gives similar result to the multiple beta without correlation structure. Thus, we here focus on the bivariate case. The test data size is $m = 200$ for each replication. We consider sample size $n = 250$ and 500 in continuous case and the multiple beta kernel which is the more suitable for bivariate rates data. Table 3.2 reports the average $\widehat{\text{MR}}$ which we denote $\overline{\widehat{\text{MR}}}$. We can observe low values that prove the appropriateness of the discriminant analysis algorithm mentioned above in this section. Furthermore, the errors become smaller when the sample size increases.

	m	n	Beta×Beta
A	200	250	0.0421(0.0146)
		500	0.0329(0.0097)

TABLE 3.1 – Some expected values of $\overline{\widehat{\text{MR}}}$ and their standard errors in parentheses with $N_{sim} = 100$ of the multiple beta kernel with test data size $m = 200$.

Since the binomial kernel is the most interesting of discrete (count) associated kernels for small sample sizes, we consider the samples sizes $n = 50, 100$ and 200 ; see Somé and Kokonendji (2015) for comparisons with the discrete (symmetric) triangular kernel. In each replication, the test data size is $m = 100$. Table 3.2 reports the values of $\overline{\widehat{\text{MR}}}$ for this appropriate multiple binomial kernel. Once again, we observe low values of $\overline{\widehat{\text{MR}}}$ which has a tendency to become better when the sample size increases. For scenario C of mixed

	m	n	Binomial×Binomial
B	100	50	0.0747(0.0334)
		100	0.0677(0.0255)
		200	0.0581(0.0221)

TABLE 3.2 – Some expected values of $\overline{\widehat{\text{MR}}}$ and their standard errors in parentheses with $N_{sim} = 100$ of the multiple binomial kernel with test data size $m = 100$.

case, we have to take into account the effects of the continuous and discrete sample size. Thus, we consider sample sizes $n = 80, 250$ and 500 and test data size $m = 150$. The values of $\overline{\widehat{\text{MR}}}$ in Table 3.3 show the effectiveness of the profile cross-validation method (3.16). Also, the errors rate is getting improved when the sample size increases.

For now, it is not possible to make simulations studies with more than two variables mainly due to the time consuming of the cross-validation methods.

	m	n	Beta×Binomial
		80	0.0182(0.0116)
C	150	250	0.0171(0.0080)
		500	0.0152(0.0042)

TABLE 3.3 – Some expected values of $\widehat{\overline{\text{MR}}}$ and their standard errors in parentheses with $N_{sim} = 100$ of the beta×binomial kernel with test data size $m = 150$.

3.3.2 Real data analysis

The algorithm of AKDR in Section 3.3 is applied using multiple associated kernels (3.11). The dataset represents a retrospective sample of males in a heart-disease high-risk region of the Western Cape, South Africa. These data are taken from a larger dataset, described in Rousseauw *et al.* (1983) and available in Hastie *et al.* (2009) ; see also the `R` package `ElemStatLearn` of Halvorsen (2015. It has 462 observations on 10 variables composed of five continous (positive) variables, three count variables, one categorical variable and the classification variable Coronary Heart Disease (CHD). This CHD variable has two groups : the "group 1" of patients with CHD and the "group 2" of those without CHD. The used multiple associated kernel is composed by gamma kernels for continuous (positive) variables, binomial kernels for count variables and DiracDU kernel for the categorical one. We must here use the cross-validation estimate MR_{cv} of (3.17) since we do not have test data. Also, the profile cross- validation is appropriate for this mixed dataset. This method is already computationally intense for two variables, and thus is not recommended for this dataset of ten mixed variables. Ttwo diagonal bandwidth matrices $\mathbf{H}_{\text{group1}} = \mathbf{Diag}(0.01, 0.02, 0.03, 0.015, 0.004, 0.02, 0.03, 0.02, 0.02)$ and $\mathbf{H}_{\text{group2}} = \mathbf{Diag}(0.87, 0.3, 0.2, 0.15, 0.4, 0.2, 0.4, 0.2, 0.56, 0.6)$ are chosen for both groups of CHD and the cross-validation estimate of the misclassification rate MR_{cv} is 30.090%. Also, we use the LSCV selector with full bandwidth matrices and multivariate Gaussian kernel of Duong (2007), and the misclassification rate is equal to 30.952%. Thus, the associated kernel discriminant analysis with chosen bandwidth matrices is slightly better than the classical one of Duong (2007).

3.4 Concluding remarks

We have presented associated kernels for discriminant analysis and in presence of a mixture of discrete and continuous explanatory variables. Two particular cases including the continuous classical and the multiple (or product of) associated kernels are highlight. Also, six univariate associated kernels and a bivariate beta with correlation structure are presented. Four of them, namely binomial, DiracDU, beta and gamma are used for simulations studies and real data analysis. The bandwidth selection is obtained by classical cross-validation for homogeneous (continuous or discrete) data while a profile version is provided for mixed data. The bandwidth selections methods can be improved by Bayesian methods ; see, e.g., Zougab *et al.* (2014) for a choice of the

bandwidth matrix by Bayesian methods.

Simulation experiments and analysis of a real dataset provide insight into the appropriateness of associated kernel for small and moderate sample sizes and also for homogeneous or mixed data. Table 3.1 and 3.2 shows the efficiency of the discriminant analysis with appropriate kernels while Table 3.3 gives an effective bandwidth selection using profile cross-validation. The method needs some improvements, particularly in terms of computation time, to be applicable to real mixed datasets with more than two variables. Further research can be done on the bandwidth matrix selection, especially to speed it up for example by parallelism processing.

Chapitre 4

Effects of associated kernels in nonparametric multiple regressions

Abstract. Associated kernels have been introduced to improve the classical continuous kernels for smoothing any functional on several kinds of supports such as bounded continuous and discrete sets. This work deals with the effects of combined associated kernels on nonparametric multiple regression functions. Using the Nadaraya-Watson estimator with optimal bandwidth matrices selected by cross-validation procedure, different behaviours of multiple regression estimations are pointed out according the type of multivariate associated kernels with correlation or not. Through simulation studies, there are no effect of correlation structures for the continuous regression functions and also for the associated continuous kernels ; however, there exist really effects of the choice of multivariate associated kernels following the support of the multiple regression functions bounded continuous or discrete. Applications are made on two real datasets.

4.1 Introduction

Considering the relation between a response variable Y and a d-vector ($d \geq 1$) of explanatory variables $\mathbf{x}$ given by

$$Y = m(\mathbf{x}) + \epsilon, \tag{4.1}$$

where m is the unknown regression function from $\mathbb{T}_d \subseteq \mathbb{R}^d$ to $\mathbb{R}$ and ϵ the disturbance term with null mean and finite variance. We are interested in estimating or smoothing this regression function m by taking into account especially its structure. Let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ be a sequence of independent and identically distributed (iid) random vectors on $\mathbb{T}_d \times \mathbb{R} (\subseteq \mathbb{R}^{d+1})$ with $m(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ of (4.1). The Nadaraya (1964) and Watson (1964) estimator $\widehat{m}_n$ of m, using continuous classical (symmetric) kernels is

$$\widehat{m}_n(\mathbf{x}; K, \mathbf{H}) = \sum_{i=1}^{n} \frac{Y_i K\left\{\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)\right\}}{\sum_{i=1}^{n} K\left\{\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)\right\}} = \widehat{m}_n(\mathbf{x}; \mathbf{H}), \quad \forall \mathbf{x} \in \mathbb{T}_d := \mathbb{R}^d, \tag{4.2}$$

where $\mathbf{H}$ is the symmetric and positive definite bandwidth matrix of dimension $d \times d$ and the function $K(\cdot)$ is the multivariate kernel assumed to be spherically symmetric probability density function. Since the choice of the kernel K is not important in classical case, we use the common notation $\widehat{m}_n(\mathbf{x}; \mathbf{H})$ for classical kernel regression. The multivariate classical kernel (e.g. Gaussian) suits only for regression functions on unbounded supports (i.e. $\mathbb{R}^d$) ; see also Scott (1992) and Wand and Jones (1995). Racine and Li (2004) proposed product of kernels composed by univariate Gaussian kernels for continuous variables and Aitchison and Aitken (1976) kernels for categorical variables ; see also Hayfield and Racine (2007) for some implementations and uses of these multiple kernels. Notice that the use of symmetric kernels gives weights outside variables with unbounded supports. In the univariate continuous case, Chen (1999, 2000) is one of the pioneers who has proposed asymmetric kernels (i.e. beta and gamma) which supports coincide with those of the functions to be estimated. Zhang (2010) and Zhang and Karunamuni (2010) studied the performance of these beta and gamma kernel estimators at the boundaries in comparison with those of the classical kernels. Recently, Libengué (2013) investigated several families of these univariate continuous kernels that he called univariate associated kernels ; see also Kokonendji *et al.* (2007), Kokonendji and Senga Kiessé (2001), Zougab *et al.* (2012) and Wansouwé *et al.* (2014b) for univariate discrete situations. A continuous multivariate version of these associated kernels have been studied by Kokonendji and Somé (2015) for density estimation.

The main goal of this work is to consider multivariate associated kernels and then to investigate the importance of their choice in multiple regression. These associated kernels are appropriated for both continuous and count explanatory variables. In fact, in order to estimate the regression function m in (4.1), we propose multiple (or product of) associated kernels composed by univariate discrete associated kernels (e.g. binomial, discrete triangular) and continuous ones (e.g. beta, Epanechnikov). We will also use a bivariate beta kernel with correlation structure. Another motivation of this work is to investigate the effect of correlation structure for explanatory variables in continuous regression estimation. These associated kernels suit for this situation of mixing axes as they fully respect the support of each explanatory variable. In other words, we will measure the effect of type of associated kernels, denoted $\boldsymbol{\kappa}$, in multiple regression by simulations and applications.

The rest of the paper is organized as follows. Section 4.2 gives a general definition of multivariate associated kernels which includes the continuous classical symmetric and the multiple composed by univariate discrete and continuous. For each definition, the corresponding kernel regression appropriated for both continuous and discrete explanatory variables are given ; asymptotic bias and variance are also presented.. In Section 4.3, we explore the importance of the choice of appropriated associated kernels according to the support of the variables through simulations studies and real data analysis. Finally, summary and final remarks are drawn in Section 4.4.

4.2 Multiple regression by associated kernels

4.2.1 Definition and properties

In order to include both discrete and continuous regressors, we assume $\mathbb{T}_d$ is any subset of $\mathbb{R}^d$. More precisely, for $j = 1, \ldots, n$, let us consider on $\mathbb{T}_d = \otimes_{j=1}^{d} \mathbb{T}_1^{[j]}$ the measure $\boldsymbol{\nu} = \nu_1 \otimes \ldots \otimes \nu_d$ where ν_j is a Lesbesgue or count measure on the corresponding univariate support $\mathbb{T}_1^{[j]}$. Under these assumptions, the associated kernel $K_{\mathbf{x},\mathbf{H}}(\cdot)$ which replaces the classical kernel $K(\cdot)$ of (4.2) is a probability density function (pdf) in relation to a measure $\boldsymbol{\nu}$. This kernel $K_{\mathbf{x},\mathbf{H}}(\cdot)$ can be defined as follows.

Définition 4.2.1 *Let $\mathbb{T}_d \left(\subseteq \mathbb{R}^d\right)$ be the support of the regressors, $\mathbf{x} \in \mathbb{T}_d$ a target vector and $\mathbf{H}$ a bandwidth matrix. A parametrized pdf $K_{\mathbf{x},\mathbf{H}}(\cdot)$ of support $\mathbb{S}_{\mathbf{x},\mathbf{H}} \left(\subseteq \mathbb{R}^d\right)$ is called "multivariate (or general) associated kernel" if the following conditions are satisfied :*

$$\mathbf{x} \in \mathbb{S}_{\mathbf{x},\mathbf{H}}, \tag{4.3}$$

$$\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x} + \mathbf{a}(\mathbf{x},\mathbf{H}), \tag{4.4}$$

$$\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{B}(\mathbf{x},\mathbf{H}), \tag{4.5}$$

where $\mathcal{Z}_{\mathbf{x},\mathbf{H}}$ denotes the random vector with pdf $K_{\mathbf{x},\mathbf{H}}$ and both $\mathbf{a}(\mathbf{x},\mathbf{H}) = (a_1(\mathbf{x},\mathbf{H}), \ldots, a_d(\mathbf{x},\mathbf{H}))^\top$ and $\mathbf{B}(\mathbf{x},\mathbf{H}) = \left(b_{ij}(\mathbf{x},\mathbf{H})\right)_{i,j=1,\ldots,d}$ tend, respectively, to the null vector $\mathbf{0}$ and the null matrix $\mathbf{0}_d$ as $\mathbf{H}$ goes to $\mathbf{0}_d$.

From this definition and in comparison with (4.2), the Nadaraya-Watson estimator using associated kernels is

$$\widetilde{m}_n(\mathbf{x}; K_{\mathbf{x},\mathbf{H}}) = \sum_{i=1}^{n} \frac{Y_i K_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i)}{\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i)} = \widetilde{m}_n(\mathbf{x}; \boldsymbol{\kappa}, \mathbf{H}), \quad \forall \mathbf{x} \in \mathbb{T}_d \subseteq \mathbb{R}^d, \tag{4.6}$$

where $\mathbf{H} \equiv \mathbf{H}_n$ is the bandwidth matrix such that $\mathbf{H}_n \to \mathbf{0}$ as $n \to \infty$, and $\boldsymbol{\kappa}$ represents the type of associated kernel $K_{\mathbf{x},\mathbf{H}}$, parametrized by $\mathbf{x}$ and $\mathbf{H}$. Without loss of generality and to point out the effect of $\boldsymbol{\kappa}$, we will in hereafter use $\widetilde{m}_n(\mathbf{x}; \boldsymbol{\kappa}, \mathbf{H}) = \widetilde{m}_n(\mathbf{x}; \boldsymbol{\kappa})$ since the bandwidth matrix is here investigated only by cross validation. Asymptotic bias and variance of the nonparametric estimation $\widetilde{m}_n(\mathbf{x}; \boldsymbol{\kappa})$ of (5.23) are given below in Theorem 6.2.1 for the specific multiple kernel to be used, for which its extension in multivariate situation follows the similar way as for density estimation in Kokonendji and Somé (2015).

The following two examples provide the well-known and also interesting particular cases of multivariate associated kernel estimators. The first can be seen as an interpretation of classical associated kernels through continuous symmetric kernels. The second deals on non-classical associated kernels without correlation structure.

Given a target vector $\mathbf{x} \in \mathbb{R}^d =: \mathbb{T}_d$ and a bandwidth matrix $\mathbf{H}$, it follows that the classical kernel in (4.2) with support $\mathbb{S}_d \subseteq \mathbb{R}^d$, with null mean vector and covariance matrix $\boldsymbol{\Sigma}$ induces the so-called (multivariate) classical associated kernel :

$$(i) \quad K_{\mathbf{x},\mathbf{H}}(\cdot) = \frac{1}{\det \mathbf{H}} K\left\{\mathbf{H}^{-1}(\mathbf{x} - \cdot)\right\} \tag{4.7}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbf{x} - \mathbf{H}\mathbb{S}_d$ with $\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x}$ (i.e. $\mathbf{a}(\mathbf{x},\mathbf{H}) = \mathbf{0}$) and $\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}$;

$$(ii)\ \ K_{\mathbf{x},\mathbf{H}}(\cdot) = \frac{1}{(\det \mathbf{H})^{1/2}} K\left\{\mathbf{H}^{-1/2}(\mathbf{x} - \cdot)\right\}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbf{x} - \mathbf{H}^{1/2}\mathbb{S}_d$ with $\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x}$ (i.e. $\mathbf{a}(\mathbf{x},\mathbf{H}) = \mathbf{0}$) and $\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{H}^{1/2}\boldsymbol{\Sigma}\mathbf{H}^{1/2}$.

A second particular case of Definition 4.2.1, appropriate for both continuous and count explanatory variables without correlation structure is presented as follows.

Let $\mathbf{x} = (x_1, \ldots, x_d)^\top \in \times_{j=1}^d \mathbb{T}_1^{[j]} =: \mathbb{T}_d$ and $\mathbf{H} = \mathbf{Diag}(h_{11}, \ldots, h_{dd})$ with $h_{jj} > 0$. Let $K_{x_j,h_{jj}}^{[j]}$ be a (discrete or continuous) univariate associated kernel (see Definition 4.2.1 for $d = 1$) with its corresponding random variable $\mathcal{Z}_{x_j,h_{jj}}^{[j]}$ on $\mathbb{S}_{x_j,h_{jj}} (\subseteq \mathbb{R})$ for all $j = 1, \ldots, d$. Then, the multiple associated kernel is also a multivariate associated kernel :

$$K_{\mathbf{x},\mathbf{H}}(\cdot) = \prod_{j=1}^{d} K_{x_j,h_{jj}}^{[j]}(\cdot) \tag{4.8}$$

on $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \times_{j=1}^d \mathbb{S}_{x_j,h_{jj}}$ with $\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = (x_1 + a_1(x_1, h_{11}), \ldots, x_d + a_d(x_d, h_{dd}))^\top$ and $\mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right)$ $= \mathbf{Diag}\left(b_{jj}(x_j, h_{jj})\right)_{j=1,\ldots,d}$. In other words, the random variables $\mathcal{Z}_{x_j,h_{jj}}^{[j]}$ are independent components of the random vector $\mathcal{Z}_{\mathbf{x},\mathbf{H}}$.

Here, in addition to the Nadaraya-Watson estimator using general associated kernels given in (4.6), we proposed a slight one. In fact, for multivariate supports composed of continuous and discrete univariate support, we lack appropriate general associated kernels. Therefore, the estimator (4.6) becomes with multiple associated kernels (4.8) :

$$\widetilde{m}_n(\mathbf{x};\boldsymbol{\kappa}) = \sum_{i=1}^{n} \frac{Y_i \prod_{j=1}^d K_{x_j,h_{jj}}^{[j]}(X_{ij})}{\sum_{i=1}^n \prod_{j=1}^d K_{x_j,h_{jj}}^{[j]}(X_{ij})}, \quad \forall \mathbf{x} = (x_1, \ldots, x_d)^\top \in \mathbb{T}_d := \times_{j=1}^d \mathbb{T}_1^{[j]}. \tag{4.9}$$

In theory and in practice, one often uses (4.9) from multiple associated kernels (4.8) which are more manageable than (4.6) ; see, e.g., Scott (1992) and also Bouerzmarni and Rombouts (2010) for density estimation. Thus, we now present the main asymptotic properties of $\widetilde{m}_n(\mathbf{x};\boldsymbol{\kappa})$ in (4.9) for ℓ-discrete and $(d-\ell)$-continuous associated kernels, which can derive from the mean squared error (MSE) as sum of squared bias and variance.

Theorem 4.2.2 *Let $f = (f_1, f_2, \ldots, f_d)$ such that f_j is a univariate pdf of a X_j for $j = 1, \ldots, \ell$ of discrete parts and $j = 1, \ldots, d-\ell$ of continuous ones. Assume that $f_j(x_j) > 0$ for $j = 1, \ldots, d-\ell$ and $f_j(x_j) = \Pr(X_j = x_j) > 0$ for $j = 1, \ldots, \ell$. Furthermore, suppose that the bandwidth $\mathbf{H}_n = \mathbf{Diag}(h_{11}, \ldots, h_{dd}) \to \mathbf{0}$ as $n \to \infty$, and that the multiple associated kernel fulfills* (5.1)-(5.3) *of Definition 4.2.1. Then, the bias and variance of $\widetilde{m}_n(\mathbf{x};\boldsymbol{\kappa})$ in* (4.9) *admit the following behaviours :*

$$\mathrm{Bias}\{\widetilde{m}_n(\mathbf{x};\boldsymbol{\kappa})\} \approx \frac{1}{2}\sum_{j=1}^{d}\left\{m_{jj}^{(2)}(\mathbf{x}) + 2m_j^{(1)}(\mathbf{x})\left(\frac{f_j^{(1)}}{f}\right)(\mathbf{x})\right\}\mathrm{Var}\left(\mathcal{Z}_{x_j,h_{jj}}^{[j]}\right),$$

$$\mathrm{Var}\{\widetilde{m}_n(\mathbf{x};\boldsymbol{\kappa})\} \approx \frac{\mathrm{Var}(Y|\mathbf{X}=\mathbf{x})}{nf(\mathbf{x})}\prod_{j=1}^{d}\|K^{[j]}_{x_j,h_{jj}}\|_2^2$$

with

$$\|K_{x,h}\|_2^2 = \begin{cases} [\Pr(\mathcal{Z}_{x,h}=x)]^2 & \text{for discrete} \\ \int_{\mathbb{S}_{x,h}} K^2_{x,h}(u)du & \text{for continuous,} \end{cases}$$

and where $f_j^{(1)}$, $m_j^{(1)}$ and $m_{jj}^{(2)}$ are usual j-partial differentiation on $\mathbb{R}$ for $j=1,2,\ldots,d-\ell$ (i.e., $\partial/\partial x_j$ of order 1 and $\partial^2/\partial x_j^2$ of order 2), and the corresponding j-partial finite differences for $j=1,2,\ldots,\ell$, which are defined in the sense of any univariate count component $g:\mathbb{N}\to\mathbb{R}$ by

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

and

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

Proof. Without loss of generality, we consider the univariate case leading to the multiple case (4.9). For the discrete parts of bias and variance with $j=1,2,\ldots,\ell$, we refer to Proposition 3.1 of Kokonendji *et al.* (2009). As for $j=1,2,\ldots,d-\ell$ of univariate continuous associated (classical or not) kernels, it is enough to adapt the previous discrete (associated kernel) expansions of bias and variance to the continuous ones and replace finite differences to their corresponding usual differentiations on $\mathbb{R}$. ■

4.2.2 Associated kernels for illustration

In order to point out the importance of the type of kernel $\boldsymbol{\kappa}$ in a regression study, we motivate below some kernels that will be used in simulations. These concern seven basic associated kernels for which three of them are univariate discrete, three others are univariate continuous and the last one is a bivariate beta with correlation structure.

– The binomial kernel (`Bin`) is defined on the support $\mathbb{S}_x=\{0,1,\ldots,x+1\}$ with $x\in\mathbb{T}_1:=\mathbb{N}=\{0,1,\ldots\}$ and then $h\in(0,1]$:

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

where $\mathbb{1}_A$ denotes the indicator function of any given event A. Note that $B_{x,h}$ is the probability mass function (pmf) of the binomial distribution $\mathcal{B}(x+1;(x+h)/(x+1))$ with its number of trials $x+1$ and its success probability in each trial $(x+h)/(x+1)$. It is appropriated for count data with small or moderate sample sizes and, also, it does not satisfy (5.3) ; see Kokonendji & Senga Kiéssé (2011) and also Zougab *et al.* (2012) for a bandwidth selection by Bayesian method.

– For fixed arm $a \in \mathbb{N}$, the discrete triangular kernel (`DTra`) is defined on $\mathbb{S}_{x,a} = \{x, x \pm 1, \ldots, x \pm a\}$ with $x \in \mathbb{T}_1 = \mathbb{N}$:

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

where $h > 0$ and $P(a,h) = (2a+1)(a+1) - 2\sum_{k=0}^{a} k^h$ is the normalizing constant. It is symmetric around the target x, satisfying Definition 4.2.1 and suitable for count variables ; see Kokonendji *et al.* (2007) and also Kokonendji & Zocchi (2010) for an asymmetric version.

– From Aitchison & Aitken (1976), Kokonendji & Senga Kiéssé (2011) deduced the following discrete kernel that we here label DiracDU (`DirDU`) as "Dirac Discrete Uniform". For fixed $c \in \{2, 3, \ldots\}$ the number of categories, we define $\mathbb{S}_c = \{0, 1, \ldots, c-1\}$ and

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

where $h \in (0,1]$ and $x \in \mathbb{T}_1$. This DiracDU kernel is symmetric around the target, satisfying Definition 4.2.1 and appropriated for categorical set $\mathbb{T}_1$. See, e.g., Li & Racine (2004) for some uses.

– From the well known Epanechnikov (1969) kernel $K^E(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{[-1,1]}(u)$, we define its associated version (`Epan`) on $\mathbb{S}_{x,h} = [x-h, x+h]$ with $x \in \mathbb{T}_1 := \mathbb{R}$ and $h > 0$:

$$K^E_{x,h}(u) = \frac{3}{4h}\left\{1 - \left(\frac{u-x}{h}\right)^2\right\}\mathbb{1}_{[x-h,x+h]}(u).$$

It is obtained through (4.7) and is well adapted for continuous variables with unbounded supports.

– The gamma kernel (`Gamma`) is defined on $\mathbb{S}_{x,h} = [0,\infty) = \mathbb{T}_1$ with $x \in \mathbb{T}_1$ and $h > 0$:

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

where $\Gamma(\cdot)$ is the classical gamma function. It is the pdf of the gamma distribution $\mathcal{G}a(1 + x/h, h)$ with scale parameter $1 + x/h$ and shape parameter h. It satisfies Definition 4.2.1 and suits for non-negative real set $\mathbb{T}_1$; see Chen (2000a).

– The beta kernel (`Beta`) is however defined on $\mathbb{S}_{x,h} = [0,1] = \mathbb{T}_1$ with $x \in \mathbb{T}_1$ and $h > 0$:

$$BE_{x,h}(u) = \frac{u^{x/h}(1-u)^{(1-x)/h}}{\mathscr{B}(1+x/h, 1+(1-x)/h)}\mathbb{1}_{[0,1]}(u),$$

where $\mathscr{B}(r,s) = \int_0^1 t^{r-1}(1-t)^{s-1}dt$ is the usual beta function with $r > 0$ and $s > 0$. It is the pdf of the beta distribution $\mathcal{B}e(1 + x/h, (1-x)/h)$ with shape parameters $1 + x/h$ and $(1-x)/h$. This pdf satisfies Definition 4.2.1 and is appropriated for rates, proportions and percentages dataset $\mathbb{T}_1$; see Chen (1999).

– We finally consider the bivariate beta kernel (`Bivariate beta`) defined by

$$\begin{aligned} BS_{\mathbf{x},\mathbf{H}}(u_1,u_2) \;=\; & \left(\frac{u_1^{x_1/h_{11}}(1-u_1)^{(1-x_1)/h_{11}}}{\mathscr{B}(1+x_1/h_{11}, 1+(1-x_1)/h_{11})}\right)\left(\frac{u_2^{x_2/h_{22}}(1-u_2)^{(1-x_2)/h_{22}}}{\mathscr{B}(1+x_2/h_{22}, 1+(1-x_2)/h_{22})}\right) \\ & \times\left(1 + h_{12} \times \frac{u_1 - \widetilde{\mu}_1(x_1,h_{11})}{h_{11}^{1/2}\widetilde{\sigma}_1(x_1,h_{11})} \times \frac{u_2 - \widetilde{\mu}_2(x_2,h_{22})}{h_{22}^{1/2}\widetilde{\sigma}_2(x_2,h_{22})}\right)\mathbb{1}_{[0,1]^2}(u_1,u_2), \quad (4.10) \end{aligned}$$

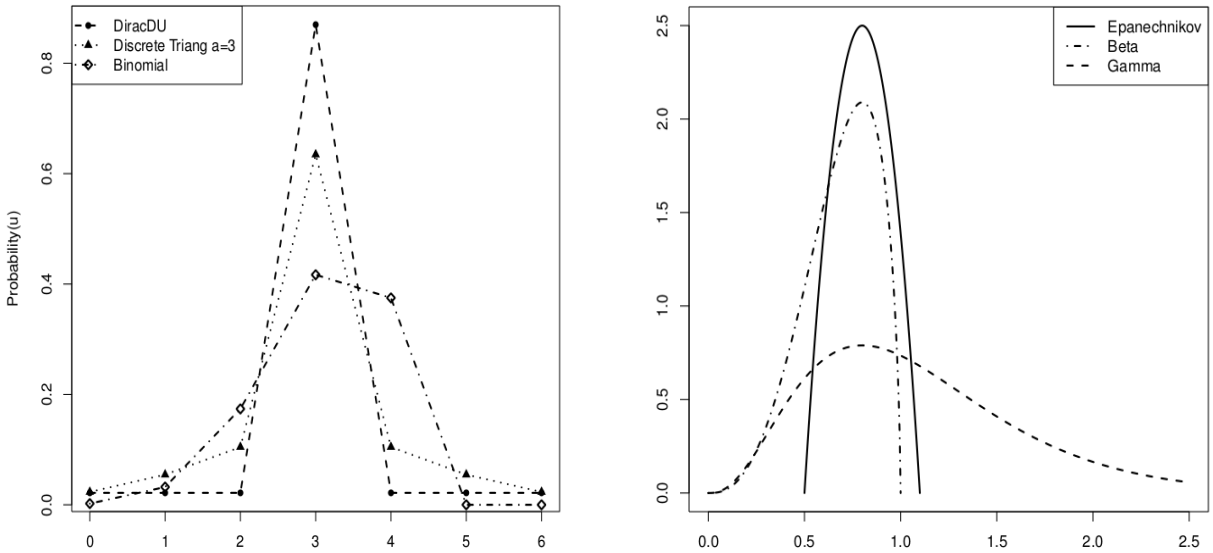
with $\mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbb{T}_2 = [0,1]^2$, $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{T}_2$ and $\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}$. For $j = 1, 2$, the characteristics in (4.10) are given by $h_{jj} > 0$, $\widetilde{\mu}_j(x_j, h_{jj}) = (x_j + h_{jj})/(1 + 2h_{jj})$, $\widetilde{\sigma}_j^2(x_j, h_{jj}) = (x_j + h_{jj})(1 + h_{jj} - x_j)(1 + 2h_{jj})^{-2}(1 + 3h_{jj})^{-1}h_{jj}$, and the constraints

$$h_{12} \in [-\beta, \beta'] \cap \left(-\sqrt{h_{11}h_{22}}\,,\, \sqrt{h_{11}h_{22}}\right) \tag{4.11}$$

with $\beta = \left(\max_{v_1,v_2} \left\{ \dfrac{v_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2}\widetilde{\sigma}_1(x_1, h_{11})} \times \dfrac{v_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2}\widetilde{\sigma}_2(x_2, h_{22})} \right\} \right)^{-1}$ and

$\beta' = \left| \left(\min_{v_1,v_2} \left\{ \dfrac{v_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2}\widetilde{\sigma}_1(x_1, h_{11})} \times \dfrac{v_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2}\widetilde{\sigma}_2(x_2, h_{22})} \right\} \right)^{-1} \right|$.

It satisfies Definition 4.2.1 and is adapted for bivariate rates. The full bandwidth matrix $\mathbf{H}$ allows any orientation of the kernel. Therefore, it can reach any point of the space which might be inaccessible with diagonal matrix. This type of kernel is called beta-Sarmanov kernel by Kokonendji & Somé (2015) ; see Sarmanov (1966) and also Lee (1996) for this construction of multivariate densities with correlation structure from independent components.



[(a)] [(b)]

FIGURE 4.1 – Shapes of univariate (discrete and continuous) associated kernels : (a) DiracDU, discrete triangular $a = 3$ and binomial with same target $x = 4$ and bandwidth $h = 0.13$; (b) Epanechnikov, beta and gamma with same $x = 0.8$ and $h = 0.3$.

Figure 4.1 shows some forms of the above-mentioned univariate associated kernels. The plots highlight the h, the classical associated kernel of Epanechnikov, and also the categorical DiracDU kernel, keep their respective same shapes

4.2.3 Bandwidth matrix selection by cross validation

In the context of multivariate kernel regression, the bandwidth matrix selection is here obtained by the well-known least squares cross-validation. In fact, for a given associated kernel, the optimal bandwidth matrix is $\widehat{\mathbf{H}} = \arg\min_{\mathbf{H}\in\mathcal{H}} \mathrm{LSCV}(\mathbf{H})$ with

$$\mathrm{LSCV}(\mathbf{H}) = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \widetilde{m}_{-i}(\mathbf{X}_i;\boldsymbol{\kappa})\right\}^2, \tag{4.12}$$

where $\widetilde{m}_{-i}(\mathbf{X}_i;\boldsymbol{\kappa})$ is computed as $\widetilde{m}_n$ of (4.6) excluding $\mathbf{X}_i$ and, $\mathcal{H}$ is the set of bandwidth matrices $\mathbf{H}$; see, e.g., Kokonendji *et al.* (2009) in univariate case and also *et al.* (2014) and Zougab *et al.* (2014a) for univariate bandwidth estimation by sampling algorithm methods. For diagonal bandwidth matrices (i.e. multiple associated kernels) the LSCV method use the set of diagonal matrices $\mathcal{D}$. Concerning the beta-Sarmanov kernel (4.10) with full bandwidth matrix, this LSCV method is used under $\mathcal{H}_1$, a subset of $\mathcal{H}$ verifying the constraint (4.11) of the associated kernel. Their algorithms are described below and used for numerical studies in the following section.

Algorithms of LSCV method (4.12) for some type of associated kernels and their correponsding bandwidth matrices

A1. Bivariate beta (4.10) with full bandwidth matrices and dimension $d = 2$.

1. Choose two intervals H_{11} and H_{22} related to h_{11} and h_{22}, respectively.
2. For $\delta = 1,\ldots,\ell(H_{11})$ and $\gamma = 1,\ldots,\ell(H_{22})$,
   (a) Compute the interval $H_{12}[\delta,\gamma]$ related to h_{12} from constraints in (4.11);
   (b) For $\lambda = 1,\ldots,\ell(H_{12}[\delta,\gamma])$,
   Compose the full bandwidth matrix $\mathbf{H}(\delta,\gamma,\lambda) := \left(h_{ij}(\delta,\gamma,\lambda)\right)_{i,j=1,2}$ with $h_{11}(\delta,\gamma,\lambda) = H_{11}(\delta)$, $h_{22}(\delta,\gamma,\lambda) = H_{22}(\gamma)$ and $h_{12}(\delta,\gamma,\lambda) = H_{12}[\delta,\gamma](\lambda)$.
3. Apply LSCV method on the set $\mathcal{H}_1$ of all full bandwidth matrices $\mathbf{H}(\delta,\gamma,\lambda)$.

A2. Multiple associated kernels (i.e. diagonal bandwidth matrices) for $d \geq 2$.

1. Choose two intervals $H_{11}, \ldots, H_{dd}$ related to $h_{11}, \ldots, h_{dd}$, respectively.
2. For $\delta_1 = 1,\ldots,\ell(H_{11}), \ldots, \delta_d = 1,\ldots,\ell(H_{dd})$,
   Compose the diagonal bandwidth matrix $\mathbf{H}(\delta_1,\ldots,\delta_d) := \mathbf{Diag}\left(H_{11}(\delta_1),\ldots,H_{dd}(\delta_d)\right)$.
3. Apply LSCV method on the set $\mathcal{D}$ of all diagonal bandwidth matrices $\mathbf{H}(\delta_1,\ldots,\delta_d)$.

For a given interval I, the notation $\ell(I)$ is the total number of subdivisions of I and $I(\eta)$ denotes the real value at the subdivision η of I. Also, for practical uses of (A1) and (A2), the intervals $H_{11},\ldots,H_{dd}$ are taken generally according to the chosen associated kernel. It should be noted that, other bandwidth matrices selection by Bayesian methods is possible; see, e.g., Ziane *et al.* (2015) for the adaptive case using asymmetric kernel density estimation for heavy tailed data.

4.3 Simulation studies and real data analysis

We apply the multivariate associated kernel estimators $\widetilde{m}_n$ of (4.6) and (4.9) to some simulated target regressions functions m and then to two real datasets. The multivariate and multiple associated kernels used are built from those of Section 4.2.2. The optimal bandwidth matrix is here chosen by LSCV method (4.12) using Algorithms A1 and A2 of Section 4.2.3 and their indications. Besides the criterion of kernel support, we retain three measures to examine the effect of different associated kernels $\boldsymbol{\kappa}$ on multiple regression. In simulations, it is the average squared errors (ASE) defined as

$$ASE(\boldsymbol{\kappa}) = \frac{1}{n}\sum_{i=1}^{n}\left\{m(\mathbf{x}_i) - \widetilde{m}_n(\mathbf{x}_i;\boldsymbol{\kappa})\right\}^2.$$

For real datasets, we use the root mean squared error (RMSE) which linked to ASE through squared root and by changing the simulated value $m(x_i)$ into the observed value y_i :

$$RMSE(\boldsymbol{\kappa}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{y_i - \widetilde{m}_n(\mathbf{x}_i;\boldsymbol{\kappa})\right\}^2}.$$

Also, we consider the practical coefficient of determination R^2 which quantifies the proportion of variation of the response variable Y_i explained by the non-intercept regressor $\mathbf{x}_i$

$$R^2(\boldsymbol{\kappa}) = \frac{\sum_{i=1}^{n}\left\{\widetilde{m}_n(\mathbf{x}_i;\boldsymbol{\kappa}) - \overline{y}\right\}^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2},$$

with $\overline{y} = n^{-1}(y_1 + \ldots + y_n)$. All these criteria above have their simulated or real data counterparts by replacing y_i with $m(\mathbf{x}_i)$ and vice versa. Computations have been performed on the supercomputer facilities of the Mésocentre de calcul de Franche-Comté using the R software ; see R Development Core Team (2015).

4.3.1 Simulation studies

Expect as otherwise, each result is obtained with the number of replications $N_{sim} = 100$.

Bivariate cases

We consider seven target regression functions labelled A, B, C, D and E with dimension $d = 2$. Our interest is to estimate or smooth the regression function m in (4.1) by taking into account especially its structure.

– Function A is a bivariate beta without correlation $\rho(x_1, x_2) = 0$:

$$m(x_1, x_2) = \frac{x_1^{p_1-1}(1-x_1)^{q_1-1}x_2^{p_2-1}(1-x_2)^{q_2-1}}{\mathscr{B}(p_1,q_1)\mathscr{B}(p_2,q_2)}\,\mathbb{1}_{[0,1]}(x_1)\mathbb{1}_{[0,1]}(x_2),$$

with $(p_1, q_1) = (3, 2)$ and $(p_2, q_2) = (5, 2)$ as parameter values in univariate beta density.

– Function B is the bivariate Dirichlet density

$$m(x_1,x_2)=\frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}x_1^{\alpha_1-1}x_2^{\alpha_2-1}(1-x_1-x_2)^{\alpha_3-1}\mathbb{1}_{\{x_1,\ x_2\geq 0,\ x_1+x_2\leq 1\}}(x_1,x_2),$$

where $\Gamma(\cdot)$ is the classical gamma function, with parameter values $\alpha_1=\alpha_2=5$, $\alpha_3=6$ and, therefore, the moderate value of $\rho(x_1,x_2)=-(\alpha_1\alpha_2)^{1/2}(\alpha_1+\alpha_3)^{-1/2}(\alpha_2+\alpha_3)^{-1/2}=-0.454$.

– Function C is a bivariate Poisson with null correlation $\rho(x_1,x_2)=0$:

$$m(x_1,x_2)=\frac{e^{-5}2^{x_1}3^{x_2}}{x_1!x_2!}\mathbb{1}_{\mathbb{N}}(x_1)\mathbb{1}_{\mathbb{N}}(x_2).$$

– Function D is a bivariate Poisson with correlation structure

$$m(x_1,x_2)=e^{-(\theta_1+\theta_2+\theta_{12})}\sum_{i=0}^{min(x_1,x_2)}\frac{\theta_1^{x_1+i}\theta_2^{x_2+i}\theta_{12}^{i}}{(x_1+i)!(x_2+i)!i!}\mathbb{1}_{\mathbb{N}\times\mathbb{N}}(x_1,x_2),$$

with parameter values $\theta_1=2$, $\theta_2=3$ and $\theta_{12}=4$ and, therefore, the moderate value of $\rho(x_1,x_2)=\theta_{12}(\theta_1+\theta_{12})^{-1/2}(\theta_2+\theta_{12})^{-1/2}=0.617$; see, e.g., Yahav and Shmueli (2012).

– Function E is a bivariate beta without correlation $\rho(x_1,x_2)=0$:

$$m(x_1,x_2)=\frac{x_1^{p_1-1}(1-x_1)^{q_1-1}3^{x_2}}{e^3\mathscr{B}(p_1,q_1)x_2!}\ \mathbb{1}_{[0,1]}(x_1)\mathbb{1}_{\mathbb{N}}(x_2),$$

with $(p_1,q_1)=(3,3)$.

n	Bivariate beta	Beta×Beta
50	276.198	7.551
100	647.255	30.081

TABLE 4.1 – Typical Central Processing Unit (CPU) times (in seconds) for one replication of LSCV method (4.12) by using Algorithms A1 and A2 of Section 4.2.3.

Table 4.1 presents the execution times needed for computing the LSCV method for both bivariate beta kernels with respect to only one replication of sample sizes $n=50$ and 100 for the target function A. The computational times of the LSCV method for the bivariate beta with correlation structure (4.10) are obviously longer than those without correlation structure. Let us note that for full bandwidth matrices, the execution times become very large when the number of observations is large ; however, these CPU times can be considerably reduced by parallelism processing, in particular for the bivariate beta kernel with full LSCV method (4.12). These constraints (4.11) reflect the difficulty for finding the appropriate bandwidth matrix with correlation structure by LSCV method.

Table 4.2 reports the average $ASE(\boldsymbol{\kappa})$ which we denote $\overline{ASE}(\boldsymbol{\kappa})$ for three continuous associated kernels $\boldsymbol{\kappa}$ with respect to functions A and B and according to sample sizes

	n	Bivariate beta	Beta×Beta	*Epan*×*Epan*
A	50	0.4368(0.3754)	0.4266(0.3724)	0.7483(0.2342)
	100	0.1727(0.0664)	0.1952(0.0816)	0.6727(0.1413)
B	50	1.2564(0.5875)	1.4267(0.4024)	1.6675(2.0353)
	100	0.3041(0.1151)	0.3362(0.1042)	1.3975(1.5758)

TABLE 4.2 – Some expected values of $\overline{ASE}(\boldsymbol{\kappa})$ and their standard errors in parentheses with $N_{sim} = 100$ of some multiple associated kernel regressions for simulated continuous data from functions A with $\rho(x_1, x_2) = 0$ and B with $\rho(x_1, x_2) = -0.454$.

$n \in \{50, 100\}$. We can see that both beta kernels in dimension $d = 2$ work better than the multiple Epanechnikov kernel for all sample sizes and all correlation structure in the regressors. This reflects the appropriateness of the beta kernels which are suitable to the support of rate regressors. Then, the explanatory variables with correlation structure give larger $\overline{ASE}(\boldsymbol{\kappa})$ than those without correlation structure. Also, both beta kernels give quite similar results. Furthermore, all $\overline{ASE}(\boldsymbol{\kappa})$ are better when the sample size increases.

Finally, Tables 4.1 and 4.2 highlight that the use of bivariate beta kernels with correlation structure is not recommend in regression with rates explanatory variables. Thus, we focus on multiple associated kernels for the rest of the simulations studies.

	n	DTr2×DTr2	DTr3×DTr3	Bin×Bin	*Epan*×*Epan*	*DirDU*×*DirDU*
C	20	1.5e-6(2.2e-6)	3.3e-6(4.1e-6)	3.6e-5(9.7e-6)	4.0e-5(3.5e-5)	1.6e-8(1.8e-8)
	50	3.1e-7(6.9e-7)	4.7e-7(9.7e-7)	3.6e-5(7.4e-6)	3.8e-5(2.8e-5)	3.7e-9(2.3e-9)
	100	8.6e-8(1.2e-7)	2.9e-7(3.1e-7)	3.7e-5(4.8e-6)	3.6e-5(2.3e-5)	4.1e-10(3.5e-10)
D	20	2.4e-6(2.8e-6)	4.5e-6(4.9e-6)	7.1e-6(2.6e-6)	4.2e-6(2.5e-6)	2.7e-8(2.1e-8)
	50	2.5e-7(3.4e-7)	1.8e-7(2.5e-7)	8.1e-5(4.3e-6)	5.1e-6(1.2e-6)	4.3e-9(3.2e-9)
	100	2.6e-8(6.2e-8)	4.8e-8(9.5e-8)	9.3e-6(8.2e-7)	7.2e-6(7.8e-7)	5.3e-10(4.6e-10)

TABLE 4.3 – Some expected values of $\overline{ASE}(\boldsymbol{\kappa})$ and their standard errors in parentheses with $N_{sim} = 100$ of some multiple associated kernel regressions for simulated count data from functions C with $\rho(x_1, x_2) = 0$ and D with $\rho(x_1, x_2) = 0.617$.

Table 4.3 shows the values $\overline{ASE}(\boldsymbol{\kappa})$ with respect to five associated kernels $\boldsymbol{\kappa}$ for sample size $n = 20, 50$ and 100 and count datasets generated from C and D. Globally, the discrete associated kernels in multiple case perform better than the multiple Epanechnikov kernel for all sample sizes and correlation structure in the regressors. The use of categorical DiracDU kernels gives the best result in term of $\overline{ASE}(\boldsymbol{\kappa})$ but DiracDU does not suit for these count datasets. Also, the discrete triangular kernels gives the most interesting result with an advantage to the discrete triangular with small arm $a = 2$. This discrete triangular is the best since it concentrates always on the target and a few observations around it ; see Figure 4.1(a). The results become much better when

the sample size increases. The values $\overline{ASE}(\boldsymbol{\kappa})$ for regressors with or without correlation structure are comparable ; and thus, we can focus on target regression functions without correlation structure for the remaining simulations.

	n	Beta×DTr2	Beta× DTr3	Beta×Bin	Beta×*Epan*	Beta×*DirDU*
	30	3.738(1.883)	1.966(1.382)	3.884(1.298)	6.361(2.134)	0.162(0.201)
E	50	3.978(1.404)	2.106(1.119)	3.683(0.833)	7.143(1.732)	0.138(0.171)
	100	3.951(1.052)	1.956(0.806)	3.835(0.834)	7.277(1.574)	0.113(0.147)

TABLE 4.4 – Some expected values ($\times 10^3$) of $\overline{ASE}(\boldsymbol{\kappa})$ and their standard errors in parentheses with $N_{sim} = 100$ of some multiple associated kernel regressions of simulated mixed data from function E with $\rho(x_1, x_2) = 0$.

Table 4.4 presents the values for sample sizes $n \in \{30, 50, 100\}$ and for five associated kernels $\boldsymbol{\kappa}$. The datasets are generated from E and the beta kernel is applied on the continuous rate variable of E. We observe the superiority of the multiple associated kernels using discrete kernels over those defined with the Epanechnikov kernel for all sample sizes. Then, the multiple associated kernel with the categorical DiracDU gives the best $\overline{ASE}(\boldsymbol{\kappa})$ but it is not appropriate for the count variable of E. Also, the values $\overline{ASE}(\boldsymbol{\kappa})$ are getting better when the sample size increases.

From Tables 4.2, 4.3 and 4.4, the importance of the type of associated kernel $\boldsymbol{\kappa}$ which respect the support of the explanatory variables is proven.

Multivariate cases

Since the appropriate associated kernels perform better than the inappropriate ones, we focus in higher dimension $d > 2$ on regression with only suitable associated kernels. Then, we consider two target regression functions labelled F and G for $d = 3$ and 4 respectively. The formulas of the functions are given below.

– Function F is a 3-variate with null correlation :

$$m(x_1, x_2, x_3) = \frac{x_1^{p_1-1}(1-x_1)^{q_1-1}2^{x_2}3^{x_3}}{e^5\mathscr{B}(p_1, q_1)x_2!x_3!}\mathbb{1}_{[0,1]}(x_1)\mathbb{1}_{\mathbb{N}}(x_2)\mathbb{1}_{\mathbb{N}}(x_3),$$

with $(p_1, q_1) = (3, 2)$.

– Function G is a 4-variate without correlation :

$$m(x_1, x_2, x_3, x_4) = \frac{x_1^{p_1-1}(1-x_1)^{q_1-1}x_2^{p_2-1}(1-x_2)^{q_2-1}2^{x_3}3^{x_4}}{e^5\mathscr{B}(p_1, q_1)\mathscr{B}(p_2, q_2)x_3!x_4!}\mathbb{1}_{[0,1]}(x_1)\mathbb{1}_{[0,1]}(x_2)\mathbb{1}_{\mathbb{N}}(x_3)\mathbb{1}_{\mathbb{N}}(x_4),$$

with $(p_1, q_1) = (3, 2)$ and $(p_2, q_2) = (5, 2)$.

Table 4.5 presents the regression study for dimension $d = 3$ and 4 with respect to functions F and G and for sample size $n \in \{30, 50, 100\}$. The values $\overline{ASE}(\boldsymbol{\kappa})$ show the superiority of the multiple associated kernels using the discrete triangular kernel with $a = 3$ over the one with the binomial kernel. Some results with respect to function G for an associated kernel $\boldsymbol{\kappa}$ composed by two beta and two discrete triangular kernels with $a = 3$ are also provided. The errors become smaller when the sample size increases.

n	Beta×DTr3×DTr3	Beta×Bin×DTr3	Beta×Beta×DTr3×DTr3
30	0.2501(0.1264)	0.3038(0.1258)	0.7448(0.5481)
50	0.2381(0.0661)	0.2895(0.0162)	0.6055(0.2291)
100	0.2282(0.0649)	0.2822(0.0608)	0.5012(0.2166)

TABLE 4.5 – Some expected values ($\times 10^3$) of $\overline{ASE}(\boldsymbol{\kappa})$ and their standard errors in parentheses with $N_{sim} = 100$ of some multiple associated kernel regressions of simulated mixed data from 3-variate F and 4-variate G.

4.3.2 Real data analysis

The dataset consists on a sample of 38 family economies from a US large city and is available as the *FoodExpenditure* object in the *betareg* package of Cribari and Neto (2010). The dataset in its current form gives not available (NA) responses for associated kernel regressions especially when we use the discrete triangular or the DiracDU kernel. Then, we extend the original *FoodExpenditure* dataset with its first 20 observations which guarantees some results for the regression, and thus $n = 58$. The dependent variable is *food/income*, the proportion of household *income* spent on *food*. Two explanatory variables are available : the previously mentioned household *income* (x_1) and the *number of residents* (x_2) living in the household with $\widehat{\rho}(x_1, x_2) = 0.028$. We use the Gamma or the Epanechnikov kernel for the continuous variable *income* and the discrete (of Figure 4.1(a)) or the Epanechnikov for the count variable *number of residents*.

The results of the multiple associated kernels for regression are divided in two in Table 4.6. The appropriate associated kernels which strictly follow the support of each variable give comparable results in terms of both $RMSE(\boldsymbol{\kappa})$ and $R^2(\boldsymbol{\kappa})$. In fact, the associated kernels that use the discrete triangular with arm $a = 2$ and 3 give some $R^2(\boldsymbol{\kappa})$ approximately equal to 64%. The inappropriate kernels give various results. The multiple Epanechnikov kernel and the type of kernel with DiracDU give $R^2(\boldsymbol{\kappa})$ higher than 80% while the Gamma×Epanechnikov gives $R^2(\boldsymbol{\kappa})$ less than 50%. Then, a little difference in terms of $RMSE(\boldsymbol{\kappa})$ can induce a high incidence on the $R^2(\boldsymbol{\kappa})$.

	Gamma×DTr2	Gamma×DTr3	Gamma×Bin
Appropriate	0.01409	0.01426	0.01730
	64.2681	64.2708	56.1091
	Epan×*Epan*	Gamma×*Epan*	Gamma×*DirDU*
Inappropriate	0.01451	0.03266	0.01278
	86.0011	47.0181	89.3462

TABLE 4.6 – Some expected values of $RMSE(\boldsymbol{\kappa})$ and in percentages $R^2(\boldsymbol{\kappa})$ of some multiple associated kernel regressions for the *FoodExpenditure* dataset with $n = 58$.

Table 4.7 of the second dataset aims to explain the turnover of a large company by two proportions explanatory variables obtained by survey. The first variable x_1 is the

x_{1i}	x_{2i}	y_i	x_{1i}	x_{2i}	y_i	x_{1i}	x_{2i}	y_i
68.1	54.6	0.8	80.3	9.1	1.5	78.6	31.0	1.6
60.8	4.4	1.3	23.1	83.1	2.3	44.9	3.2	1.0
34.4	36.2	1.2	16.9	90.4	2.0	78.2	13.9	1.8
59.4	27.5	1.3	9.4	79.2	2.8	60.2	35.2	1.1
4.7	81.0	2.9	55.8	21.9	1.3	65.6	26.1	1.6
19.9	97.4	1.2	27.5	75.0	2.2	74.4	12.6	1.6
20.6	73.6	2.4	59.1	12.9	1.4	83.5	13.3	1.8
16.4	42.9	1.1	2.7	93.9	2.4	10.9	83.5	2.6
29.9	74.4	2.0	13.9	56.9	1.4	27.0	77.1	2.2
84.8	26.6	1.6	14.0	92.9	2.1	3.1	67.0	2.2
46.1	66.9	1.2	22.9	43.9	1.1	14.8	72.9	2.5
10.2	86.3	2.5	53.8	56.2	1.0	80.6	16.5	1.6
89.4	32.5	1.6	23.7	61.5	1.5	64.1	28.6	1.5
30.9	46.3	1.1	39.6	67.2	1.4	15.6	90.5	2.0
24.3	37.8	1.2	59.5	45.1	0.9	3.9	68.6	2.5
27.4	74.6	1.9	17.3	81.2	2.6	66.9	43.7	0.9
47.7	61.7	1.1	93.7	28.5	1.5	1.5	65.8	2.3
33.1	83.8	1.5	28.7	82.7	2.0	35.6	43.7	1.0
0.3	83.3	3.0	61.3	70.9	0.6	13.9	25.0	0.8
76.9	35.4	1.2	67.1	24.0	1.7	13.2	70.8	2.2
29.5	44.6	1.3	85.8	36.5	1.2	34.5	73.7	1.8
19.6	67.7	1.9	35.5	76.9	1.8	55.6	6.9	1.3
96.2	26.1	1.7	18.8	55.9	1.3	30.7	9.1	0.9
85.9	28.0	1.5	50.4	17.7	1.4	43.5	15.1	1.0
5.6	39.1	1.1	67.2	8.7	1.5	31.5	36.7	1.2
99.9	7.15	1.3	13.1	59.4	1.7	30.0	21.5	0.8
61.0	31.1	1.4	13.7	75.8	2.5			

TABLE 4.7 – Proportions (in %) of folks who like the company, those who like its strong product and turnover of a company, designed respectively by the variables x_{1i}, x_{2i} and y_i, with $\widehat{\rho}(x_1, x_2) = -0.6949$ and $n = 80$.

rate of people who like the company and the second one x_2 is the percentage of people who like the strong product of this company. The dataset is obtained in 80 branch of this company. Obviously, there is a significant correlation between these explanatory variables : $\widehat{\rho}(x_1, x_2) = -0.6949$.

Table 4.8 presents the results for the nonparametric regressions with three associated kernels $\boldsymbol{\kappa}$. Both beta kernels offer the most interesting results with $R^2(\boldsymbol{\kappa})$ approximately

equal to 86%. Note that, the multiple Epanechnikov kernel gives lower performance mainly because this continuous unbounded kernel does not suit for these bounded explanatory variables.

Bivariate beta	Beta×Beta	*Epan*×*Epan*
0.10524	0.10523	0.18886
86.6875	86.6874	76.3431

TABLE 4.8 – Some expected values of $RMSE(\boldsymbol{\kappa})$ and in percentages $R^2(\boldsymbol{\kappa})$ of some bivariate associated kernel regressions for tunover dataset in Table 4.7 with $\widehat{\rho}(x_1, x_2) = -0.6949$ and $n = 80$.

4.4 Summary and final remarks

We have presented associated kernels for nonparametric multiple regression and in presence of a mixture of discrete and continuous explanatory variables ; see, e.g., Zougab *et al.* (2014b) for a choice of the bandwidth matrix by Bayesian methods. Two particular cases including the continuous classical and the multiple (or product of) associated kernels are highlight with the bandwidth matrix selection by cross-validation. Also, six univariate associated kernels and a bivariate beta with correlation structure are presented and used for computational studies.

Simulation experiments and analysis of two real datasets provide insight into the behaviour of the type of associated kernel $\boldsymbol{\kappa}$ for small and moderate sample sizes. Tables 4.1, 4.2 and 4.8 on bivariate rate regressions can be conceptually summarized as follows. The use of associated kernels with correlation structure is not recommend. In fact, it is time consuming and have the same performance as the multiple beta kernel. Also, these appropriate beta kernels are better than the inappropriate multiple Epanechnikov. For count regressions, the multiple associated kernels built from the binomial and the discrete triangular with small arms are superior to those with the optimal continuous Epanechnikov. Furthermore, the categorical DiracDU kernel gives misleading results since it does not suit for count variables, see Tables 4.3 and 4.4. We advise beta kernels for rates variables and gamma kernels for non-negative dataset for small and moderate sample sizes, and also for all dimension $d \geq 2$; see, e.g., Tables 4.5 and 4.6. Finally, more than the performance of the regression, it is the correct choice of the associated kernel according to the explanatory variables which is the most important. In other words, the criterion for choosing an associated kernel is the support ; however, for several kernels matching the support, we use common measures such as the mean integrated squared error. It should be noted that a large coefficient of determination R^2 does not mean good adjustment of the data ; see Tables 4.6 and 4.8.

Acknowledgements

We sincerely thank the Editor, an Associate Editor and two anonymous referees for their valuable comments.

Chapitre 5

Ake : An R package for discrete and continuous associated kernel estimations

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

5.1 Introduction

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

The `Ake` package recently developed, implements associated kernels that seam-

lessly deal with continuous (bounded, positive) and discrete (categorical, count) data types often found in applied settings ; see, for example, Libengué (2013) and Kokonendji & Senga Kiessé (2011). These associated kernels are used to smooth probability density functions (p.d.f.), probability mass functions (p.m.f.) or regression functions. The package is available from the Comprehensive `R` Archive Network (CRAN) at `http://cran.r-project.org/web/packages/` ; see R Development Core Team (2015). The coming versions of this package will contain, among others, p.d.f estimation of heavy tailed data (e.g., Ziane *et al.*, 2015) and others functionals estimations. The bandwidth selection remains crucial in associated kernel estimations of p.d.f., p.m.f. or regression function. Some methods have been investigated for selecting bandwidth parameter but the commonly used is the least squared cross-validation. A Bayesian approach has been also recently introduced by Zougab *et al.* (2012) in the case of binomial kernel. This method can be extended to various associated kernels with others functionnals. Despite the great number of packages implemented for nonparametric estimation in continuous cases with unbounded kernels, to the best of our knowledge, the `R` packages to estimate p.m.f. with categorical or count variables, p.d.f. with bounded or positive datasets, and regression have been far less investigated. We can refer to Wansouwé *et al.* (2015ab) and who implemented associated kernels for estimating p.m.f. and p.d.f. respectively, and partially included in this `Ake` package.

The rest of the paper is organized as follows. In Section 5.2, we recall the definition of associated kernels and then illustrate examples in both continuous and discrete cases which are discussed. Then, the associated kernel estimator for p.d.f. or p.m.f. is presented and illustrated with some `R` codes in Section 5.3. In particular, three bandwidth selection methods are available : cross-validation for any (continuous or discrete) associated kernel, Bayesian local for binomial and also a new theoretical Bayesian adaptive method for gamma kernel. Also, some practical and theoretical aspects of the normalizing constant in both p.d.f. and p.m.f. estimations are given. Section 5.4 investigates the case of regression functions with two bandwidth selection techniques : cross-validation and also Bayesian global for binomial kernel. Section 5.5 concludes with summary and final remarks.

5.2 Non-classical associated kernels

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

Définition 5.2.1 *(Kokonendji & Senga Kiessé 2011 ; Libengué 2013) Let $\mathbb{T}$ ($\subseteq \mathbb{R}$) be the support of the p.m.f., p.d.f. or regression function, to be estimated, $x \in \mathbb{T}$ a target and h a bandwidth. A parametrized p.m.f. (resp. p.d.f.) $K_{x,h}(\cdot)$ of support $\mathbb{S}_{x,h}$ ($\subseteq \mathbb{R}$) is called "associated kernel" if the following conditions are satisfied :*

$$x \in \mathbb{S}_{x,h}, \tag{5.1}$$

$$\mathbb{E}\left(\mathcal{Z}_{x,h}\right) = x + a(x,h), \tag{5.2}$$

$$Var\left(\mathcal{Z}_{x,h}\right) = b(x,h), \tag{5.3}$$

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

This definition has some interesting interpretations listed below.

Remarque 5.2.2 *(i) The function $K_{x,h}(\cdot)$ is not necessary symmetric and is intrinsically linked to x and h.*

(ii) The support $\mathbb{S}_{x,h}$ is not necessary symmetric around of x ; it can depend or not on x and h.

(iii) The condition (5.1) can be viewed as $\cup_{x\in\mathbb{T}}\mathbb{S}_{x,h} \supseteq \mathbb{T}$ and it implies that the associated kernel takes into account the support $\mathbb{T}$ of the density f, to be estimated.

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

(v) Both conditions (5.2) and (5.3) indicate that the associated kernel is more and more concentrated around of x as h goes to 0*. This highlights the peculiarity of associated kernel which can change its shape according to the target position.*

In order to construct an associated kernel $K_{x,h}(\cdot)$ from a parametric (discrete or continuous) probability distribution K_θ, $\theta \in \Theta \subset \mathbb{R}^d$ on the support $\mathbb{S}_\theta$ such that $\mathbb{S}_\theta \cap \mathbb{T} \neq \emptyset$, we need to establish a correspondence between $(x,h) \in \mathbb{T}\times(0,\infty)$ and $\theta \in \Theta$; see Kokonendji & Senga Kiessé (2011). In what follows, we will call $K \equiv K_\theta$ the *type of kernel* to make a difference from the classical notion of continuous symmetric (e.g. Gaussian) kernel. In this context, the choice of the associated kernel becomes important as well as that of the bandwidth. Moreover, we distinguish the associated kernels said sometimes of "second order" of those said of "first order" which verify the two first conditions (5.1) and (5.2).

5.2.1 Discrete associated kernels

Among discrete associated kernels found in literature, we here use the best in sense of Definition 5.2.1. Negative binomial and Poisson kernels are respectively overdispersed (i.e. $\mathrm{Var}(\mathcal{Z}_{x,h}) > \mathbb{E}(\mathcal{Z}_{x,h})$) and equisdispersed (i.e. $\mathrm{Var}(\mathcal{Z}_{x,h}) = \mathbb{E}(\mathcal{Z}_{x,h})$) and thus are not recommended ; see Kokonendji & Senga Kiessé (2011) for further details. The first associated kernel listed below, namely binomial kernel, is the best of the first order or *standard* kernels which satisfies

$$\lim_{h\to 0} \mathrm{Var}(\mathcal{Z}_{x,h}) \in \mathcal{V}(0), \tag{5.4}$$

where $\mathcal{V}(0)$ is a neighborhood of 0 which does not depend on x. The two others discrete associated kernels satisfies all conditions of Definition 5.2.1.

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

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

where $\mathbb{1}_A$ denotes the indicator function of any given event A. Note that $B_{x,h}$ is the p.m.f. of the binomial distribution $\mathcal{B}(x+1;(x+h)/(x+1))$ with its number of trials

$x+1$ and its success probability in each trial $(x+h)/(x+1)$. It is appropriated for count data with small or moderate sample sizes and, also, it satisfies (5.4) rather than (5.3); see Kokonendji & Senga Kiessé (2011) and also Zougab *et al.* (2012) for a bandwidth selection by Bayesian method.

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

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

where $P(a,h) = (2a+1)(a+1) - 2\sum_{k=0}^{a} k^h$ is the normalizing constant. It is symmetric around the target x, satisfying Definition 5.2.1 and suitable for count variables; see Kokonendji & Zocchi (2010) for an asymmetric version. Note that $h \to 0$ gives the Dirac kernel.

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

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

where $h \in (0,1]$ and $x \in \mathbb{T} = \mathbb{S}_c$. In addition, the target x can be considered as the reference point of f to be estimated; and, the smoothing parameter h is such that $1-h$ is the success probability of the reference point. This DiracDU kernel is symmetric around the target, satisfying Definition 5.2.1 and appropriated for categorical set $\mathbb{T}$. See, e.g., Racine & Li (2004) for some uses. Note that $h = 0$ provides the Dirac kernel.

5.2.2 Continuous associated kernels

One can find several continuous associated kernels in literature among the Birnbaum-Saunders of Jin and Kawczak (2003). Here, we present seven associated kernels well adapted for the estimations of density or regression functions on any compact or non-negative support of dataset. All these associated kernels satisfies Definition 5.2.1.

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

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

where $B(r,s) = \int_0^1 t^{r-1}(1-t)^{s-1}dt$ is the usual beta function with $r > 0$, $s > 0$; see Libengué (2013). For $a = 0$ and $b = 1$, it corresponds to the beta kernel Chen (1999)

which is the p.d.f. of the beta distribution with shape parameters $1 + x/h$ and $(1 - x)/h$. The extended beta kernel is appropriated for any compact support of observations.

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

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

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

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

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

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

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

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

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

Remarque 5.2.3 *The three continuous associated kernels inverse gamma, inverse Gaussian and Gaussian are not adapted for density estimation on supports* $[0, \infty)$ *and thus are not included in the* Ake *package ; see Part (b) of Figure 5.1.*

Indeed :

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

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

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

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

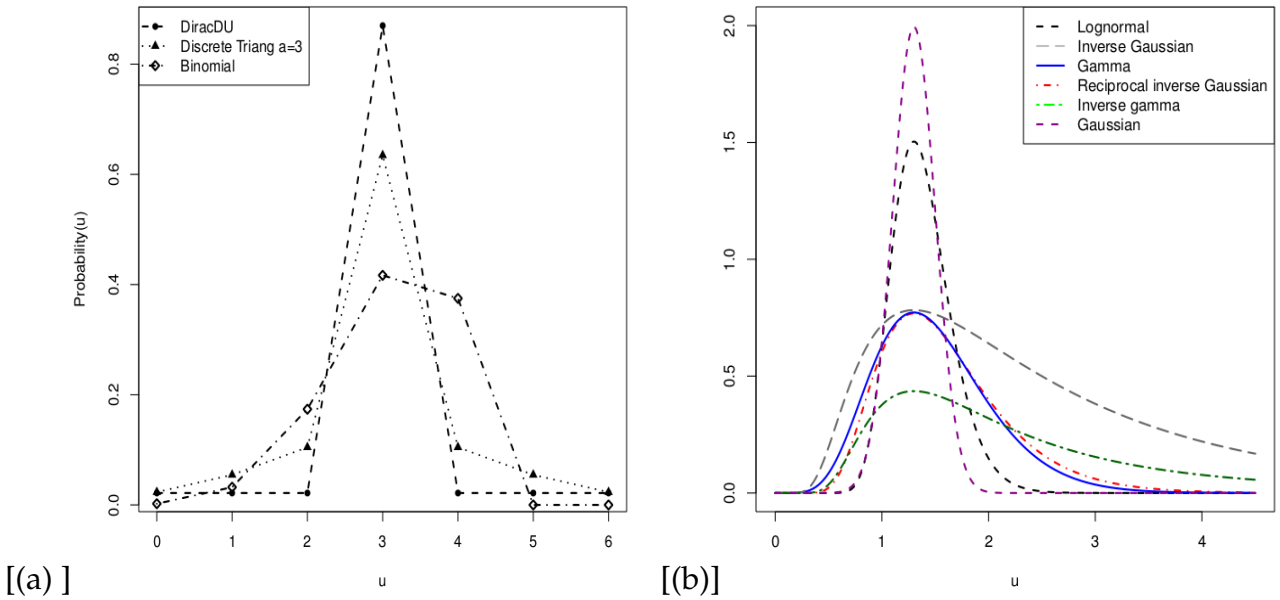
$$IG_{x,h}(u) = \frac{1}{\sqrt{2\pi h u}} \exp\left\{-\frac{(1 - 3xh)^{1/2}}{2h}\left(\frac{u}{(1 - 3xh)^{1/2}} - 2 + \frac{(1 - 3xh)^{1/2}}{u}\right)\right\} \mathbb{1}_{\mathbb{S}_{x,h}}(u)$$

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

– From the well known Gaussian kernel $K^G(u) = (h\sqrt{2\pi})^{-1}\exp(u^2)\mathbb{1}_{\mathbb{R}}(u)$, we define its associated version (`Gaussian`) on $\mathbb{S}_{x,h} = \mathbb{R}$ with $x \in \mathbb{T} := \mathbb{R}$ and $h > 0$:

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

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



[(a)] [(b)]

FIGURE 5.1 – Shapes of univariate (discrete and continuous) associated kernels : (a) DiracDU, discrete triangular $a = 3$ and binomial with same target $x = 4$ and bandwidth $h = 0.13$; (b) lognormal, inverse Gaussian, gamma, reciprocal inverse Gaussian, inverse gamma and Gaussian with same target $x = 1.3$ and $h = 0.2$.

Figure 5.1 shows some forms of the above-mentioned univariate associated kernels. The plots highlight the importance given to the target point and around it in discrete (a) and continuous (b) cases. Furthermore, for a fixed bandwidth h, the Gaussian keeps its same shape along the support ; however, they change according to the target for the others non-classical associated kernels. This explains the inappropriateness of the Gaussian kernel for density or regression estimation in any bounded interval and of the DiracDU kernel for count regression estimation ; see Part (b) of Figure 5.1. From Part (v) of Remark 5.2.2, the inverse gamma and inverse Gaussian are the worst since they do not well concentrate on the target x ; see Remark 5.2.3. These previous associated kernels can be applied to various functionals.

We have implemented in `R` the method `kern.fun` for both discrete and continuous associated kernels. Seven possibilities are allowed for the kernel function. We enumerate

the arguments and results of the default `kern.fun.default` function in Table 5.1. The `kern.fun` is used as follows for the binomial kernel :

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

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

TABLE 5.1 – Summary of arguments and results of `kern.fun`.

5.3 Density or probability mass function estimations

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

$$\widehat{f_n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{T}. \tag{5.5}$$

Here, we point out some pointwise properties of the estimator (5.5) in both discrete and continuous cases.

Proposition 5.3.1 *(Kokonendji & Senga Kiessé 2011, Libengué 2013) Let $X_1, X_2, \ldots, X_n$ be an n random sample i.i.d. from the unknown p.m.f. (resp. p.d.f.) f on $\mathbb{T}$. Let $\widehat{f_n} = \widehat{f_{n,h,K}}$ be an estimator* (5.5) *of f with an associated kernel. Then, for all $x \in \mathbb{T}$ and $h > 0$, we have*

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

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

$$\int_{x\in\mathbb{T}} \widehat{f_n}(x)\nu(dx) = C_n, \tag{5.6}$$

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

It is easy to see that $C_n = 1$ for the estimators (5.5) with DiracDU kernel or any classical (symmetric) associated kernel. Indeed, for the DiracDU kernel estimation we have

$$\begin{aligned}\sum_{x=0}^{c-1} \widehat{f_n}(x) &= \sum_{x=0}^{c-1} \left\{(1-h)\,\mathbb{1}_{\{x\}}(X_1) + \frac{h}{c-1}\mathbb{1}_{\mathbb{S}_c\setminus\{x\}}(X_1)\right\} \\ &= (1-h) + \frac{h}{c-1}(c-1) \\ &= 1.\end{aligned}$$

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

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

[1] 0.9888231
```

Without loss of generality, we study $x \mapsto \widehat{f_n}(x)$ up to normalizing constant which is used at the end of the density estimation process. Notice that that non-classical associated kernel estimators $\widehat{f_n}$ are improper density estimates or as kind of "balloon estimators" ; see Sain (2002). There are two ways to normalize these estimators (5.5). The first method is the *global* normalization using C_n of (5.6) :

$$\widetilde{f_n}(x) = \frac{\widehat{f_n}(x)}{\int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \widehat{f_n}(x)\nu(dx)}, \quad x \in \mathbb{T}. \tag{5.7}$$

Another alternative is to use an *adaptive* normalization of (5.5) according to each target x :

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

but this approach, with similar results than (6.4), is not used here. The representations are done with the global normalization (6.4). In the package, we also compute the normalizing constant (5.6) for any data set.

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

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

is the criteria used to measure numerically the discrete smoothness of $\widetilde{f_n}$ from (6.4) with the empirical or naive p.m.f. f_0 such that $\sum_{x\in\mathbb{N}} f_0(x) = 1$; see, e.g., Kokonendji & Senga Kiessé (2011) and also Wansouwé *et al.* (2015a). Concerning the continuous variables, the histogram gives a graphical measure of comparison with $\widetilde{f_n}$; see, for example, Figure 5.2.

5.3.1 Some theoretical aspects of the normalizing constant

In this section, we present some theoretical aspects of the normalizing constant C_n of (5.6) and two examples in continuous and discrete cases. We first recall the following result on pointwise properties of the estimator (5.5).

Lemme 5.3.2 *Kokonendji & Senga Kiéss´ 2011, Libengué 2013) Let $x \in \mathbb{T}$ be a target and $h \equiv h_n$ a bandwidth. Assuming f in the class $\mathscr{C}^2(\mathbb{T})$ in continuous case, then*

$$\text{Bias}\left\{\widehat{f_n}(x)\right\} = A(x,h)f'(x) + \frac{1}{2}\left\{A^2(x,h) + B(x,h)\right\}f''(x) + o(h^2). \tag{5.8}$$

Similar expressions (5.8) *holds in discrete case, except that f' and f'' are finite differences of first and second order respectively.*

Furthermore, for continuous case, if f is bounded on $\mathbb{T}$ then there exists $r_2 = r_2(K_{x,h}) > 0$ the largest real number such that $\left\|K_{x,h}\right\|_2^2 := \int_{\mathbb{S}_{x,h}} K_{x,h}^2(u)du \leq c_2(x)h^{-r_2}$, $0 \leq c_2(x) \leq \infty$ and

$$\text{Var}\left\{\widehat{f_n}(x)\right\} = \frac{1}{n}f(x)\left\|K_{x,h}\right\|_2^2 + o\left(\frac{1}{nh^{r_2}}\right). \tag{5.9}$$

For discrete situations, the result (5.9) *becomes*

$$\text{Var}\left\{\widehat{f_n}(x)\right\} = \frac{1}{n}f(x)[\{(\mathcal{Z}_{x,h} = x)\}^2 - f(x)],$$

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

It is noticeable that the bias (5.8) is bigger than the one with symmetric kernels and thus can be reduced ; see, e.g., Zhang (2010), Zhang & Karunamuni (2010) and Libengué (2013).

Proposition 5.3.3 *Following notations in Lemma 5.3.2, the mean and variance of C_n of* (5.6) *are respectively :*

$$\mathbb{E}(C_n) \simeq 1 + \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left\{A(x,h)f'(x) + \frac{1}{2}\left[A^2(x,h) + B(x,h)\right]f''(x)\right\}\nu(dx), \tag{5.10}$$

$$\mathrm{Var}\,(C_n) \simeq \begin{cases} \dfrac{1}{n}\displaystyle\int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left(f(x)\left\|K_{x,h}\right\|_2^2\right)dx & \text{if } \mathbb{T} \text{ is continuous} \\ \dfrac{1}{n}\displaystyle\sum_{x\in\mathbb{T}} \left(f(x)[\{(\mathcal{Z}_{x,h}=x)\}^2 - f(x)]\right) & \text{if } \mathbb{T} \text{ is discrete,} \end{cases} \tag{5.11}$$

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

Proof. From Lemma 5.3.2 and the Fubini theorem, we successively show (6.6) as follows :

$$\begin{aligned}
\mathbb{E}\,(C_n) &= \mathbb{E}\left(\int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \widehat{f_n}(x)\nu(dx)\right) = \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \mathbb{E}\left(\widehat{f_n}(x)\right)\nu(dx) \\
&= \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left(\mathrm{Bias}\left\{\widehat{f_n}(x)\right\} + f(x)\right)\nu(dx) \\
&\simeq \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left\{A(x,h)f'(x) + \frac{1}{2}\left[A^2(x,h) + B(x,h)\right]f''(x) + f(x)\right\}\nu(dx) \\
&\simeq \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} f(x)\nu(dx) + \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left\{A(x,h)f'(x) + \frac{1}{2}\left[A^2(x,h) + B(x,h)\right]f''(x)\right\}\nu(dx) \\
&\simeq 1 + \int_{inf(\mathbb{T})}^{sup(\mathbb{T})} \left\{A(x,h)f'(x) + \frac{1}{2}\left[A^2(x,h) + B(x,h)\right]f''(x)\right\}\nu(dx).
\end{aligned}$$

The variance (5.11) is trivial from Lemma 5.3.2. □

Example 5.3.4 *Let f be an exponential density with parameter $\gamma > 0$. Thus, one has :*

$$f(x) = \gamma\exp(-\gamma x), \quad f'(x) = -\gamma^2\exp(-\gamma x) \quad \text{and} \quad f''(x) = \gamma^3\exp(-\gamma x).$$

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

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

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

Example 5.3.5 *Let f be a Poisson p.m.f. with parameter λ and thus $f(x) = \lambda^x\exp(-\lambda)/x!$. The finite differences $f^{(k)}(x)$ of order $k \in \{1,2,\ldots\}$ at $x \in \mathbb{N}$ are given by the recursive relation :*

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

and

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

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

$$\begin{aligned}
\mathbb{E}(C_n) &\simeq 1 + \frac{h}{4}\Bigg(f(3) + 3f(2) - 3f(1) - 2f(0) \\
&\quad + \sum_{x=2}^{\infty} 2\{f(x+1) - f(x-1)\} + \frac{(x+3)\{f(x+2) - 2f(x) + f(x-2)\}}{x+1} \Bigg) \\
&\quad + \Bigg(\frac{3f(3) + 8f(2) - 17f(1) + 3f(0)}{2} \\
&\quad + \sum_{x=2}^{\infty} \frac{x\{f(x+1) - f(x-1)\}}{2} + \frac{(x^3 + x^2 + 1)\{f(x+2) - 2f(x) + f(x-2)\}}{x+1} \Bigg) \\
&\simeq 1 + \frac{h\exp(-\lambda)}{4}\Bigg[\frac{\lambda^3}{3!} + 3\frac{\lambda^2}{2!} - 3\lambda - 2 \\
&\quad + \sum_{x=2}^{\infty} \left\{ \frac{2\lambda^{x+1}}{(x+1)!} - \frac{2\lambda^{x-1}}{(x-1)!} + \frac{x+3}{x+1}\left(\frac{\lambda^{x+2}}{(x+2)!} - 2\frac{\lambda^x}{x!} + \frac{\lambda^{x-2}}{(x-2)!} \right) \right\} \Bigg] \\
&\quad + \Bigg[\frac{\lambda^3}{4} + 2\lambda^2 - \frac{17\lambda}{2} + 3 \\
&\quad + \sum_{x=2}^{\infty} \left\{ \frac{x}{2}\left(\frac{\lambda^{x+1}}{(x+1)!} - \frac{\lambda^{x-1}}{(x-1)!} \right) + \frac{x^3 + x^2 + 1}{x+1}\left(\frac{\lambda^{x+2}}{(x+2)!} - 2\frac{\lambda^x}{x!} + \frac{\lambda^{x-2}}{(x-2)!} \right) \right\} \Bigg]
\end{aligned}$$

and

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

5.3.2 Bandwidth selection

Now, we consider the bandwidth selection problems which are generally crucial in nonparametric estimation. Several methods already existing for continuous kernels can be adapted to the discrete case as the classical least-squares cross-validation method ; see, for example, Bowman (1984), Marron (1987) and references therein. Here, we simply propose three procedures for the bandwidth selection : cross-validation, Bayesian local for binomial and adaptive for the gamma kernel. The smoothing parameter selection is done with the non-normalized version $\widehat{f_n}$ of the estimator (5.5) before the global normalization $\widetilde{f_n}$ of (6.4).

Cross-validation for any associated kernel

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

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

where $\widehat{f_{n,-i}}(X_i) = (n-1)^{-1}\sum_{j\neq i} K_{X_i,h}(X_j)$ is being computed as $\widehat{f_n}(X_i)$ by excluding the observation X_i and ν is the Lebesgue or count measure. This method is applied to all estimators (5.5) with associated kernels cited in this paper, independently on the support $\mathbb{T}$ of f to be estimated. Table 5.2 gives the arguments and results of the cross-

Arguments	Description
`Vec`	The data sample
`seqbws`	The sequence of bandwidths where to compute the cross-validation function.
`ker`	The associated kernel.
`a0,a1`	The bounds of the support of extended beta kernel.
`a`	The arm of the discrete triangular kernel.
`c`	The number of categories in DiracDU kernel.
Results	Description
`hcv`	The optimal bandwidth obtained by cross-validation.
`seqh`	The sequence of bandwidths used to compute `hcv`.
`CV`	The values of the cross-validation function.

TABLE 5.2 – Summary of arguments and results of `hcvc.fun`.

validation function `hcvc.fun` defined for continuous data are below. The `hcvd.fun` is the corresponding function for discrete data. The `hcvc.fun` is performed with the Old Faithful geyser data described in Azzalini & Bowman (1990) and Hardle & James (1985). The dataset concerns waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. The following codes and Figure 5.2 give smoothing density estimation with various associated kernels of the waiting time variable.

```
R> data("faithful", package = "datasets")
R> x <- faithful$waiting
R> f1 = dke.fun(x, 0.1, "continuous", ker = "GA")
R> f2 = dke.fun(x, 0.036, "continuous", ker = "LN")
R> f3 = dke.fun(x, 0.098, "continuous", ker = "RIG")
R> f4 = dke.fun(x, 0.01, "continuous", ker = "BE", a0 = 40, a1 = 100)
R> t = seq(min(x), max(x), length.out = 100)
```

```
R> hist(x, probability = TRUE, xlab = "Waiting times (in min.)",
+ ylab = "Frequency", main = "", border = "gray")
R> lines(t, f1$fn, lty = 2, lwd = 2, col = "blue")
R> lines(t, f2$fn, lty = 5, lwd = 2, col = "black")
R> lines(t, f3$fn, lty = 1, lwd = 2, col = "green")
R> lines(t, f4$fn, lty = 4, lwd = 2, col = "grey")
R> lines(density(x, width = 12), lty = 8, lwd = 2, col = "red")
legend("topleft", c("Gamma", "Lognormal",
+  "Reciprocal inverse Gaussian", "Extended beta", "Gaussian"),
+  col = c("blue", "black", "green", "grey", "red"),
+  lwd = 2, lty = c(2, 5, 1, 4, 8), inset = .0)
```

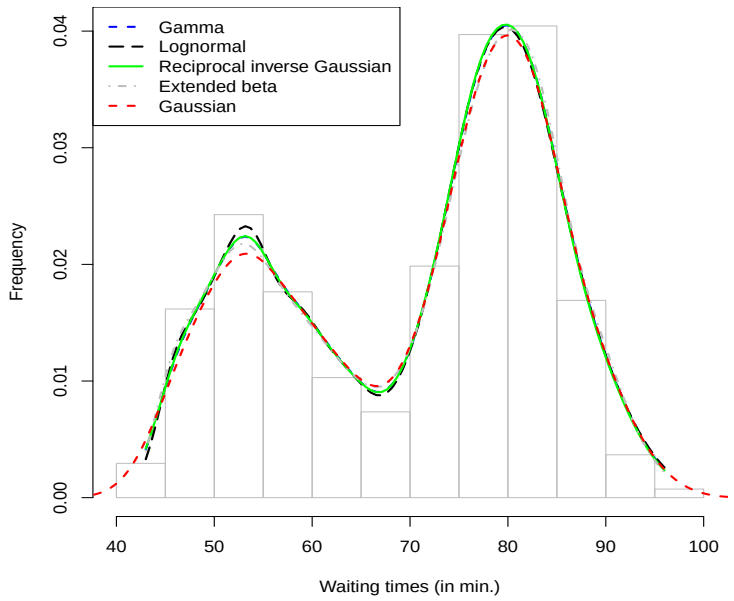


FIGURE 5.2 – Smoothing density estimation of the Old Faithful geyser data (Azzalini & Bowman, 1990) by some continous associated kernels with the support of observations $[43, 96] = \mathbb{T}$.

Bayesian local for binomial kernel

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

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

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

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

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

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

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

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

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

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

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

Bayesian adaptive for gamma kernel

The last choice of h is the Bayesian adaptive for gamma kernel. This approach gives a variable bandwidth h_i for each observation X_i in place of the initial fixed bandwidth h. Following Zougab *et al.* (2014b), we suggest a Bayesian estimation of adaptive bandwidths to select the variable bandwidths $h_i, i = 1, ..., n$. Thus, we treat h_i as a random variable with a prior distribution $\pi(\cdot)$. The estimator (5.5) with gamma kernel of Section 5.2.2 and variable bandwidths is reformulated as

$$\widehat{f_n}(x) = \frac{1}{n}\sum_{i=1}^{n} GA_{x,h_i}(X_i). \tag{5.12}$$

The leave-one-out kernel estimator of $f(X_i)$ deduced from (5.12) is

$$\widehat{f}(X_i \mid \{X_{-i}\}, h_i) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} GA_{X_i,h_i}(X_j), \tag{5.13}$$

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

$$\pi(h_i \mid X_i) = \frac{\widehat{f}(X_i \mid \{X_{-i}\}, h_i)\pi(h_i)}{\int_0^\infty \widehat{f}(X_i \mid \{X_{-i}\}, h_i)\pi(h_i)dh_i}. \tag{5.14}$$

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

$$\widetilde{h}_i = \mathbb{E}(h_i \mid X_i). \tag{5.15}$$

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

$$\Phi_{a,b}(z) = \frac{b^a}{\Gamma(a)} z^{-a-1} \exp(-b/z)\mathbb{1}_{(0,\infty)}(z). \tag{5.16}$$

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

Théorème 5.3.6 *For fixed $i \in \{1, 2, ..., n\}$, consider each observation X_i with its corresponding bandwidth h_i. Using the gamma kernel estimator* (5.12) *and the inverse gamma prior distribution $\mathcal{IGA}(\alpha, \beta)$ given in* (5.16) *with $\alpha > 1/2$ and $\beta > 0$ for each h_i, then :*

*(i) the posterior density (*5.14*) is the following weighted sum of inverse gamma*

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

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

*(ii) the Bayesian estimator $\widetilde{h}_i$ of h_i, given in (*5.15*), is*

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

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

Proof. (i) Let us represent $\pi(h_i \mid X_i)$ of (5.14) as the ratio of $N(h_i \mid X_i) := \widehat{f}(X_i \mid \{X_{-i}\}, h_i)\pi(h_i)$ and $\int_0^\infty N(h_i \mid X_i)dh_i$. From (5.13) and (5.16) the numerator is, first, equal to

$$\begin{aligned} N(h_i \mid X_i) &= \left(\frac{1}{n-1} \sum_{j=1, j\neq i}^{n} GA_{X_i, h_i}(X_j) \right) \left(\frac{\beta^\alpha}{\Gamma(\alpha)} h_i^{-\alpha-1} \exp(-\beta/h_i) \right) \\ &= \frac{[\Gamma(\alpha)]^{-1}}{(n-1)} \sum_{j=1, j\neq i}^{n} \frac{GA_{X_i, h_i}(X_j)}{\beta^{-\alpha} h_i^{\alpha+1}} \exp(-\beta/h_i). \end{aligned} \tag{5.17}$$

Following Chen (2000a), we assume that for all $X_i \in (0,\infty)$ one has $1 + (X_i/h_i) \to \infty$ as $n \to \infty$. Using the Stirling formula $\Gamma(z+1) \simeq \sqrt{2\pi} z^{z+1/2} \exp(-z)$ as $z \to \infty$, the term of sum in (5.17) can be successively calculated as

$$\begin{aligned}
\frac{GA_{X_i,h_i}(X_j)}{\beta^{-\alpha} h_i^{\alpha+1}} \exp(-\beta/h_i) &= \frac{X_j^{(X_i/h_i)} \exp(-X_j/h_i)}{h_i^{1+(X_i/h_i)} \Gamma[1+(X_i/h_i)] \beta^{-\alpha} h_i^{\alpha+1}} \exp(-\beta/h_i) \\
&= \frac{\exp[-(X_j + \beta - X_i \log X_j)/h_i]}{\beta^{-\alpha} h_i^{(X_i/h_i)+\alpha+2} \sqrt{2\pi} \exp(-X_i/h_i) (X_i/h_i)^{(X_i/h_i)+1/2}} \\
&= \frac{\Gamma(\alpha+1/2)}{\beta^{-\alpha} X_i^{1/2} \sqrt{2\pi} B_{ij}^{\alpha+1/2}} \times \frac{B_{ij}^{\alpha+1/2} \exp[-B_{ij}/h_i]}{h_i^{\alpha+3/2} \Gamma(\alpha+1/2)} \\
&= A_{ij}\, \Phi_{\alpha+1/2, B_{ij}}(h_i), \qquad (5.18)
\end{aligned}$$

with $B_{ij} = X_i \log X_i - X_i \log X_j + X_j - X_i + \beta$, $A_{ij} = [X_j^{-1}\Gamma(\alpha+1/2)]/(\beta^{-\alpha} X_i^{-1/2} \sqrt{2\pi} B_{ij}^{\alpha+1/2})$ and $\Phi_{\alpha+1/2,B_{ij}}(h_i)$ is given in (5.16).

Also, for $X_i = 0$, the term of sum (5.17) can be expressed as follows

$$\begin{aligned}
\frac{GA_{0,h_i}(X_j)}{\beta^{-\alpha} h_i^{\alpha+1}} \exp(-\beta/h_i) &= \frac{\exp(-X_j/h_i)}{\beta^{-\alpha} h_i^{\alpha+2}} \exp(-\beta/h_i) \\
&= \frac{\Gamma(\alpha+1)}{\beta^{-\alpha}(X_j+\beta)^{\alpha+1}} \times \frac{(X_j+\beta)^{\alpha+1} \exp[-(X_j+\beta)/h_i]}{h_i^{\alpha+2}\Gamma(\alpha+1)} \\
&= C_j\, \Phi_{\alpha+1, X_j+\beta}(h_i), \qquad (5.19)
\end{aligned}$$

with $C_j = \Gamma(\alpha+1)/[\beta^{-\alpha}(X_j+\beta)^{\alpha+1}]$ and $\Phi_{\alpha+1,X_j+\beta}(h_i)$ is given in (5.16). Combining (5.18) and (5.19), the expression of $N(h_i \mid X_i)$ in (5.17) becomes

$$N(h_i \mid X_i) = \frac{[\Gamma(\alpha)]^{-1}}{(n-1)} \sum_{j=1, j\neq i}^{n} \left\{ A_{ij}\, \Phi_{\alpha+1/2,B_{ij}}(h_i) \mathbb{1}_{(0,\infty)}(X_i) + C_j\, \Phi_{\alpha+1,X_j+\beta}(h_i) \mathbb{1}_{\{0\}}(X_i) \right\}. \qquad (5.20)$$

From (5.20), the denominator is successively computed as follows :

$$\begin{aligned}
\int_0^\infty N(h_i \mid X_i)\, dh_i &= \frac{[\Gamma(\alpha)]^{-1}}{(n-1)} \sum_{j=1, j\neq i}^{n} \left(A_{ij} \int_0^\infty \Phi_{\alpha+1/2,B_{ij}}(h_i) \mathbb{1}_{(0,\infty)}(X_i)\, dh_i \right. \\
&\quad \left. + C_j \int_0^\infty \Phi_{\alpha+1,X_j+\beta}(h_i) \mathbb{1}_{\{0\}}(X_i)\, dh_i \right) \\
&= \frac{[\Gamma(\alpha)]^{-1}}{(n-1)} \sum_{j=1, j\neq i}^{n} \left\{ A_{ij} \mathbb{1}_{(0,\infty)}(X_i) + C_j \mathbb{1}_{\{0\}}(X_i) \right\} \\
&= \frac{[\Gamma(\alpha)]^{-1}}{(n-1)} D_{ij}, \qquad (5.21)
\end{aligned}$$

with $D_{ij} = \sum_{j=1, j\neq i}^{n} \left(A_{ij} \mathbb{1}_{(0,\infty)}(X_i) + C_j \mathbb{1}_{\{0\}}(X_i) \right)$. Finally, the ratio of (5.20) and (5.21) leads to the result of Part (i).

(ii) Let us remember that the mean of inverse gamma distribution $\mathcal{IG}(\alpha, \beta)$ is $\beta/(\alpha-1)$. Thus, the expression of $\pi(h_i \mid X_i)$ in (5.14) is written by

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

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

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

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

5.4 Regression function estimations

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

$$Y = m(x) + \epsilon, \tag{5.22}$$

where m is an unknown regression function from $\mathbb{T} \subseteq \mathbb{R}$ to $\mathbb{R}$ and ϵ the disturbance term with null mean and finite variance. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sequence of i.i.d. random variables on $\mathbb{T} \times \mathbb{R} (\subseteq \mathbb{R}^2)$ with $m(x) = \mathbb{E}(Y|X = x)$ of (5.22). Using (continuous or discrete) associated kernels, the Nadaraya (1969) and Watson (1969) estimator $\widehat{m}_n$ of m is

$$\widehat{m}_n(x; h) = \sum_{i=1}^{n} \frac{Y_i K_{x,h}(X_i)}{\sum_{i=1}^{n} K_{x,h}(X_i)} = \widehat{m}_n(x), \quad \forall x \in \mathbb{T} \subseteq \mathbb{R}, \tag{5.23}$$

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

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

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

and

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

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

In discrete cases, the `reg.fun` function for (5.23) is used with the binomial kernel on milk data as follows. This dataset is about average daily fat (kg/day) yields from milk of a single cow for each of the first 35 weeks.

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

Bandwidth h:0.1        Coef_det=0.9726

Number of points: 35; Kernel =  Binomial

      data            y
 Min.   : 1.0   Min.   :0.0100
 1st Qu.: 9.5   1st Qu.:0.2750
 Median :18.0   Median :0.3600
 Mean   :18.0   Mean   :0.3986
 3rd Qu.:26.5   3rd Qu.:0.6150
 Max.   :35.0   Max.   :0.7200
  eval.points        m_n
 Min.   : 1.0   Min.   :0.01542
 1st Qu.: 9.5   1st Qu.:0.27681
 Median :18.0   Median :0.35065
 Mean   :18.0   Mean   :0.39777
 3rd Qu.:26.5   3rd Qu.:0.60942
 Max.   :35.0   Max.   :0.70064
```

The above `reg.fun` is also used for continuous cases ; see Figure 5.3 and Table 5.4 for the motorcycle impact data of Silvermann (1985).

Bandwidth selection

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

Cross-validation for any associated kernel

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

$$\text{LSCV}(h) = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \widehat{m}_{-i}(X_i)\right\}^2, \tag{5.24}$$

where $\widehat{m}_{-i}(X_i)$ is computed as $\widehat{m}_n$ of (5.23) excluding X_i ; see, e.g., Kokonendji *et al.* (2009) and Wansouwé *et al.* (2015b). The `hcvreg.fun` function to compute this optimal bandwidth is described in Table 5.3.

Arguments	Description
Vec	The explanatory data sample.
y	The response variable.
ker	The associated kernel.
h	The sequence of bandwidths where to compute the optimal bandwidth.
a0, a1	The bounds of the support of extended beta kernel.
a	The arm of the discrete triangular kernel.
c	The number of categories in DiracDU kernel.
Results	**Description**
kernel	The associated kernel.
hcv	The optimal bandwidth obtained by cross-validation.
CV	The values of the cross-validation function.
seqbws	The sequence of bandwidths used to compute hcv.

TABLE 5.3 – Summary of arguments and results of hcvreg.fun.

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

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

```
[1] 1.141073
```

When we consider continuous associated kernel, one needs to set the type of data parameter to "continuous" in the hcvreg.fun function. Thus, the hcvreg.fun and reg.fun functions are used with gamma, lognormal, reciprocal inverse Gaussian and Gaussian kernel on the motor cycle impact data described in Silvermann (1985). The observations consist of accelerometer reading taken through time in an experimentation on the efficiency of crash helmets. The results in Table 5.4 agree with the shapes of continuous associated kernels of Part (b) of Figure 5.1 ; see also Figure 5.3. In fact, since the lognormal kernel is well concentrated around the target x, it gives the best R^2 which is 75.9%. The gamma and the reciprocal inverse Gaussian give similar R^2 in the order 73%. Although the Gaussian kernel is well concentrated on the target, it gives the lower result of $R^2 = 70.90\%$. This is mainly due to the symmetry of the kernel which cannot change its shapes according the target.

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

TABLE 5.4 – Some expected values of R^2 of of nonparametric regressions of the motor cycle impact data Silvermann (1985) by some continuous associated kernels.

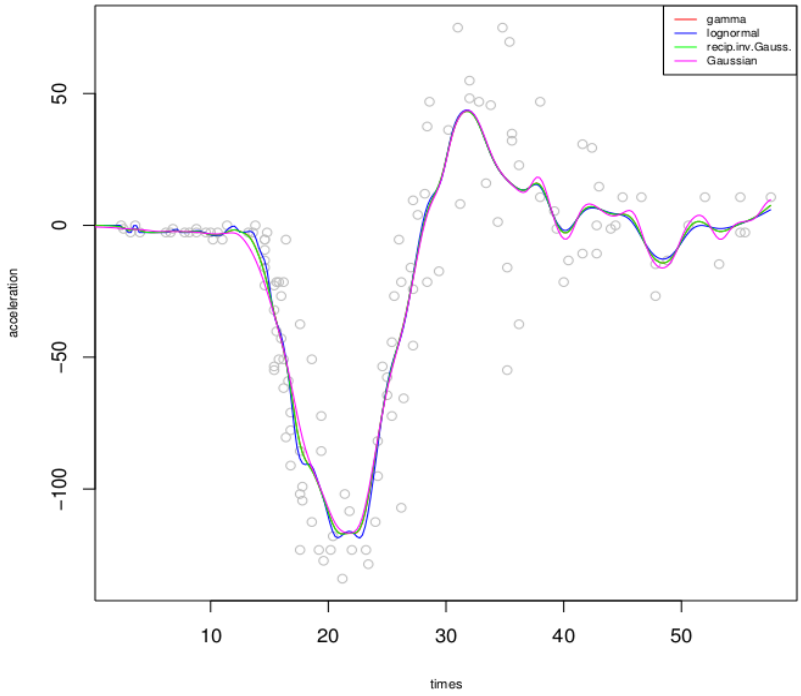


FIGURE 5.3 – Nonparametric regressions of the motors cycle impact data (Silvermann, 1985) by some continous associated kernels.

Bayesian global for binomial kernel

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

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

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

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

where N_0 is the burn-in period and N the number of iterations ; see Zougab *et al.* (2012) for further details. It will be implemented in the future version of the `Ake` package.

5.5 Summary and final remarks

The Ake package offers to the users of R a first associated kernel based method for p.d.f., p.m.f. and regression functions that are capable of handling all categorical, count and real positive datasets. Figure 5.1 shows the importance of the associated kernel choice as well as the bandwidth selection. In fact, symmetric (e.g. Gausssian) kernel estimator (resp. empirical estimator) do not suit for bounded or positive continuous dataset (resp. discrete small sample). We then need an appropriate associated kernel. The binomial is suitable for small size count data while the discrete triangular or the naive kernel is more indicated for large sample sizes. In continuous cases, the lognormal and gamma kernels give the best estimation for positive data while the extended beta suit for any compact support.

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

Chapitre 6

Compléments

Dans cette partie, quelques aspects de travaux réalisés sont discutés. Il s'agit d'abord d'aspects théoriques de la constante de normalisation des estimateurs de densité par noyaux associés multivariés. Enfin, on étend la technique de sélection de fenêtre par la méthode bayésienne adaptative pour l'estimateur à noyau gamma univarié au cas multiple.

6.1 Constantes de normalisation pour fonctions de densités

Soit $\mathbf{X}_1, \cdots, \mathbf{X}_n$ une suite de vecteurs aléatoires i.i.d. de densité inconnue f sur $\mathbb{T}_d$ ($\subseteq \mathbb{R}^d$) avec $d \geq 1$. On rappelle que l'estimateur à noyaux associés multivariés est donné par

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i), \quad \forall \mathbf{x} \in \mathbb{T}_d \subseteq \mathbb{R}^d, \tag{6.1}$$

où $\mathbf{H}$ est une matrice des fenêtres d'ordre $d \times d$ (i.e. symétrique et définie positive telle que $\mathbf{H} \equiv \mathbf{H}_n \to \mathbf{0}$ quand $n \to +\infty$) et $K_{\mathbf{x},\mathbf{H}}(\cdot)$ est appelé le noyau associé multivarié, paramétré par $\mathbf{x}$ et $\mathbf{H}$. Ainsi, la masse totale est

$$\Lambda_n = \int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x})\boldsymbol{\nu}(d\mathbf{x}), \tag{6.2}$$

où $\boldsymbol{\nu}$ la mesure de référence sur $\mathbb{T}_d$ et que Λ_n est positive et finie (i.e.

$$\int_{\mathbb{T}_d} K_{\mathbf{x},\mathbf{H}}(\mathbf{u})\boldsymbol{\nu}(d\mathbf{x}) < +\infty \text{ pour tout } \mathbf{u} \in \mathbb{S}_{\mathbf{x},\mathbf{H}} \cap \mathbb{T}_d). \tag{6.3}$$

Comme en univarié, rien ne garantit que le noyau associé dans (6.3) est une densité de probabilité par rapport à $\mathbf{x}$. En ce qui concerne les cas multivarié continu et univarié (continu ou discret), on a montré par simulations que Λ_n est autour de 1. Théoriquement, cela peut se montrer par les inégalités de concentration comme l'a fait Kokonendji & Varron (2015) en univarié discret. Le lecteur pourra se référer à Boucheron *et al.* (2013)

pour une théorie générale de ces outils. Toujours en univarié, une autre approche consiste à utiliser les notions de normalisation globale et adaptative telle qu'énoncé par Wansouwé *et al.* (2015d) au Chapitre 5. On étend ici ces notions aux cas multivariés. Ainsi, les normalisations *globale* et *adaptative* deviennent respectivement

$$\widetilde{f_n}(\mathbf{x}) = \frac{\widehat{f_n}(\mathbf{x})}{\int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x})\boldsymbol{\nu}(d\mathbf{x})}, \quad \forall \mathbf{x} \in \mathbb{T}_d$$

et

$$\widetilde{\widetilde{f_n}}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{K_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i)}{\int_{\mathbb{T}_d} K_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i)\boldsymbol{\nu}(d\mathbf{x})}, \quad \forall \mathbf{x} \in \mathbb{T}_d.$$

La masse totale Λ_n de (6.2) est plus ou moins égale à 1 comme l'atteste la Proposition 6.1.2 ci-dessous pour le cas des fonctions de densité. On rappelle d'abord les biais et variance de l'estimateur (6.1) par la proposition suivante.

Proposition 6.1.1 *(Kokonendji & Somé, 2015) Soit un vecteur cible $\mathbf{x} \in \mathbb{T}_d$ et une matrice de lissage $\mathbf{H} \equiv \mathbf{H}_n \to \mathbf{0}_d$ quand $n \to +\infty$. Si f est de classe $\mathscr{C}^2(\mathbb{T}_d)$ et estimée par $\widehat{f_n}$ de* (6.1)*, alors*

$$\begin{aligned}\text{Biais}\left\{\widehat{f_n}(\mathbf{x})\right\} &= \mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H})\nabla f(\mathbf{x}) + \frac{1}{2}\,\text{trace}\left[\left\{\mathbf{a}_\theta(\mathbf{x},\mathbf{H})\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H}) + \mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right\}\nabla^2 f(\mathbf{x})\right] \\ &\quad + \text{o}\left\{\text{trace}\left(\mathbf{H}^2\right)\right\}. \end{aligned} \tag{6.4}$$

De plus, si f est bornée sur $\mathbb{T}_d$ alors il existe $r_2 = r_2(K_\theta)$ le plus grand réel tel que $\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 \lesssim c_2(\mathbf{x})(\det\mathbf{H})^{-r_2}$, $0 \le c_2(\mathbf{x}) \le +\infty$ et

$$\text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} = \frac{1}{n}\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x}) + \text{o}\left(n^{-1}\,(\det\mathbf{H})^{-r_2}\right), \tag{6.5}$$

avec $\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 := \int_{\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}} K^2_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{u})d\mathbf{u}$ et où "$\lesssim$" désigne "$\le$ et ensuite une approximation quand $n \to +\infty$".

Proposition 6.1.2 *Partant de Λ_n de* (6.2)*, sa moyenne et sa variance sont respectivement données par :*

$$\begin{aligned}\mathbb{E}(\Lambda_n) &= 1 + \int_{\mathbb{T}_d} \text{Biais}\left\{\widehat{f_n}(\mathbf{x})\right\} d\mathbf{x} \\ &\simeq 1 + \int_{\mathbb{T}_d} \left\{\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H})\nabla f(\mathbf{x}) + \frac{1}{2}\,\text{trace}\left[\left\{\mathbf{a}_\theta(\mathbf{x},\mathbf{H})\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H}) + \mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right\}\nabla^2 f(\mathbf{x})\right]\right\} d\mathbf{x}, \end{aligned} \tag{6.6}$$

$$\begin{aligned}\text{Var}(\Lambda_n) &= \int_{\mathbb{T}_d} \text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} d\mathbf{x} \\ &\simeq \frac{1}{n}\int_{\mathbb{T}_d} \left(\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x})\right) d\mathbf{x} \end{aligned} \tag{6.7}$$

où "$\simeq$" désigne une approximation.

Démonstration De la Proposition 6.1.1 et du thèorème de Fubini, on obtient successivement (6.6) comme suit :

$$\begin{aligned}
\mathbb{E}\left(\Lambda_n\right) &= \mathbb{E}\left[\int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x})d\mathbf{x}\right] = \int_{\mathbb{T}_d} \mathbb{E}\left(\widehat{f_n}(\mathbf{x})\right) d\mathbf{x} \\
&= \int_{\mathbb{T}_d} \left(\text{Biais}\left\{\widehat{f_n}(\mathbf{x})\right\} + f(\mathbf{x})\right) d\mathbf{x} = 1 + \int_{\mathbb{T}_d} \text{Biais}\left\{\widehat{f_n}(\mathbf{x})\right\} d\mathbf{x} \\
&\simeq 1 + \int_{\mathbb{T}_d} \left\{\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H})\nabla f(\mathbf{x}) + \frac{1}{2}\,\text{trace}\left[\left\{\mathbf{a}_\theta(\mathbf{x},\mathbf{H})\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H}) + \mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right\}\nabla^2 f\left(\mathbf{x}\right)\right]\right\} d\mathbf{x} \\
&\simeq 1 + \int_{\mathbb{T}_d} f(\mathbf{x})d\mathbf{x} + \int_{\mathbb{T}_d} \left\{\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H})\nabla f(\mathbf{x})\right. \\
&\quad \left. + \frac{1}{2}\,\text{trace}\left[\left\{\mathbf{a}_\theta(\mathbf{x},\mathbf{H})\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H}) + \mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right\}\nabla^2 f\left(\mathbf{x}\right)\right]\right\} d\mathbf{x} \\
&\simeq 1 + \int_{\mathbb{T}_d} \left\{\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H})\nabla f(\mathbf{x}) + \frac{1}{2}\,\text{trace}\left[\left\{\mathbf{a}_\theta(\mathbf{x},\mathbf{H})\mathbf{a}_\theta^\top(\mathbf{x},\mathbf{H}) + \mathbf{B}_\theta(\mathbf{x},\mathbf{H})\right\}\nabla^2 f\left(\mathbf{x}\right)\right]\right\} d\mathbf{x}.
\end{aligned}$$

La variance est aussi obtenue de manière similaire :

$$\begin{aligned}
\text{Var}\left(\Lambda_n\right) &= \text{Var}\left(\int_{\mathbb{T}_d} \widehat{f_n}(\mathbf{x})d\mathbf{x}\right) = \int_{\mathbb{T}_d} \text{Var}\left\{\widehat{f_n}(\mathbf{x})\right\} d\mathbf{x} \\
&\simeq \frac{1}{n}\int_{\mathbb{T}_d} \left(\left\|K_{\theta(\mathbf{x},\mathbf{H})}\right\|_2^2 f(\mathbf{x})\right) d\mathbf{x}.\blacksquare
\end{aligned}$$

Des illustrations de la Proposition 6.1.2 peuvent être obtenues en utilisant les versions multiples de celles univariés données en Section 5.3, voire Wansouwé *et al.* (2015d). On peut montrer que la constante de normalisation Λ_n du noyau DiracDU multiple est égale à 1. Il en est de même pour le noyau multiple de Wang & Van Rizin (1981) dans la Table 1.3 ; en effet, on a successivement

$$\begin{aligned}
\sum_{\mathbf{x}\in\mathbb{Z}^d} \widehat{f_n}(\mathbf{x}) &= \sum_{\mathbf{x}\in\mathbb{Z}^d} \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\left\{\left(1-h_j\right)\mathbb{1}_{\{X_{ij}=x_j\}} + \frac{1}{2}\left(1-h_j\right)h_j^{|X_{ij}-x_j|}\mathbb{1}_{\{|X_{ij}-x_j|\geqslant 1\}}\right\} \\
&= \sum_{\mathbf{x}\in\mathbb{Z}^d}\prod_{j=1}^{d}\left\{\left(1-h_j\right)\mathbb{1}_{\{X_{1j}=x_j\}} + \frac{1}{2}\left(1-h_j\right)h_j^{|X_{1j}-x_j|}\mathbb{1}_{\{|X_{1j}-x_j|\geqslant 1\}}\right\} \\
&= \prod_{j=1}^{d}\left\{\left(1-h_j\right) + \frac{1}{2}\left(1-h_j\right)\sum_{y_j\in\mathbb{Z}\setminus\{0\}} h_j^{|y_j|}\right\} \\
&= \prod_{j=1}^{d}\left\{\left(1-h_j\right) + \frac{1}{2}\left(1-h_j\right)\frac{2h_j}{(1-h_j)}\right\} \\
&= 1.
\end{aligned}$$

6.2 Choix bayésien adaptatif pour noyau gamma multiple

Cette partie étend la sélection de fenêtre par la méthode bayésienne adaptative pour un estimateur à noyau gamma univarié au cas multivarié (multiple). Comme dans le cas univarié de Wansouwé *et al*. (2015d), cette approche consiste à sélectionner un vecteur de fenêtres $\mathbf{h}_i$ correspondant à l'observation $\mathbf{X}_i$ en lieu et place du vecteur $\mathbf{h} = (h_1, \ldots, h_d)^\top$ initial. Soit $\mathbf{X}_1, \ldots, \mathbf{X}_n$ une suite de vecteurs aléatoires i.i.d. de densité inconnue f sur $[0, +\infty[^d := \mathbb{T}_d$ avec $d \geq 1$, l'estimateur à noyau gamma multiple avec vecteurs de fenêtres variables $\mathbf{h}_i = (h_{i1}, ..., h_{id})^\top$ est donné par

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{\ell=1}^{d} GA_{x_\ell, h_{i\ell}}(X_{i\ell}), \quad \forall \mathbf{x} = (x_1, \ldots, x_d)^\top \in \mathbb{T}_d := [0, +\infty[^d, \tag{6.8}$$

où $GA_{x,h}(u) = \left[u^{x/h}\exp(-u/h)\right]\left[\Gamma(1 + x/h)\, h^{1+x/h}\right]^{-1} \mathbb{1}_{[0,+\infty[}(u)$ (voir Table 1.2).

De (6.8), on déduit l'estimateur par validation croisée de $f(\mathbf{X}_i)$ par

$$\widehat{f_{-i}}(\mathbf{X}_i \mid \mathbf{h}_i) = \widehat{f}(\mathbf{X}_i \mid \{\mathbf{X}_{-i}\}, \mathbf{h}_i) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n}\prod_{\ell=1}^{d} GA_{X_{i\ell}, h_{i\ell}}(X_{j\ell}), \tag{6.9}$$

où $\{\mathbf{X}_{-i}\}$ est l'ensemble de toutes les observations sans l'observation $\mathbf{X}_i$. C'est ce $\widehat{f_{-i}}(\mathbf{X}_i \mid \mathbf{h}_i) = \widehat{f}(\mathbf{X}_i \mid \{\mathbf{X}_{-i}\}, \mathbf{h}_i)$ qui apporte l'information dont on a besoin pour avoir une loi *a posteriori* de $\mathbf{h}_i$ étant donné une loi *a priori* $\pi(\mathbf{h}_i)$:

$$\pi(\mathbf{h}_i \mid \mathbf{X}_i) = \frac{\widehat{f}(\mathbf{X}_i \mid \{\mathbf{X}_{-i}\}, \mathbf{h}_i)\pi(\mathbf{h}_i)}{\int_\chi \widehat{f}(\mathbf{X}_i \mid \{\mathbf{X}_{-i}\}, \mathbf{h}_i)\pi(\mathbf{h}_i)d\mathbf{h}_i}, \tag{6.10}$$

où χ est l'espace de vecteurs positifs. On obtient l'estimateur bayésien $\widetilde{\mathbf{h}_i}$ de $\mathbf{h}_i$ en utilisant la fonction de perte quadratique (en anglais "quadratic loss function")

$$\widetilde{\mathbf{h}_i} = \mathbb{E}(\mathbf{h}_i \mid \mathbf{X}_i) = (\mathbb{E}(h_{i1} \mid \mathbf{X}_i), ..., \mathbb{E}(h_{id} \mid \mathbf{X}_i))^\top. \tag{6.11}$$

Dans la suite, on suppose que chaque composante $h_{i\ell} = h_{i\ell}(n), \ell = 1, ..., d$ de $\mathbf{h}_i$ a comme loi *a priori* la loi gamma inverse $\mathcal{IG}(\alpha, \beta_\ell)$ avec les mêmes paramètres de forme $\alpha > 0$ et éventuellement différents paramètres d'échelle $\beta_\ell > 0$. On rappelle que la densité de $\mathcal{IG}(a, b)$ avec $a, b > 0$ est définie par

$$\Phi_{a,b}(z) = \frac{b^a}{\Gamma(a)} z^{-a-1}\exp(-b/z)\mathbb{1}_{]0,+\infty[}(z). \tag{6.12}$$

De ces considérations, on obtient une forme explicite de la loi *a posteriori* et de l'estimateur bayésien de $\mathbf{h}_i$ par le résultat suivant.

Théorème 6.2.1 *Pour $i \in \{1, 2, ..., n\}$ fixé, on considère chaque observation $\boldsymbol{X}_i = (X_{i1}, ..., X_{id})^\top$ avec son vecteur de fenêtres univariés $\boldsymbol{h}_i = (h_{i1}, ..., h_{id})^\top$ et on définit le sous ensemble $\mathbb{I}_i =$*

$\{k \in \{1,...,d\}\,;X_{ik}=0\}$ *et son complémentaire* $\mathbb{I}_i^c=\{\ell\in\{1,...,d\}\,;X_{i\ell}\in]0,+\infty[\}$. *Avec l'estimateur à noyau gamma multiple* (6.8) *et la distribution a priori gamma inverse* $\mathcal{IG}(\alpha,\beta_\ell)$ *donné en* (6.12) *pour chaque composante* $h_{i\ell}$ *de* $\boldsymbol{h}_i$, *alors :*

(i) la loi a posteriori (6.10) *est la somme pondérée des densités gamma inverse*

$$\pi(\boldsymbol{h}_i\mid\boldsymbol{X}_i)=\frac{1}{D_{ij}}\sum_{j=1,j\neq i}^{n}\left(\prod_{k\in\mathbb{I}_i}C_{jk}\,\Phi_{\alpha+1,X_{jk}+\beta_k}(h_{ik})\right)\left(\prod_{\ell\in\mathbb{I}_i^c}A_{ij\ell}\,\Phi_{\alpha+1/2,B_{ij\ell}}(h_{i\ell})\right),$$

avec $A_{ij\ell}=[\Gamma(\alpha+1/2)]/(\beta_\ell^{-\alpha}X_{i\ell}^{1/2}\sqrt{2\pi}B_{ij\ell}^{\alpha+1/2})$, $C_{jk}=[\Gamma(\alpha+1)]/[\beta_k^{-\alpha}(X_{jk}+\beta_k)^{\alpha+1}]$, $B_{ij\ell}=X_{i\ell}\log X_{i\ell}-X_{i\ell}\log X_{j\ell}+X_{j\ell}-X_{i\ell}+\beta_\ell$ *et* $D_{ij}=\sum_{j=1,j\neq i}^{n}\left(\prod_{k\in\mathbb{I}_i}A_{ijk}\right)\left(\prod_{\ell\in\mathbb{I}_i^c}B_{ij\ell}\right)$;

(ii) le vecteur de l'estimateur bayésien $\widetilde{\boldsymbol{h}}_i=\left(\widetilde{h}_{i1},...,\widetilde{h}_{id}\right)^\top$ *de* $\boldsymbol{h}_i$, *donné en* (6.11), *est composé de*

$$\widetilde{h}_{im}=\frac{1}{D_{ij}}\sum_{j=1,j\neq i}^{n}\left(\prod_{k\in\mathbb{I}_i}C_{jk}\right)\left(\prod_{\ell\in\mathbb{I}_i^c}A_{ij\ell}\right)\left(\frac{X_{jm}+\beta_m}{\alpha}\mathbb{1}_{\{0\}}(X_{im})+\frac{B_{ijm}}{\alpha-1/2}\mathbb{1}_{]0,+\infty[}(X_{im})\right),$$

pour $m=1,2,\ldots,d$, *avec les notations précédentes de* $A_{j\ell}$, B_{ijm}, C_{jk} *et* D_{ij}.

Démonstration (i) On réécrit $\pi(\mathbf{h}_i\mid\mathbf{X}_i)$ de (6.10) comme quotient de $N(\mathbf{h}_i\mid\mathbf{X}_i):=\widehat{f}(\mathbf{X}_i\mid\{\mathbf{X}_{-i}\},\mathbf{h}_i)\pi(\mathbf{h}_i)$ et $\int_{[0,+\infty[^d}N(\mathbf{h}_i\mid\mathbf{X}_i)d\mathbf{h}_i$. De (6.9) et (6.12) le numérateur est d'abord égal à

$$\begin{aligned}N(\mathbf{h}_i\mid\mathbf{X}_i) &= \left(\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\prod_{\ell=1}^{d}K_{X_{i\ell},h_{i\ell}}(X_{j\ell})\right)\left(\prod_{\ell=1}^{d}\frac{\beta_\ell^\alpha}{\Gamma(\alpha)}h_{i\ell}^{-\alpha-1}\exp(-\beta_\ell/h_{i\ell})\right)\\ &= \frac{[\Gamma(\alpha)]^{-d}}{(n-1)}\sum_{j=1,j\neq i}^{n}\prod_{\ell=1}^{d}\frac{K_{X_{i\ell},h_{i\ell}}(X_{j\ell})}{\beta_\ell^{-\alpha}h_{i\ell}^{\alpha+1}}\exp(-\beta_\ell/h_{i\ell}).\end{aligned}\tag{6.13}$$

Considérons la grande partie $\mathbb{I}_i^c=\{\ell\in\{1,...,d\}\,;X_{i\ell}\in]0,+\infty[\}$. Suivant Chen (2000), on suppose que pour tout $X_{i\ell}\in]0,+\infty[$ on a $1+(X_{i\ell}/h_{i\ell})\to+\infty$ quand $n\to+\infty$ pour tout $\ell\in\{1,2,...,d\}$. Avec la formule de Stirling $\Gamma(z+1)\simeq\sqrt{2\pi}z^{z+1/2}\exp(-z)$ quand $z\to+\infty$, le terme de la somme dans (6.13) peut être calculé comme suit :

$$\begin{aligned}\frac{GA_{X_{i\ell},h_i}(X_{j\ell})}{\beta_\ell^{-\alpha}h_{i\ell}^{\alpha+1}}\exp(-\beta/h_{i\ell}) &= \frac{X_{j\ell}^{(X_{i\ell}/h_{i\ell})}\exp(-X_{j\ell}/h_{i\ell})}{h_{i\ell}^{1+(X_{i\ell}/h_{i\ell})}\Gamma[1+(X_{i\ell}/h_{i\ell})]\beta^{-\alpha}h_{i\ell}^{\alpha+1}}\exp(-\beta/h_{i\ell})\\ &= \frac{\exp[-(X_{j\ell}+\beta-X_{i\ell}\log X_{j\ell})/h_{i\ell}]}{\beta^{-\alpha}h_{i\ell}^{(X_{i\ell}/h_{i\ell})+\alpha+2}\sqrt{2\pi}\exp(-X_{i\ell}/h_{i\ell})(X_{i\ell}/h_{i\ell})^{(X_{i\ell}/h_{i\ell})+1/2}}\\ &= \frac{\Gamma(\alpha+1/2)}{\beta^{-\alpha}X_{i\ell}^{1/2}\sqrt{2\pi}B_{ij}^{\alpha+1/2}}\times\frac{B_{ij}^{\alpha+1/2}\exp[-B_{ij}/h_{i\ell}]}{h_{i\ell}^{\alpha+3/2}\Gamma(\alpha+1/2)}\\ &= A_{ij\ell}\,\Phi_{\alpha+1/2,B_{ij}}(h_{i\ell}),\end{aligned}\tag{6.14}$$

avec $A_{ij\ell}=[X_{j\ell}^{-1}\Gamma(\alpha+1/2)]/(\beta^{-\alpha}X_{i\ell}^{-1/2}\sqrt{2\pi}B_{ij}^{\alpha+1/2})$, $B_{ij\ell}=X_{i\ell}\log X_{i\ell}-X_{i\ell}\log X_{j\ell}+X_{j\ell}-X_{i\ell}+\beta$ et $\Phi_{\alpha+1/2,B_{ij}}(h_{i\ell})$ est donné en (6.12).

De même, considérant la petite partie $\mathbb{I}_i = \{k \in \{1, ..., d\} \, ; X_{ik} = 0\}$, pour chaque $X_{ik} = 0$, le terme de la somme (6.13) peut s'exprimer par

$$\begin{aligned}
\frac{GA_{0,h_{ik}}(X_{jk})}{\beta_k^{-\alpha} h_{ik}^{\alpha+1}} \exp(-\beta_k/h_{ik}) &= \frac{\exp(-X_{jk}/h_{ik})}{\beta^{-\alpha} h_{ik}^{\alpha+2}} \exp(-\beta_k/h_{ik}) \\
&= \frac{\Gamma(\alpha+1)}{\beta^{-\alpha}(X_{jk}+\beta_k)^{\alpha+1}} \times \frac{(X_{jk}+\beta_k)^{\alpha+1} \exp[-(X_{jk}+\beta_k)/h_{ik}]}{h_{ik}^{\alpha+2}\Gamma(\alpha+1)} \\
&= C_{jk}\, \Phi_{\alpha+1, X_{jk}+\beta_k}(h_{ik}), \qquad (6.15)
\end{aligned}$$

En combinant les résultats (6.14) et (6.15), l'expression $N(\mathbf{h}_i \mid \mathbf{X}_i)$ en (6.13) devient

$$N(\mathbf{h}_i \mid \mathbf{X}_i) = \frac{[\Gamma(\alpha)]^{-d}}{(n-1)} \sum_{j=1, j\neq i}^{n} \left(\prod_{k\in\mathbb{I}_i} C_{jk}\, \Phi_{\alpha+1, X_{jk}+\beta_k}(h_{ik}) \right) \left(\prod_{\ell\in\mathbb{I}_i^c} A_{ij\ell}\, \Phi_{\alpha+1/2, B_{ij\ell}}(h_{i\ell}) \right). \qquad (6.16)$$

De (6.16), le dénominateur est obtenu successivement comme suit :

$$\begin{aligned}
\int_{[0,+\infty[^d} N(\mathbf{h}_i \mid \mathbf{X}_i)\, d\mathbf{h}_i &= \frac{[\Gamma(\alpha)]^{-d}}{(n-1)} \sum_{j=1, j\neq i}^{n} \left(\prod_{k\in\mathbb{I}_i} C_{jk} \int_0^\infty \Phi_{\alpha+1, X_{jk}+\beta_k}(h_{ik})\, dh_{ik} \right) \\
&\quad \times \left(\prod_{\ell\in\mathbb{I}_i^c} A_{ij\ell} \int_0^\infty \Phi_{\alpha+1/2, B_{ij\ell}}(h_{i\ell})\, dh_{i\ell} \right) \\
&= \frac{[\Gamma(\alpha)]^{-d}}{(n-1)} \sum_{j=1, j\neq i}^{n} \left(\prod_{k\in\mathbb{I}_i} C_{jk} \right) \left(\prod_{\ell\in\mathbb{I}_i^c} A_{ij\ell} \right) \qquad (6.17) \\
&= \frac{[\Gamma(\alpha)]^{-d}}{(n-1)} D_{ij}, \qquad (6.18)
\end{aligned}$$

avec $D_{ij} = \sum_{j=1, j\neq i}^{n} \left(\prod_{k\in\mathbb{I}_i} C_{jk} \right) \left(\prod_{\ell\in\mathbb{I}_i^c} A_{ij\ell} \right)$. Finalement, le quotient formé de (6.16) et (6.17) conduit à (i).

(ii) On rappelle que la moyenne de la loi gamma $\mathcal{IG}(\alpha, \beta)$ est $\beta/(\alpha-1)$ et que $\mathbb{E}(h_{i\ell} \mid \mathbf{X}_i) = \int_0^\infty h_{i\ell} \pi(h_{im} \mid \mathbf{X}_i)\, dh_{im}$ avec $\pi(h_{im} \mid \mathbf{X}_i)$ est la distribution marginale de h_{im} obtenue en intégrant $\pi(\mathbf{h}_i \mid \mathbf{X}_i)$ sur toutes les composantes de $\mathbf{h}_i$ autre que h_{im}. Ensuite, $\pi(h_{im} \mid \mathbf{X}_i) = \int_{[0,+\infty[^{d-1}} \pi(\mathbf{h}_i \mid \mathbf{X}_i)\, d\mathbf{h}_{i(-m)}$ où $d\mathbf{h}_{i(-m)}$ est le vecteur $d\mathbf{h}_i$ sans sa $m^{\text{ème}}$ composante. Si $m \in \mathbb{I}_i$:

$$\pi(h_{im} \mid \mathbf{X}_i) = \frac{1}{D_{ij}} \sum_{j=1, j\neq i}^{n} \left(\prod_{k\in\mathbb{I}_i} C_{jk} \right) \left(\prod_{\ell\in\mathbb{I}_i^c} A_{ij\ell} \right) \Phi_{\alpha+1, X_{jm}+\beta_m}(h_{im})$$

et

$$\widetilde{h}_{im} = \mathbb{E}(h_{im} \mid \mathbf{X}_i) = \frac{1}{D_{ij}} \sum_{j=1, j\neq i}^{n} \left(\prod_{k\in\mathbb{I}_i} A_{jk} \right) \left(\prod_{\ell\in\mathbb{I}_i^c} C_{j\ell} \right) \left(\frac{X_{jm}+\beta_m}{\alpha} \right). \qquad (6.19)$$

Si $m \in \mathbb{I}_i^c$:

$$\pi(h_{im} \mid \mathbf{X}_i) = \frac{1}{D_{ij}} \sum_{j=1, j\neq i}^{n} \left(\prod_{k \in \mathbb{I}_i} C_{jk} \right) \left(\prod_{\ell \in \mathbb{I}_i^c} A_{ij\ell} \right) \Phi_{\alpha+1/2, B_{ijm}}(h_{im})$$

et

$$\widetilde{h}_{im} = \mathbb{E}(h_{im} \mid \mathbf{X}_i) = \frac{1}{D_i} \sum_{j=1, j\neq i}^{n} \left(\prod_{k \in \mathbb{I}_i} C_{jk} \right) \left(\prod_{\ell \in \mathbb{I}_i^c} A_{ij\ell} \right) \left(\frac{B_{ijm}}{\alpha - 1/2} \right). \tag{6.20}$$

On obtient le résultat (ii) en combinant (6.19) et (6.20). ■

Finalement, l'estimateur de f par noyau gamma multiple à choix bayésien adaptatif des fenêtres s'écrit :

$$\widehat{f_n}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{\ell=1}^{d} GA_{x_\ell, \widetilde{h}_{i\ell}}(X_{i\ell}), \quad \forall \mathbf{x} = (x_1, \ldots, x_d)^\top \in \mathbb{T}_d := [0, +\infty[^d. \tag{6.21}$$

Remarque. On peut remplacer l'estimateur à noyau gamma multiple (6.8) dans la formule (6.10) par l'un des noyaux lognormal, gaussien inverse, gaussien inverse réciproque où Birnbaum-Saunders avec les lois *a priori* appropriées. Toutefois, il n'est pas toujours aisé d'avoir une forme explicite de la loi *a posteriori* (6.10) et de l'estimateur bayésien (6.11) des fenêtres. Dans ce cas, l'estimateur bayésien des fenêtres peut être calculé par les méthodes de Monte-Carlo par chaînes de Markov (MCMC) ; le lecteur peut se référer par exemple à Zougab *et al.* (2014) et aux références qui s'y trouvent.

Conclusion et perspectives

Ce travail a porté sur l'estimation des fonctions de densité, de masse de probabilité et de régression en utilisant les noyaux associés appropriés tant univariés que multivariés.

Récapitulatif

Tout d'abord, l'état de l'art de l'estimation à noyaux classiques multivariés et associés univariés de Senga Kiéssé (2008) et Libengué (2013) a été fait. Ce faisant, les notions d'erreurs moyennes quadratiques ainsi que des formes et techniques de sélection de matrice de lissages ont été développées. Il faut noter que dans ce travail de rappel, la méthode des noyaux associés univariés a été rappelée. La question de la régression multiple par noyaux classiques a été aussi abordée.

Ensuite, des noyaux associés multivariés continus ont été présentés. Cette méthode étend au multivarié celle d'univarié de Libengué (2013) en tenant compte des effets de différentes formes de matrice de lissages. Leur construction par le principe "mode-dispersion multivarié" consiste à résoudre un système d'équations formé par $\mathbf{x} = \mathbf{m}$ et $\mathbf{H} = \mathbf{D}$ où $\mathbf{m}$ et $\mathbf{D}$ sont, respectivement, l'unique vecteur mode et la matrice de dispersion de $K_{\theta(\mathbf{m},\mathbf{D})}$. La fonction $K_{\theta(\mathbf{m},\mathbf{D})}$ est une densité de probabilité paramétrée et de carré intégrable. Cela conduit au noyau associé $K_{\theta(\mathbf{x},\mathbf{H})} := K_{\theta(\mathbf{m}(\mathbf{x},\mathbf{H}),\mathbf{D}(\mathbf{x},\mathbf{H}))}$ de support $\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})} = \mathbb{S}_{\theta(\mathbf{m}(\mathbf{x},\mathbf{H}),\mathbf{D}(\mathbf{x},\mathbf{H}))}$, où $\mathbf{m}(\mathbf{x},\mathbf{H})$ et $\mathbf{D}(\mathbf{x},\mathbf{H})$ sont solutions des équations précédentes. Ceci a été illustré sur le noyau bêta bivarié à structure de corrélation de Sarmanov (1966). Par la suite, l'estimateur standard à noyau associé multivarié dans le cas continu a été défini. Il a été montré que ces estimateurs sont sans effet de bords mais possèdent un biais plus grand. En effet, le nombre de termes du biais des estimateurs à noyaux associés est supérieur à celui à noyau classique. Pour corriger cela, un algorithme de réduction du biais consistant à partitionner le support $\mathbb{T}_d$ en deux régions (de bords et la plus grande partie d'intérieur) puis à modifier le noyau associé de départ de sorte que le biais de l'estimateur issu de ce noyau soit de même ordre que celui de l'estimateur à noyau classique multivarié.

De là, des études par simulations et données réelles avec l'estimateur standard du noyau bêta bivarié avec corrélation et avec trois formes de matrices des fenêtres (diagonale, Scott et pleine) ont été effectuées. De ces études, la nécessité du noyau bêta bivarié à matrice pleine a été mis en exergue pour des densités à forme complexe.

Par ailleurs, le noyau bêta bivarié à matrices de lissage pleine et diagonale ont été étudié, de manière pratique dans un problème de régression multiple. Il a été montré que l'usage des matrices pleines ou diagonales, pour le noyau bêta bivarié, avait peu d'effet sur la qualité de la régression. Pour les noyaux associés multiples composés

d'univariés (gamma, bêta, DiracDU, binomial, triangulaire discret), l'importance et l'efficacité du bon choix de noyau associé par rapport au support des variables explicatives a été montré par simulations et applications. Les noyaux associés multiples appropriés à chaque support ont alors été introduits et mis en oeuvre dans un problème d'analyse discriminante. Les choix de fenêtres par validation croisée (usuelle et profilée) montrent l'efficacité de la méthode par simulations.

Enfin, un package `R` dénommé `Ake` a été créé pour la vulgarisation de la méthode des noyaux associés univariés pour l'estimation de densités, de masse de probabilité et de régression. Ce package inclut plusieurs méthodes de sélections de fenêtres de lissage (validation croisée et bayésiennes) et contribuera à promouvoir la méthode auprès des praticiens.

Remarques finales

Les travaux réalisés offrent des perspectives très intéressantes. Tout d'abord concernant les noyaux associés non-classiques et multivariés, il serait intéressant de faire une étude fine de l'algorithme de réduction du biais sur la qualité de l'estimation, et de les comparer aux estimateurs à noyaux standards et à noyaux classiques multivariés. On s'intéressera notamment, par simulations, aux effets au bords du support des densités à estimer ; voir, par exemple, Malec & Schienle (2014) et Igarashi & Kakizawa (2015) dans le cas univarié. La question du noyau optimal dans les noyaux associés mérite aussi une attention particulière. La construction d'autres noyaux associés (continues, discrets et mixtes) multivariés avec structure de corrélation est aussi envisageable soit par la technique de Sarmanov (1966) ou par les copules (voir p. ex. Scaillet *et al.*, 2007). Un autre sujet intéressant serait d'étendre la méthode de construction de noyaux associés multivariés continus au cas discrets. L'idée de construction d'une version récursive de l'estimateur à noyau associé est de même souhaitable comme dans Amiri *et al.* (2014) pour les noyaux classiques. Dans le cas discret, les noyaux associés multivariés sont, de plus, utilisables pour des problèmes évoqués dans Kokonendji (2014) ; par exemple, séries temporelles à valeurs entières et modèles de régressions semi-paramétriques.

Puisque les matrices des fenêtres jouent un rôle crucial dans la qualité des différents lissages, il serait intéressant d'étudier la performance du choix bayésien adaptatif pour gamma multiple à d'autres noyaux associés (discrets ou continus) avec ou sans structure de corrélation. À l'image de Racine (2002) pour les noyaux classiques univariés, on pourrait accélérer la sélection des matrices de lissages par validation croisée en faisant du calcul parallèle. Une fois que les sélections de matrices de lissages seront automatiques, on pourra aussi s'intéresser à l'élaboration d'un package sous `R` pour, entre autre, l'estimation de fonctions de densité, de masse de probabilité et de régression par noyaux associés multivariés. Aussi, on pourra envisager des estimations par noyaux associés d'autres fonctionnelles telles que les quantiles.

Il serait intéressant de poursuivre la réflexion autour du noyau associé mixte (à la fois discrets et continus) univarié introduit par Libengué *et al.* (2013) et Libengué (2013). Une étude fine des différents effets de bords tant au passage du continu au discret qu'aux extrémités sont à envisager. Les outils d'échelles de temps de l'univarié serviront aussi à une généralisation au multivarié.

Références

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

[2] Agarwal, R.P. & Bohner. M. (1999). Basic calculus on time scales and some of its applications, *Results in Mathematics* **35**, 3–22.

[3] Aitchison, J. & Aitken, C.G.G. (1976), Multivariate binary discrimination by the kernel method, *Biometrika* **63**, 413–420.

[4] Amiri, A., Crambes, C. & Thiam, B. (2014). Recursive estimation of nonparametric regression with functional covariate, *Computational Statistics and Data Analysis* **69**, 154–172.

[5] Antoniadis, A. (1997). Wavelets in statistics : a review (with discussion), *Journal of the Italian Statistical Society Series B* **6**, 97–144.

[6] Azzalini A, Bowman, AW (1990). A Look at some Data on the Old Faithful Geyser, *Applied Statistics* **39**, 357–365.

[7] Balakrishnan, N. & Lai, C.D. (2009). *Continuous Bivariate Distributions*, Springer, New York.

[8] Bertin, K. & Klutchnikoff, N. (2014). Adaptative estimation of a density function using beta kernels, *ESAIM : Probability and Statistics* **18**, 400–417.

[9] Bohner, M. & Peterson, A. (2001). *Dynamic Equations on Time Scales*, Birkhera Boston Inc., Boston.

[10] Bohner, M. & Peterson, A. (2003). *Advances in Dynamic Equations on Time Scales*, Birkhera Boston Inc., Boston.

[11] Boucheron, S., Lugosi, G. & Massart, P. (2013). *Concentration Inequalities : A Nonasymptotic Theory of Independence*, Oxford University Press.

[12] Bouezmarni, T. & Rombouts, J.V.K. (2010). Nonparametric density estimation for multivariate bounded data, *Journal of Statistical Planning and Inference* **140**, 139–152.

[13] Brewer, M.J. (2000). A Bayesian model for local smoothing in kernel density estimation, *Statistics and Computing* **10**, 299–309.

[14] Brown, B.M. & Chen, S.X. (1998). Beta Bernstein smoothing for regression curves with compact support, *Scandinavian Journal of Statistics* **26**, 47–59.

[17] Cardot, H., De Moliner, A. & Goga, C. (2015). Estimating with kernel smoothers the mean of functional data in a finite population setting. A note on variance estimation in presence of partially observed trajectories, *Statistics and Probability Letters* **99**, 156–166.

[16] Chacón, J.E. & Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices, *Test* **19**, 375–398.

[17] Chacón, J.E. & Duong, T. (2011). Unconstrained pilot selectors for smoothed cross validation, *Australian and New Zealand Journal of Statistics* **53**, 331–351.

[18] Chacón, J.E., Duong, T. & Wand, M.P. (2011). Asymptotics for general multivariate kernel density derivative estimators, *Statistica Sinica* **21**, 807–840.

[19] Chen, S.X. (1999). A beta kernel estimation for density functions, *Computational Statistics and Data Analysis* **31**, 131–145.

[20] Chen, S.X. (2000a). Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics* **52**, 471–480.

[21] Chen, S.X. (2000b). Beta kernels smoothers for regression curves, *Statistica Sinica* **52**, 73–91.

[22] Cline, D.B.H. & Hart, J.D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives.

[23] Cohen, L. (1984). Probability distributions with given multivariate marginals, *Journal of Mathematical Physics* **25**, 2402–2403.

[24] Cribari-Neto, F. & Zeilis, A. (2010). Beta regression in R, *Journal of Statistical Software* **34**, 1–24.

[25] Dabo-Niang, S. & Yao, A-F. (2013). Spatial kernel density estimation for functional random variables, *Metrika* **76**, 19–52.

[26] Daouia, A., Gardes, L. & Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernouilli* **19**, 2557–2589.

[27] Duong, T. (2004). *Bandwidth Selectors for Multivariate Kernel Density Estimation*. Ph.D. Thesis Manuscript to University of Western Australia, Perth, Australia, October 2004.

[28] Duong, T. (2007). ks : Kernel density estimation and kernel discriminant analysis for multivariate data in R, *Journal of Statistical Software* **21**, 1–16.

[29] Epanechnikov, V.A. (1969). Nonparametric estimation of a multivariate probability density, *Theory of Probability and Its Applications* **14**, 153–158.

[30] Funke, B. & Kawka, R. (2015). Nonparametric density estimation for multivariate bounded data using two nonnegative multiplicative bias correction methods, *Computationnal Analysis and Data Analysis* **92**, 148–162.

[31] Gasser, T. & Müller, H.G. (1979). Kernel estimation of regression functions. Smoothing techniques for curve estimation *Proceedings of Workshop, Heidelberg*.

[32] Gasser, T., Müller, H.G. & Mammitzsch, V. (1985). Kernels for nonparametric curve estimation, *Journal of the Royal Statistical Society* **2**, 238-252.

[33] Girard, S., Guillou, A. & Stupfler, G. (2013). Frontier estimation with kernel regression on high order moments, *Journal of Multivariate Analysis* **116**, 172–189.

[34] Gosh, A.K. & Chaudhury, P. (2004). Optimal smoothing in kernel analysis discriminant, *Statistica Sinica* **14**, 457–483.

[35] Gosh, A.K. & Hall, P. (2008). On error-rate estimation in nonparametric classification, *Statistica Sinica* **18**, 1081–1100.

[36] González-Manteiga, W., Lombardía, M.J., Martínez-Miranda, M.D. & Sperlich, S. (2013). Kernel smoothers and bootstrapping for semiparametric mixed effects models, *Journal of Multivariate Analysis* **114**, 288–302.

[37] Gu, C. (1993). Smoothing spline density estimation : A dimensionless automatic algorithm, *Journal of the American Statistical Association* **88**, 495–504.

[38] Habbema, J.D.F., Hermans, J. & Vanden Broee, K. (1974). A stepwise discriminant analysis program using density estimation, *Compstat 1974, Proceedings in Computationnal Statistics*, Ed. G. Bruckmnann, 101–110, Physica Verlag, Vienna.

[39] Hall, P. & Kang, K.-H. (2005). Bandwidth choice for nonparametric classification, *Annals of Statistics* **33**, 284–306.

[40] Hall, P. & Wand, M.P. (1988). On nonparametric discrimination using density differences, *Biometrika* **75**, 541–547.

[41] Hayfield, T. & Racine, J.S. (2007). Nonparametric econometrics : the np package, *Journal of Statistical Software* **27**, 1–32.

[42] Hazelton, M.L., Marshall, J.C., 2009. Linear boundary kernels for bivariate density estimation, *Statistics and Probability Letters* **79**, 999–1003.

[43] Härdle, W. & James, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation, *Annals of Statistics* **13**, 1465–1481.

[44] He, J., Yang, G., Rao, H., Li, Z., Ding, X. & Chen Y. (2012). Prediction of human major histocompatibility complex class II binding peptides by continuous kernel discrimination method, *Artificial Intelligence in Medecine* **55**, 107–115.

[45] Henderson, H.V. & Searle, S.R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics, *Canadian Journal of Statistics* **7**, 65–81.

[46] Hernández-Bastida, A. & Fernández-Sanchez, M.P. (2012). A Sarmanov family with beta and gamma marginal distributions : an application to the Bayes premium in a collective risk model, *Statistical Methods & Applications* **21**, 391–409.

[47] Hielscher, R. (2013). Kernel density estimation on the rotation group and its application to crystallographic texture analysis, *Journal of Multivariate Analysis* **119**, 119–143.

[48] Hilger, S. (1990). Analysis on measure chains – a unified approach to continuous and discrete calculus, *Results in Mathematics* **18**, 18–56.

[49] Hilger, S. (1998). *Ein Maettenkalkl mit Anwendung auf Zentrumsmannigfal- tigkeiten*. Universitrzburg.

[50] Hirukawa, M. & Sakudo, M. (2014). Nonnegative bias reduction methods for density estimation using asymmetric kernels, *Computational Statistics and Data Analysis* **75**, 112–123.

[51] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* **58**, 13–30.

[52] Ichimura, T. & Fukuda, D. (2010). A fast algorithm for computing least-squares cross-validations for nonparametric conditional kernel density functions, *Computational Statistics and Data Analysis* **54**, 3404–3410.

[53] Igarashi, G. & Kakizawa, Y. (2014). Re-formulation of inverse Gaussian, reciprocal inverse Gaussian, and Birnbaum-Saunders kernel estimators, *Statistics and Probability Letters* **84**, 235–246.

[54] Igarashi, G. & Kakizawa, Y. (2015). Bias corrections for some aymmmetric kernel estimators, *Journal of Statistical Planning and Inference* **159**, 37–63.

[55] Johnson, N.L., Kotz, S. & Balakrishnan, N. (1997). *Discrete Multivariate Distributions*, John Wiley and Sons, New York.

[56] Jørgensen, B. (1997). *The Theory of Dispersion Models*, Chapman & Hall, London.

[57] Jørgensen, B. (2013). Construction of multivariate dispersion models, *Brazilian Journal of Probability and Statistics* **27**, 285–309.

[58] Jørgensen, B. & Kokonendji, C.C. (2011). Dispersion models for geometric sums, *Brazilian Journal of Probability and Statistics* **25**, 263–293.

[59] Jørgensen, B. & Kokonendji, C.C. (2015). Discrete dispersion models and their Tweedie asymptotics, *AStA Advances in Statistical Analysis*, `(DOI :10.1007/s10182-015-0250-z)`, à paraître.

[60] Klemelä, J. & Ripley, M.C. (1995). `regpro` : Nonparametric Regression, URL `http://cran.r-project.org/web/packages/regpro/index.html/`.

[61] Kokonendji, C.C. (2014). Over- and Underdispersion Models. In *The Wiley Encyclopedia of Clinical Trials - Methods and Applications of Statistics in Clinical Trials, Vol. 2 : Planning, Analysis, and Inferential Methods*, Chap. 30, pp. 506-526, (Editor N. Balakrishnan), John Wiley & Sons, Newark, NJ.

[62] Kokonendji, C.C. & Senga Kiessé, T. (2011). Discrete associated kernels method and extensions, *Statistical Methodology* **8**, 497–516.

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

[64] Kokonendji, C.C., Senga Kiessé, T. & Demétrio, C.G.B. (2009). Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics* **12**, 99–125.

[65] Kokonendji, C.C., Senga Kiessé, T. & Zocchi, S.S. (2007). Discrete triangular distributions and nonparametric estimation for probability mass function, *Journal of Nonparametric Statistics* **19**, 241–254.

[66] Kokonendji, C.C. & Somé, S.M. (2015). On multivariate associated kernels for smoothing some density function, arXiv:1502.01173.

[67] Kokonendji, C.C. & Varron, D. (2015). Performance of the discrete associated kernel estimator through the total variation distance, Submitted to publication.

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

[69] Kotz, S., Balakrishnan, N. & Johnson, L.N. (2000). *Continuous Multivariate Distributions*, John Wiley and Sons, Chicester.

[70] Kundu, D., Balakrishnan, N. & Jamalizadeh, A. (2010). Bivariate Birnbaum-Saunders distribution and associated inference, *Journal of Multivariate Analysis* **101**, 113–125.

[71] Kuruwita, C.N., Kulasekera, K.B. & Padgett, W.J. (2010). Density estimation using asymmetric kernels and Bayes bandwidths with censored data, *Journal of Statistical Planning and Inference* **140**, 1765–1774.

[72] Lee, M-L.T. (1996). Properties and applications of the Sarmanov family of bivariate distributions, *Communications in Statistics - Theory and Methods* **25**, 1207–1222.

[73] Li, Q. & Racine, J. (2007). *Nonparametric Econometrics : Theory and Practice*, Princeton University Press.

[74] Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*. Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France & Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon, URL http://hal.in2p3.fr/tel-01124288/document.

[75] Libengué, F.G., Somé, S.M. & Kokonendji, C.C. (2013). Estimation par noyaux associés mixtes d'un modèle de mélange. *Actes des 45èmes Journées de Statistique de la Société Française de Statistiques (SFdS)*, 6pages. URL `http://papersjds13.sfds.asso.fr/submission82.pdf`.

[76] Liu, B., Yang, Y., Webb, I. & Boughton, J. (2012). A comparative study of bandwidth choice in kernel density estimation for naive Bayesian classification, *Advances in Knowledge Discovery and Data Mining*, 302–313, Springer, Berlin.

[77] Magnus, J.R. & Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley and Sons, Chichester.

[78] Malec, P. & Schienle, M. (2014). Nonparametric kernel density estimation near the boundary, *Computational Statistics and Data Analysis* **72**, 57–76.

[79] Markovich, N. (2008). *Nonparametric Analysis of Univariate Heavy-Tailed Data*, Jonh Wiley & Sons, Chicester.

[80] Markovich, N. (2008). *Nonparametric Analysis of Univariate Heavy-Tailed Data*, John Wiley & Sons, Chicester.

[81] Marsh, L.C. & Mukhopadhyay, K. (1999). Discrete Poisson kernel density estimation with an application to wildcat coal strikes, *Applied Economics Letters* **6**, 393–396.

[82] Müller, H.G. & Stadtmüller, U. (1999), Multivariate boundary kernels and a continuous least squares principle, *Journal of the Royal Statistical Society B* **61**, 439–458.

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

[84] Parzen, E. (1962). On estimation of a probability density function and mode, *Annals of Mathematical Statistics* **33**, 1065–1076.

[85] R Development Core Team (2015). *R : A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://cran.r-project.org/`.

[86] Racine, J. (2002). Parrallel distributed kernel estimation, *Computationnal Statistics and Data Analysis* **40**, 293–302.

[87] Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*, Cambridge University Press.

[88] Romano, J.P. & Thombs, L.A. (1996). Inference for autocorrelations under weak assumptions, *Journal of the American Statistical Association* **91**, 590–600.

[89] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27**, 832–837.

[90] Sanyal, S. (2008), *Stochastic Dynamic Equations*, Ph.D. Thesis Manuscript to Missouri University of Sciences and Technology.

[91] Sain, S.R. (2002). Multivariate locally adaptive density estimation, *Computational Statistics and Data Analysis* **39**, 165–186.

[92] Sarmanov, O.V. (1966). Generalized normal correlation and two-dimensionnal Frechet classes, *Doklady (Soviet Mathematics)* **168**, 596–599.

[93] Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels, *Journal of Nonparametric Statistics* **16**, 217–226.

[94] Scaillet, O., Charpentier, A. & Fermanian, J.-D. (2007). *The Estimation of Copulas : Theory and Practice*, URL `http://www.angelfire.com/falcon/isinotes/mult/cop2.pdf`.

[95] Scott, W.D. (1992). *Multivariate Density Estimation*, John Wiley and Sons, New York.

[96] Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities, *Communications in Statistics - Theory and Methods* **40**, 1123–1136.

[97] Senga Kiessé, T. (2008). *Approche Non-Parametrique par Noyaux Associés Discrets des Données de Dénombrement*. Ph.D. Thesis Manuscript (in French), Université de Pau. `http://tel.archives-ouverts.fr/tel-00372180/fr/`.

[98] Silvermann, B.W. (1985). Some aspects of the smoothing approach to nonparametric regression curve fitting, *Journal of the Royal Statistical Society Series B* **47**, 1–52.

[99] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[100] Simonoff, J.S. (1996). *Smoothing Methods in Statistics*, Springer, New York.

[101] Simonof, J.S. & Tutz, G. (2000). Smoothing methods for discrete data. *In : Smoothing and Regression : Approaches, Computation, and Application (ed. M.G. Shimek)*. pp. 193-228. Wiley, New York.

[102] Somé, S.M. & Kokonendji, C.C. (2015). Effects of associated kernels in nonparametric multiple regressions, `arXiv:1502.01488`.

[103] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley and Sons, New York.

[104] Wahba, G. (1990). *Splines Models for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia.

[105] Wand, M.P. & Jones, M.C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation, *Journal of the American Statistical Association* **88**, 520–528.

[106] Wand, M.P. & Jones, M.C. (1995). *Kernel Smoothing, Vol. 60 of Monographs on statistics and applied probability*, Chapman and Hall, London.

[107] Wand, M.P. & Ripley, M.C. (1995). `KernSmooth` : Functions for Kernel Smoothing Supporting Wand Jones (1995), URL `http://cran.r-project.org/web/packages/KernSmooth/index.html/`.

[108] Wand, M.P. & Ruppert, D. (1994). Multivariate locally weighted least squares regression, *Annals of Statistics* **22**, 1346–1370.

[109] Wang, M.-C. & Van Ryzin, J. (1981). A class of smooth estimators for discrete distributions, *Biometrika* **68**, 301–309.

[110] Wansouwé, W.E. Kokonendji, C.C. & Kolyang, D.T. (2015a). `Disake` : an R package for discrete associated kernel estimator, URL `http://cran.r-project.org/web/packages/Disake/`.

[111] Wansouwé, W.E., Libengué, F.G. & Kokonendji, C.C. (2015b). `Conake` : an R package for continuous associated kernel estimators, URL `http://CRAN.R-project.org/package = Conake/`.

[112] Wansouwé, W.E., Somé, S.M. & Kokonendji, C.C. (2015c). `Ake` : Associated kernel estimations, URL `http://cran.r-project.org/web/packages/Ake/`.

[113] Wansouwé, W.E., Somé, S.M. & Kokonendji, C.C. (2015d). `Ake:` an `R` package for discrete and continuous associated kernel estimations. Submitted to *Journal of Statistical Software*.

[114] Wassermann, L. (2006). *All of Nonparametrics Statistics*, Springer, Berlin.

[115] Watson, G.S. (1964). Smooth regression analysis, *Sankhya Series A* **26**, 359–372.

[116] Yahav, I. & Shmueli, G. (2012). On generating multivariate Poisson data in management science applications, *Applied Stochastic Models in Business and Industry* **28**, 91–102.

[117] Zhang, S. (2010). A note on the performance of the gamma kernel estimators at the boundary, *Statistics and Probability Letters* **80**, 548–557.

[118] Zhang, S. & Karunamuni, R.J. (2010). Boundary performance of the beta kernel estimators, *Journal of Nonparametric Statistics* **22**, 81–104.

[119] Zhang, X., King, M.L. & Shang, H.L. (2014). A sampling algorithm for bandwidth estimation in a nonparametric regression model with a flexible error density, *Computational Statistics and Data Analysis* **78**, 218–234.

[120] Ziane, Y., Adjabi, S. & Zougab, N. (2015). Adaptive Bayesian bandwidth selection in asymmetric kernel density estimation for nonnegative heavy-tailed data, *Journal of Applied Statistics* **42**, 1645–1658.

[121] Zougab, N. (2013). *Approche Bayésienne dans l'Estimation Non-paramétrique de la Densité de Probabilité et la Courbe de Régression de la Moyenne*. Thèse à Université Abderrahmane Mira de Béjaïa, Algérie.

[122] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation, *Journal of Nonparametrics Statistics* **24**, 783–795.

[123] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2013). A Bayesian approach to bandwidth selection in univariate associate kernel estimation, *Journal of Statistical Theory and Practice* **7**, 8–23.

[124] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2014a). Bayesian approach in nonparametric count regression with binomial kernel, *Communications in Statistics - Simulation and Computation* **43**, 1052–1063.

[125] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2014b). Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation, *Computational Statistics and Data Analysis* **75**, 28–38.

Annexe : Codes sources avec R

A1. Estimateur de densité multivarié (Chapitre 2)

Cette partie présente le code R de la procédure d'estimation de densité par noyau bêta bivarié avec structure de corrélation de Sarmanov (1966). Le cas de la matrice diagonale est simplement obtenu en remplaçant en ignorant le calcul des éléments non diagonaux.

```
#######################################################################
#Grille pour calcul de l'intégrale par methode de Simpson
#######################################################################
#borne de l'intégrale bidimensionnelle
a=0;b=1;c=0;d=1
#######################################################################"
Nx=27
Ny=27
####Construction des noeuds de simpson et des différents poids pour x ###
h = (b - a)/(Nx-1);
 x = seq(a,b,h);
wx = rep(1,Nx);   wx[seq(2,Nx-1,2)] = 4; wx[seq(3,Nx-2,2)] = 2; wx = wx*h/3;
######Construction des noeuds de simpson et des différents poids pour y  ######
 y = seq(c,d,h);
wy = rep(1,Ny); wy[seq(2,Ny-1,2)] = 4; wy[seq(3,Ny-2,2)] = 2; wy = wy*h/3;
######## Combinaison des poids pour l'intégrale multiple#########
fentr=expand.grid(x=x,y=y)
x=fentr[,1]
y=fentr[,2]
x1<-x
x2<-y
w<-outer(wx,wy,"*")
w1<-as.vector(t(w))
#########################################################
#########################################################################
```

```
################Noyau beta bivarié de Sarmanov#############
#############################################################
BetaSarmanov<-function(x1,x2,h11,h22,h12,y1,y2)
{
  T=rep(0,length(y));
   # y1 : vecteur des points y2=0:1
     # x1 : centre du vecteur de points x=5
     # h1 : paramtre de lissage discret h=0.1

   # y2 : vecteur des points y2=0:1
     # x2 : centre du vecteur de points x=5
     # h2 : paramtre de lissage discret h=0.1

     for (i in 1 :length(x1)) # boucle en i pour chaque point x[i]

      {for (j in 1:length(y1))

        for (k in 1 :length(x2)) # boucle en i pour chaque point x[i]

          {for (l in 1:length(y2))
                {
 T[j,l]=((1/beta((x1[i]/h11)+1,((1-x1[i])/h11)+1))*y1[j]^((x1[i]/h11)+1-1)
 *(1-y1[j])^(((1-x1[i])/h11)+1-1))*((1/beta((x2[k]/h22)+1,((1-x2[k])/h22)+1))
 *y2[l]^((x2[k]/h22)+1-1)*(1-y2[l])^(((1-x2[k])/h22)+1-1))*
 (1+((sqrt(((1+2*h11)^2*(3*h11+1))/((x1[i]+h11)*(1+h11-x1[i])))*
 sqrt(((1+2*h22)^2*(3*h22+1))/((x2[k]+h22)*(1+h22-x2[k]))))*
 (h12/(h11*h22))*(y1[j]-((x1[i]+h11)/(1+2*h11)))*(y2[l]-((x2[k]+h22)/(1+2*h22)))))

                }
          }
      }

 return(T)
}
####################VALIDATION CROISEE################

#########################################################################
####Generation des l'ensemble des matrices pour la validation croisée#
#########################################################################
Hr=array(0,dim=c(2,2,6,length(h11),length(h22)))
     for ( k in 1 : length(h11)){
         for ( l in 1 : length(h22)){
       for ( r in 1 : length(h12r[1,,k,l])){
Hr[,,r,k,l]=matrix(c(h11[k],h12r[1,r,k,l],h12r[1,r,k,l],h22[l]),
nrow=2,ncol=2,byrow=T)
                                          }
                                    }
                                }
#########################################################################
##Validation croisé
```

```
#########################################################################
CV1bonr<-function(x1,x2,V1r,V2r,Hr,w1)
{  n<-length(V1r)
  U<-array(0,dim=c(2,2,6,length(Hr[1,1,6,,1]),length(Hr[2,2,6,1,])))
   K<-array(0,dim=c(length(x1),length(V1r)))

    for ( k in 1 : length(Hr[1,1,6,,1])){   #vecteur des h11
      for ( l in 1 : length(Hr[2,2,6,1,])){   #vecteur des h22
        for ( r in 1 : length(Hr[1,2,,k,l]))
      { m=rep(0,length(x1))

       for (i in 1 :length(x1)) # boucle en i pour chaque point x[i]
         {for (j in 1 :length(V1r)) # boucle en j pour chaque observation V[j]
           { K[i,j]<-(1/n)*dbeta(V1r[j],(x1[i]/Hr[1,1,6,k,1])+1,
           ((1-x1[i])/Hr[1,1,6,k,1])+1)*dbeta(V2r[j],(x2[i]/Hr[2,2,6,1,l])+1,
           ((1-x2[i])/Hr[2,2,6,1,l])+1)*(1+((sqrt(((x1[i]+Hr[1,1,6,k,1])^2
*(3*Hr[1,1,6,k,1]+1))/((x1[i]+Hr[1,1,6,k,1])*(1+Hr[1,1,6,k,1]-x1[i])))
*sqrt(((x2[i]+Hr[2,2,6,1,l])^2*(3*Hr[2,2,6,1,l]+1))/((x2[i]+Hr[2,2,6,1,l])
*(1+Hr[2,2,6,1,l]-x2[i]))))*(Hr[1,2,r,k,l]/(Hr[1,1,6,k,1]*Hr[2,2,6,1,l]))
*(V1r[j]-((x1[i]+Hr[1,1,6,k,1])/(1+2*Hr[1,1,6,k,1])))
*(V2r[j]-((x2[i]+Hr[2,2,6,1,l])/(1+2*Hr[2,2,6,1,l]))))# noyau bêta bivarié

                         }
}

G<-apply(K,1, sum) #
U[,,r,k,l] <- crossprod(w1,G^2) #vecteur de tous ISE selon les simulations

  }
 }
 }
  return(U)
}

#########################################################################
CV2bonr<-function(V1r,V2r,Hr)
{
 n<-length(V1r)
   K<-array(0,dim=c(length(V1r),length(V1r)))

    for ( k in 1 : length(Hr[1,1,6,,1])){   #vecteur des h11
      for ( l in 1 : length(Hr[2,2,6,1,])){   #vecteur des h22
      {  for (i in 1 : length(V1r)) # boucle en i pour chaque point x[i]

         {for (j in 1 :length(V1r)) # boucle en j pour chaque observation V[j]

                         {
K[i,j]<-2*(1/n)*(1/(n-1))*dbeta(V1r[j],(V1r[i]/Hr[1,1,6,k,1])+1,
((1-V1r[i])/Hr[1,1,6,k,1])+1)*dbeta(V2r[j],(V2r[i]/Hr[2,2,6,1,l])+1,
((1-V2r[i])/Hr[2,2,6,1,l])+1)
```

```
*sqrt(((V1r[i]+Hr[2,2,6,1,l])^2*(3*Hr[2,2,6,1,l]+1))/((V2r[i]+Hr[2,2,6,1,l])
*(1+Hr[2,2,6,1,l]-V2r[i]))))*(Hr[1,2,r,k,l]/(Hr[1,1,6,k,1]*Hr[2,2,6,1,l]))*
(V1r[j]-((V1r[i]+Hr[1,1,6,k,1]))*
(V2r[j]-((V2r[i]+Hr[2,2,6,1,l])/(1+2*Hr[2,2,6,1,l])))))# noyau beta bivarié
               }

           }

G<-apply(K,1, sum)
               W <- sum(G)
   U[,,r,k,l] <- W
    }
  }
  }
return(U)
}
##############################################################################
 Y2r<-CV2bonr(V1r,V2r,Hr)
 Y1r<-CV1bonr(x1,x2,V1r,V2r,Hr,w1)
CV_Hr<-Y1r-Y2r
##################
Hcvr<-array(0,dim=c(2,2))
Hcvr[,]<-matrix(Hr[which(CV_Hr[,,,] == min(CV_Hr[,,]))]
,byrow=T)
Hcvr
###############################################
#############################################################
#esimateur à noyau beta bivarié avec corrélation
#############################################################################
estimbetaSarmanov<-function(x1,x2,V1r,V2r,Hcvr)
   { A=0
   B=0
   n=length(V1r)
   m=length(x1)
   K<-array(0,dim=c(length(x1),length(x2),length(V1r)))
     A<-rep(0,length(x1))       #rep(0,length(x))
         for (i in 1 :length(x1)) # boucle en i pour chaque point x[i]
    {for (k in 1 :length(x2)) # boucle en i pour chaque point x[i]
      {for (l in 1:length(V1r))  # boucle en j pour chaque obs y

        {K[i,k,l]<-(1/n)*dbeta(V1r[l],(x1[i]/Hcvr[1,1])+1,((1-x1[i])
/Hcvr[1,1])+1)*dbeta(V2r[l],(x2[k]/Hcvr[2,2])+1,((1-x2[k])/Hcvr[2,2])+1)
*(1+((sqrt(((x1[i])*Hcvr[1,1]+1))/((x1[i]+Hcvr[1,1])
*(1+Hcvr[1,1]-x1[i])))*sqrt(((x2[k]+Hcvr[2,2])^2
*(3*Hcvr[2,2]+1))/((x2[k]+Hcvr[2,2])*(1+Hcvr[2,2]-x2[k]))))*
(Hcvr[1,2]/(Hcvr[1,1]*Hcvr[2,2]))*(V1r[l]-((x1[i]+Hcvr[1,1])/
(1+2*Hcvr[1,1])))*(V2r[l]-((1+2*Hcvr[2,2])))))

                  }
               }
```

```
}
G <- apply(K,c(1,2),sum)

return(G)
}
##############################################################################
```

A2. Analyse discriminate par noyaux associés (Chapitre 3)

La méthode d'analyse discriminante par noyaux associés est données uniquement en bivarié. En ce qui concerne les dimensions supérieures, il suffit de remplacer de remplacer, à chaque fois, le noyau associé bivarié.

```
#################Analyse discrimiante par noyaux associés multiple##########
##############################################################################
############Selection de la matrice optimale par validation croisée profilée
#
Hlscv <-
function(x1,x2,V1r,V2r,h11,h22,w){

Hrp=array(0,dim=c(2,2,length(h11),length(h22)))
     for ( k in 1 : length(h11)){
         for ( l in 1 : length(h22)){
Hrp[,,k,l]=matrix(c(h11[k],0,0,h22[l]),nrow=2,ncol=2,byrow=T)

                                     }
                                  }
#####################################################################
#####################################################################
CV1=function(x1,x2,V1r,V2r,Hrp,w)
{b1<-1
 ngr<-length(V1r)
  U<-array(0,dim=c(2,2,length(Hrp[1,1,,1]),length(Hrp[2,2,1,])))
   K<-array(0,dim=c(length(x1),length(x2),length(V1r)))
   K1<-array(0,dim=c(length(x2),length(V2r)))
    for ( k in 1 : length(Hrp[1,1,,1])){   #vecteur des h11
      for ( l in 1 : length(Hrp[2,2,1,])){   #vecteur des h22

       for (i in 1 :length(x1)) # boucle en i pour chaque point x[i]
for (m in 1 :length(x2)) # boucle en i pour chaque point x[i]
         {for (j in 1 :length(V1r)) # boucle en j pour chaque observation V[j]
{if (V2r[j]<=(x2[m]+1))#

               { K[i,m,j]<-(1/ngr)*dbeta(V1r[j],(x1[i]/Hrp[1,1,k,1])+1,
((1-x1[i])/Hrp[1,1,k,1])+1)*(choose(x2[m]+1,V2r[j])*((x2[m]+Hrp[2,2,1,l])
/(x2[m]+1))^(V2r[j])*((1-Hrp[2,2,1,l])/(x2[m]+1))^(x2[m]+1-V2r[j]))

                  }
```

```
   }

}
G<-apply(K,c(1,2), sum)
G2<-
G1<-apply(G,1, sum)
                  U[,,k,l] <- G

 }
 }
  return(U)
}
################################################################################
CV2=function(V1r,V2r,Hrp)
{b1<-3
 ngr<-length(V1r)
  U<-array(0,dim=c(2,2,length(Hrp[1,1,,1]),length(Hrp[2,2,,1])))
   K<-array(0,dim=c(length(V1r),length(V1r)))
       for ( k in 1 : length(Hrp[1,1,,1])){   #vecteur des h11
      for ( l in 1 : length(Hrp[2,2,1,])){   #vecteur des h22

       for (i in 1 : length(V1r)) # boucle en i pour chaque point x[i]

         {for (j in 1 :length(V1r)) # boucle en j pour chaque observation V[j]

             {if ((V2r[j]<=(V2r[i]+1)))#
               {
K[i,j]<-2*(1/ngr)*(1/(ngr-1))*dbeta(V1r[j],(V1r[i]/Hrp[1,1,k,1])+1,
((1-V1r[i])/Hrp[1,1,k,1])+1)*(choose(V2r[i]+1,V2r[j])*((V2r[i]+Hrp[2,2,1,l])
/(V2r[i]+1))^(V2r[j])*((1-Hrp[2,2,1,l])/(V2r[i]+1))^(V2r[i]+1-V2r[j]))
 K[i,i]<-0
                }
}
           }

G<-apply(K,1, sum)
                W <- sum(G)
   U[,,k,l] <- W

  }
  }
return(U)
}
#####################################################
CV= CV1(x1,x2,V1r,V2r)-CV2(V1r,V2r,Hrp)
Hcv<-array(0,dim=c(2,2))
Hcv[,]<-matrix(Hrp[which(CV[,,] == min(CV[,,]))]
)
return(Hcv)
}
```

```
##############################Generer les données###################
####################################################################
a1=1, b1=6, a2=10, b2=2, a3=6, b3=6
####
 Nsim=1
n=200
#####################données test#####################################
Ysim=array(0,dim=c(n,5,Nsim))
 for ( k in 1 : Nsim){
###set.seed(10)
 Ysim[,1,k]=rpois(n,a1)
 Ysim[,2,k]=rpois(n,a2)
 Ysim[,3,k]=rbeta(n,a3,b3)
 Ysim[,4,k]=rbeta(n,a1+1.5,b1)
 Ysim[,5,k]=rbeta(n,a2+1.5,b2)
                      }
Qr<-sample(Ysim[,1,k],86)
Qs<-sample(Ysim[,2,k],114)
Vrsim<-c(Qr,Qs)
Vssim=Ysim[,3,]
y<-cbind(Vssim,Vrsim)
####################################################################
##############################Données d'apprentissage'################
####################################################################
Nsim=1, n=80
 Xsim=array(0,dim=c(n,5,Nsim))
 for ( k in 1 : Nsim){
##set.seed(10)
 Xsim[,1,k]=rpois(n,a1)
 Xsim[,2,k]=rpois(n,a2)
 Xsim[,3,k]=rbeta(n,a3,b3)
 Xsim[,4,k]=rbeta(n,a1+1.5,b1)
 Xsim[,5,k]=rbeta(n,a2+1.5,b2)
                      }
#####################densité bimodale########
Q1<-sample(Xsim[,1,k],round(3/7*80))
Q2<-sample(Xsim[,2,k],round(4/7*80))
V2sim<-c(Q1,Q2)
V1sim=Xsim[,3,]
V1r=V1sim
V2r=V2sim
d=cbind(V1r,V2r)

V<-x
V1r<-x[,1]
V2r<-x[,2]
####################################################################
a=0
b=1
c=0
```

```
d=1
############borne de l'intégrale'
################################################################"
Nx=27
Ny=27    #Nx impair car le nombre n (Nx-1) final doit être toujours pair

### Construct Simpson nodes and weights over x ###
h = (b - a)/(Nx-1);
 x1 = seq(a,b,h);
w = rep(1,Nx);  w[seq(2,Nx-1,2)] = 4; w[seq(3,Nx-2,2)] = 2; w = w*h/3;
#########################################################
#########################################################################
#########################################################################
H1kda.diag <- function(x, x.group,H11)
{
  d <- ncol(x)
  gr <- sort(unique(x.group))
  m <- length(gr)
  HH<-array(0,c(4,2,dim(H11)[1]))##

  Hs <- numeric(0)
for (j in 1:dim(H11)[1]){
  for (i in 1:m)
  {
    y1 <- x[x.group==gr[i],]
V1r<-y1[,1]
V2r<-y1[,2]
h11<-H11[j,]
     H1 <-Hlscv4(x1,x2,V1r,V2r,h11,h22,w)
    Hs <- rbind(Hs, H1)
  }
}
  return(Hs)
}
H<-H1kda.diag(x, x.group,H11)
#########################################################################
HH<-array(0,c(4,2,dim(H11)[1]))
for (i in 1:dim(H11)[1]){
HH[,,i]<-array(H[(4*(i-1)+1):(4*i),])}
############Estimation des densités pour les j groupes###################
#########################################################################
#####################################################################
estim<-function(xev,x,Hs)

   { A=0
   B=0
   m=length(xev[,1])
   K<-array(0,dim=c(length(xev[,1]),length(x[,1])))
#############
#############
```

```
         for (i in 1 :length(xev[,1])) # boucle en i pour chaque point x[i]

      {for (j in 1:length(x[,1]))

{if (x[j,2]<=(xev[i,2]+1))

                  {K[i,j]<-(1/length(x[,1]))*dbeta(x[j,1],(xev[i,1]/Hs[1,1])+1,
((1-xev[i,1])/Hs[1,1])+1)*(choose(xev[i,2]+1,x[j,2])*((xev[i,2]+Hs[2,2])/
(xev[i,2]+1))^(x[j,2])*((1-Hs[2,2])/(xev[i,2]+1))^(xev[i,2]+1-x[j,2]))

                   }
                }
}
G <- apply(K,c(1),sum)

return(G)
}
##############################################################################
xev<-y
##############################################################################
kernelestim.points <- function(xev, x,Hs, eval.points)
{
  n <- nrow(x)
  fhat <- estim(xev,x,Hs)
  return(list(x=x, eval.points=xev, estimate=fhat, Hs=Hs, gridded=FALSE))
}
##############################################################################
##############################################################################

##############################################################################
kernelestim <- function(xev, Hs, h,x, gridsize, gridtype, xmin, xmax, supp=3.7,
 eval.points, binned=FALSE, bgridsize, positive=FALSE, adj.positive,
 w, compute.cont=FALSE, approx.cont=TRUE, unit.interval=FALSE, verbose=FALSE)
{
  d <- ncol(x); n <- nrow(x)
  if (missing(w)) w <- rep(1,n)
       if (is.data.frame(x)) x <- as.matrix(x)
         fhat <- kdea.points(xev=xev, x=x,Hs=Hs ,eval.points=eval.points)
  fhat$w <- w
  class(fhat) <- "akde"

  ## compute prob contour levels
  if (compute.cont & missing(eval.points))
    fhat$cont <- contourLevels(fhat, cont=1:99, approx=approx.cont)

  return(fhat)
 }
 ##############################################################################
##############################################################################
kerneldiscrim <- function(x, x.group, Hs, hs, prior.prob=NULL, gridsize, xmin,
```

```
xmax, supp=3.7, eval.points, binned=FALSE, bgridsize, w, compute.cont=FALSE,
 approx.cont=TRUE, kde.flag=TRUE)
{
  if (missing(eval.points)) eval.points <- x
  gr <- sort(unique(x.group))
  m <- length(gr)

      fhat.list <- kdaa.nd(x=x, x.group=x.group, Hs=Hs, prior.prob=prior.prob,
  gridsize=gridsize, supp=supp, xmin=xmin, xmax=xmax, compute.cont=compute.cont,
   approx.cont=approx.cont)

    fhat <- kdaa.nd(x=x, x.group=x.group, Hs=Hs, prior.prob=prior.prob,
    gridsize=gridsize, supp=supp,   xmin=xmin, xmax=xmax, eval.points=xev,
     compute.cont=compute.cont, approx.cont=approx.cont)
    fhat.wt <- matrix(0, ncol=m, nrow=nrow(xev))
  for (j in 1:m)
    fhat.wt[,j] <- fhat$estimate[[j]]* fhat$prior.prob[j]

  ## Assigner y suivant la plus grande densité poondérée
  disc.gr.temp <- apply(fhat.wt, 1, which.max)

  disc.gr <- gr
  for (j in 1:m)
  {
    ind <- which(disc.gr.temp==j)
    disc.gr[ind] <- gr[j]
  }
   fhat.list$x.group.estimate <- disc.gr
    return(fhat.list)
}
###############################################################################
#########################################################################
kerneldiscrim.nd <- function(x, x.group, Hs, prior.prob, gridsize, supp, eval.points,
  bgridsize, xmin, xmax, w, compute.cont, approx.cont)
{
  if (is.data.frame(x)) x <- as.matrix(x)
  gr <- sort(unique(x.group))
  m <- length(gr)
  d <- ncol(x)

  if (missing(w)) w <- rep(1, nrow(x))

  fhat.list <- list()
  for (j in 1:m)
  {
    xx <- x[x.group==gr[j],]
    ww <- w[x.group==gr[j]]
   nn<-dim(x[x.group==gr[j],])[1] #longeur de chaque groupe j

    H <- Hs[((j-1)*d+1) : (j*d),]
```

```
    fhat.temp <- kdea(xev=xev, H=H, x=xx,eval.points=xev,w=ww)

    fhat.list$estimate <- c(fhat.list$estimate, list(fhat.temp$estimate))
    fhat.list$eval.points <- fhat.temp$eval.points
    fhat.list$x <- c(fhat.list$x, list(xx))
    fhat.list$H <- c(fhat.list$H, list(H))
    fhat.list$w <- c(fhat.list$w, list(ww))
  }
  pr <- rep(0, length(gr))
  for (j in 1:length(gr)) pr[j] <- length(which(x.group==gr[j]))
  pr <- pr/nrow(x)
  fhat.list$prior.prob <- pr
  fhat.list$x.group <- x.group

  class(fhat.list) <- "kdaa"
  return (fhat.list)
}
###############################################################################
# Comparer la vraie variable de classification à celle estimer
#
# Parametres
# group - variable de classement
# est.group - estimation de la variable de classement
#
# sorties
# error - taux d'erreur de classification
###############################################################################
compare <- function(x.group, est.group, by.group=FALSE)
{
  if (length(x.group)!=length(est.group))
    stop("group label vectors not the same length")

  grlab <- sort(unique(x.group))
  m <- length(grlab)
  comp <- matrix(0, nrow=m, ncol=m)

  for (i in 1:m)
    for (j in 1:m)
      comp[i,j] <- sum((x.group==grlab[i]) & (est.group==grlab[j]))

  if (by.group)
  {
    er <- vector()
    for (i in 1:m)
      er[i] <- 1-comp[i,i]/rowSums(comp)[i]
    er <- matrix(er, ncol=1)
    er <- rbind(er, 1 - sum(diag(comp))/sum(comp))
    rownames(er) <- c(as.character(paste(grlab, "(true)")), "Total")
    colnames(er) <- "error"
```

```
  }
  else
    er <- 1 - sum(diag(comp))/sum(comp)

  comp <- cbind(comp, rowSums(comp))
  comp <- rbind(comp, colSums(comp))

  colnames(comp) <- c(as.character(paste(grlab, "(est.)")), "Total")
  rownames(comp) <- c(as.character(paste(grlab, "(true)")), "Total")

  TN <- comp[1,1]; FP <- comp[1,2]; FN <- comp[2,1]; TP <- comp[2,2]
  spec <- 1-(FP/(FP+TN))
  sens <- 1-(FN/(TP+FN))

  return(list(cross=comp, error=er, TP=TP, FP=FP, FN=FN, TN=TN, spec=spec, sens=sens))

}
#########################################################################
###########Valeur des matrices de lissages par validation croisée profilée########
Profile<-rep(0,dim(H11)[1])
for (j in 1:dim(H11)[1]){

fhatt<-kdaa(x, x.group, Hs=HH[,,j],eval.points=xev)

est.group<-fhatt$x.group.estimate
Profile[j]<-compare(y.group, est.group, by.group=FALSE)$error
}
A1<-min(Profile)
H1<-HH[,,which.min(Profile)]
#########################################################################################
```

A3. Regression multiple par noyaux associés (Chapitre 4)

La procédure pour la régression multiple est donnée ici en dimension avec le noyau multiple composée de deux bêta et deux triangulaires discrets. Pour les autres cas, il suffit de changer de noyau associé multivarié.

```
#####################################Regression multiple#########################
#############################################################################
#Validation croisée pour regression à quatre variables
#############################################################################
#########################################################################
 hcvBetaBinTrian1<-function(a1,V1,V2,V3,V4,h11,h22,h33,h44,y)
   {
n<-length(V1)
  U<-array(0,dim=c(length(h11)))
Kern<-array(0,dim=c(length(V1),length(V1)))
```

```
K<-array(0,dim=c(length(V1),length(V1)))
    for ( r in 1 : length(h11)){
              for (i in 1 :length(V1)) # boucle en i pour chaque point x[i]
                {for (j in 1 :length(V1)) # boucle en j pour chaque observation V[j]

                      {if ( (
                       (V4[j]>=(V4[i]-a1) & V4[j]<=(V4[i]+a1))) # Support
                        Kern[i,j]<-dbeta(V1[j],(V1[i]/h11[r])+1,((1-V1[i])/h11[r])+1)
                          *dbeta(V2[j],(V2[i]/h22[r])+1,((1-V2[i])/h22[r])+1)*
                          (((a1+1)^h33[r] - (abs(V3[j]-V3[i]))^h33[r]))*
                          (((a1+1)^h44[r] - (abs(V4[j]-V4[i]))^h44[r]))

K[i,j]<-y[j]*Kern[i,j]

                      }
                }
diag(Kern)<-0
diag(K)<-0
G<-apply(K,1, sum)

E<-apply(Kern,1, sum)

F<-(1/n)*sum((y-(G/E))^2)
  U[r] <- F #vecteur de tous ISE selon les simulations
        }
   return(U)
 }
 ##################################################
##################################################
 y1<-y
h1<-h11[which.min(dens3d$hcv)]
 h2<-h22[which.min(dens3d$hcv)]
  h3<-h33[which.min(dens3d$hcv)]
    h4<-h44[which.min(dens3d$hcv)]
#############################################################################
R2<-function(a1,V1,V2,V3,V4,h1,h2,h3,h4,y1)
{
  s=rep(0,length(V1))
  S=rep(0,length(V1))

   for (i in 1 :length(V1)) # boucle en i pour chaque observation x[i]
     {for (j in 1:length(V1)) # boucle en j pour chaque observation V[j]

{if ( (V3[j]>=(V3[i]-a1) & V3[j]<=(V3[i]+a1))
 & (V4[j]>=(V4[i]-a1) & V4[j]<=(V4[i]+a1)))#

                 { K <- dbeta(V1[j],(V1[i]/h1)+1,((1-V1[i])/h1)+1)
                 *dbeta(V2[j],(V2[i]/h2)+1,((1-V2[i])/h2)+1)
                 *(((a1+1)^h3 - (abs(V3[j]-V3[i]))^h3))
                 *(((a1+1)^h4 - (abs(V4[j]-V4[i]))^h4))
```

```
                  A<-K
                 B<-K*y1[j]
                 s[i]<-s[i]+A
                 S[i]<-S[i]+B
                   } }


      F<-S/s
    }
     moyY<-sum(y1)/length(V1)
    R<-sum(((F-moyY)^2))/sum(((y1-moyY)^2)) # Coefficient de détermination
 return(R)
}
###############
Rsquare<-R2(a1,V1,V2,V3,V4,h1,h2,h3,h4,y1)
################Average Squared Error#####
ASE<-function(a1,V1,V2,V3,V4,h1,h2,h3,h4,y1)
{
  s=rep(0,length(V1))
  S=rep(0,length(V1))
      for (i in 1 :length(V1)) # boucle en i pour chaque observation x[i]
     {for (j in 1:length(V1))

{if ((V3[j]>=(V3[i]-a1) & V3[j]<=(V3[i]+a1))
 & (V4[j]>=(V4[i]-a1) & V4[j]<=(V4[i]+a1)))
              K <- dbeta(V1[j],(V1[i]/h1)+1,((1-V1[i])/h1)+1)
                  *dbeta(V2[j],(V2[i]/h2)+1,((1-V2[i])/h2)+1)*
                  (((a1+1)^h3 - (abs(V3[j]-V3[i]))^h3))*
                  (((a1+1)^h4 - (abs(V4[j]-V4[i]))^h4))

                   A<-K
                  B<-K*y1[j]
                  s[i]<-s[i]+A
                  S[i]<-S[i]+B
                    }

            F<-S/s
    }

     moyY<-sum(y1)/length(V1)
       T<-sum(((F-y1)^2))/length(V1) # Coefficient de détermination
 return(T)
}
###############
ASE(a1,V1,V2,V3,V4,h1,h2,h3,h4,y1)
#########################################################
```

A4. R manual of Ake : Associated kernel estimations (Chapitre 5)

Le manuel suivant présente les différentes routines, sous le logiciel R, relatives au Chapitre 5 du package Ake. Ce package est disponible sur le site "Comprehensible R Archive Network" (CRAN) via le lien :
http://cran.r-project.org/web/packages/Ake/Ake.pdf.

Package 'Ake'

March 30, 2015

Encoding UTF-8

Type Package

Title Associated Kernel Estimations

Version 1.0

Date 2015-03-29

Author W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

Maintainer W. E. Wansouwé <ericwansouwe@gmail.com>

Description Continuous and discrete (count or categorical) estimation of density, probability mass function (p.m.f.) and regression functions are performed using associated kernels. The cross-validation technique and the local Bayesian procedure are also implemented for bandwidth selection.

License GPL (>= 2)

LazyLoad yes

URL www.r-project.org

NeedsCompilation no

Repository CRAN

Date/Publication 2015-03-30 00:04:27

R topics documented:

Ake-package *Associated kernel estimations*

Description

Continuous and discrete estimation of density `dke.fun`, probability mass function (p.m.f.) `kpmfe.fun` and regression `reg.fun` functions are performed using continuous and discrete associated kernels. The cross-validation technique `hcvc.fun`, `hcvreg.fun` and the Bayesian procedure `hbay.fun` are also implemented for bandwidth selection.

Details

The estimated density or p.m.f: The associated kernel estimator $\widehat{f}_n$ of f is defined as

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i),$$

where $K_{x,h}$ is one of the kernels `kef` defined below. In practice, we first calculate the global normalizing constant

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

where T is the support of the density or p.m.f. function and ν is the Lebesgue or count measure on T. For both continuous and discrete associated kernels, this normalizing constant is not generally equal to 1 and it will be computed. The represented density or p.m.f. estimate is then $\tilde{f}_n = \widehat{f}_n/C_n$.

For **discrete data**, the integrated squared error (ISE) defined by

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

is the criteria used to measure the smoothness of the associated kernel estimator $\tilde{f}_n$ with the empirical p.m.f. f_0; see Kokonendji and Senga Kiessé (2011).

The estimated regressor: Both in continuous and discrete cases, considering the relation between a response variable y and an explanatory variable x given by

$$y = m(x) + \epsilon,$$

where m is an unknown regession function on T and ϵ the disturbance term with null mean and finite variance. Let $(x_1, y_1), \ldots, (x_n, y_n)$ be a sequence of independent and identically distributed (iid) random vectors on $T \times R$ with $m(x) = E(y|x)$. The well-known Nadaraya-Watson estimator using associated kernels is $\widehat{m}_n$ defined as

$$\widehat{m}_n(x) = \sum_{i=1}^{n} \omega_x(X_i) Y_i,$$

where $\omega_x(X_i) = K_{x,h}(X_i) / \sum_{i=1}^{n} K_{x,h}(X_i)$ and $K_{x,h}$ is one of the associated kernels defined below.

Beside the criterion of kernel support, we retain the root mean squared error (RMSE) and also the practical coefficient of determination R^2 defined respectively by

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

and

$$R^2 = \frac{\sum_{i=1}^{n} \{\widehat{m}_n(x_i) - \bar{y}\}^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},$$

where $\bar{y} = n^{-1}(y_1 + \ldots + y_n)$; see Kokonendji et al. (2009).

Given a data sample, the package allows to compute the density or p.m.f. and regression functions using one of the seven associated kernels: extended beta, lognormal, gamma, reciprocal inverse Gaussian for continuous data, DiracDU for categorical data, and binomial and discrete triangular for count data. The bandwidth parameter is computed using the cross-validation technique. When the associated kernel function is binomial, the bandwidth parameter is also computed using the local Bayesian procedure. The associated kernel functions are defined below. The first four kernels are for continuous data and the last three kernels are for discrete case.

Extended beta kernel: The extended beta kernel is defined on $S_{x,h,a,b} = [a, b] = T$ with $a < b < \infty$, $x \in T$ and $h > 0$:

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

where $B(r, s) = \int_0^1 t^{r-1}(1-t)^{s-1} dt$ is the usual beta function with $r > 0$, $s > 0$ and 1_A denotes the indicator function of A. For $a = 0$ and $b = 1$, it corresponds to the beta kernel which is the probability density function of the beta distribution with shape parameters $1+x/h$ and $(1-x)/h$; see Libengué (2013).

Gamma kernel: The gamma kernel is defined on $S_{x,h} = [0, \infty) = T$ with $x \in T$ and $h > 0$ by

$$GA_{x,h}(y) = \frac{y^{x/h}}{\Gamma(1 + x/h) h^{1+x/h}} exp\left(-\frac{y}{h}\right) 1_{S_{x,h}}(y),$$

where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ is the classical gamma function. The probability density function $GA_{x,h}$ is the gamma distribution with scale parameter $1 + x/h$ and shape parameter h; see Chen (2000).

Lognormal kernel: The lognormal kernel is defined on $S_{x,h} = [0, \infty) = T$ with $x \in T$ and $h > 0$ by

$$LN_{x,h}(y) = \frac{1}{yh\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{1}{h}log(\frac{y}{x}) - h\right)^2\right\} 1_{S_{x,h}}(y).$$

It is the probability density function of the classical lognormal distribution with parameters $log(x) + h^2$ and h; see Libengué (2013).

Binomial kernel: Let $x \in N := \{0, 1, \ldots\}$ and $S_x = \{0, 1, \ldots, x + 1\}$. The Binomial kernel is defined on the support S_x by

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

where $h \in (0, 1]$. Note that $B_{x,h}$ is the p.m.f. of the binomial distribution with its number of trials $x + 1$ and its success probability $(x + h)/(x + 1)$; see Kokonendji and Senga Kiessé (2011).

Discrete triangular kernel: For fixed arm $a \in N$, we define $S_{x,a} = \{x - a, \ldots, x, \ldots, x + a\}$. The discrete triangular kernel is defined on $S_{x,a}$ by

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

where $x \in N$, $h > 0$ and $P(a, h) = (2a+1)(a+1)^h - 2(1+2^h+\cdots+a^h)$ is the normalizing constant. For $a = 0$, the Discrete Triangular kernel $DT_{x,h;0}$ corresponds to the Dirac kernel on x; see Kokonendji et al. (2007), and also Kokonendji and Zocchi (2010) for an asymmetric version of discrete triangular.

DiracDU kernel: For fixed number of categories $c \in \{2, 3, ...\}$, we define $S_c = \{0, 1, \ldots, c - 1\}$. The DiracDU kernel is defined on S_c by

$$DU_{x,h;c}(y) = (1-h)1_{\{x\}}(y) + \frac{h}{c-1} 1_{S_c \setminus \{x\}}(y),$$

where $x \in S_c$ and $h \in (0, 1]$. See Kokonendji and Senga Kiessé (2011), and also Aitchison and Aitken (1976) for multivariate case.

Note that the global normalizing constant is 1 for DiracDU.

The bandwidth selection: Two functions are implemented to select the bandwidth: cross-validation and local Bayesian procedure. The cross-validation technique is used for all the associated kernels both in density and regression; see Kokonendji and Senga Kiessé (2011). The local Bayesian procedure is implemented to select the bandwidth in the estimation of p.m.f. when using binomial kernel; see Zougab et al. (2014).

In the coming versions of the package, adaptive Bayesian procedure will be included for bandwidth selection in density estimation when using gamma kernel. A global Bayesian procedure will also be implemented for bandwidth selection in regression when using binomial kernel.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

Maintainer: W. E. Wansouwé <ericwansouwe@gmail.com>

References

Aitchison, J. and Aitken, C.G.G. (1976). Multivariate binary discrimination by the kernel method, *Biometrika* **63**, 413 - 420.

Chen, S. X. (1999). Beta kernels estimators for density functions, *Computational Statistics and Data Analysis* **31**, 131 - 145.

Chen, S. X. (2000). Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics* **52**, 471 - 480.

Igarashi, G. and Kakizawa, Y. (2015). Bias correction for some asymmetric kernel estimators, *Journal of Statistical Planning and Inference* **159**, 37 - 63.

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

Kokonendji, C.C., Senga Kiessé, T. and Demétrio, C.G.B. (2009). Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics* **12**, 99 - 125.

Libengue, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2014). Bayesian approach in nonparametric count regression with binomial kernel, *Communications in Statistics - Simulation and Computation* **43**, 1052 - 1063.

`dke.fun` *Function for density estimation*

Description

The (S3) generic function `dkde.fun` computes the density. Its default method does so with the given kernel and bandwidth h.

Usage

```
dke.fun(Vec, ...)
## Default S3 method:
dke.fun(Vec, h, type_data = c("discrete", "continuous"),
ker = c("BE", "GA", "LN", "RIG"), x = NULL, a0 = 0, a1 = 1, ... )
```

Arguments

`Vec`	The data sample from which the estimate is to be computed.
`h`	The bandwidth or smoothing parameter.
`type_data`	The data sample type. Data can be continuous or discrete (categorical or count). Here, in this function , we deal with continuous data.
`ker`	A character string giving the smoothing kernel to be used which is the associated kernel: "BE" extended beta, "GA" gamma, "LN" lognormal and "RIG" reciprocal inverse Gaussian.

`x`	The points of the grid at which the density is to be estimated.
`a0`	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`a1`	The right bound of the support used for extended beta kernel. Default value is 1 for beta kernel.
`...`	Further arguments.

Details

The associated kernel estimator $\widehat{f_n}$ of f is defined in the above sections. We recall that in general, the sum of the estimated values on the support is not equal to 1. In practice, we compute the global normalizing constant C_n before computing the estimated density $\widetilde{f_n}$; see e.g. Libengué (2013).

Value

Returns a list containing:

`data`	The data - same as input Vec.
`n`	The sample size.
`kernel`	The asssociated kernel used to compute the density estimate.
`h`	The bandwidth used to compute the density estimate.
`eval.points`	The coordinates of the points where the density is estimated.
`est.fn`	The estimated density values.
`C_n`	The global normalizing constant.
`hist`	The histogram corresponding to the observations.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Examples

```
## A sample data with n=100.
V<-rgamma(100,1.5,2.6)
##The bandwidth can be the one obtained by cross validation.
h<-0.052
## We choose Gamma kernel.

est<-dke.fun(V,h,"continuous","GA")
```

`hbay.fun` *Local Bayesian procedure for bandwidth selection*

Description

The (S3) generic function `hbay.fun` computes the local Bayesian procedure for bandwidth selection.

Usage

```
hbay.fun(Vec, ...)
## Default S3 method:
hbay.fun(Vec, x = NULL, ...)
```

Arguments

`Vec`	The data sample from which the estimate is to be computed.
`x`	The points of the grid where the density is to be estimated.
`...`	Further arguments for (non-default) methods.

Details

`hbay.fun` implements the choice of the bandwidth h using the local Bayesian approach of a kernel density estimator.

Value

Returns the bandwidth selected using the local Bayesian procedure.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Chen, S. X. (1999). Beta kernels estimators for density functions, *Computational Statistics and Data Analysis* **31**, 131 - 145.

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2014). Bayesian approach in nonparametric count regression with binomial kernel, *Communications in Statistics - Simulation and Computation* **43**, 1052 - 1063.

`hcvc.fun` *Cross-validation function for bandwidth selection for continuous data*

Description

The (S3) generic function `hcvc.fun` computes the cross-validation bandwidth selector.

Usage

```
hcvc.fun(Vec,...)
## Default S3 method:
hcvc.fun(Vec, bw = NULL, type_data, ker, a0 = 0, a1 = 1, ...)
```

Arguments

`Vec`	The data sample from which the estimate is to be computed.
`bw`	The sequence of bandwidths where to compute the cross-validation. Default value is NULL.
`type_data`	The sample data type.
`ker`	The associated kernel.
`a0`	The left bound of the extended beta. Default value is 0.
`a1`	The right bound of the extended beta.Default value is 1.
`...`	Further arguments.

Details

`hcvc.fun` implements the choice of the bandwidth h using the cross-validation approach of a kernel density estimator.

Value

Returns a list containing:

`hcv`	value of bandwidth parameter.
`CV`	the values of cross-validation function.
`seq_h`	the sequence of bandwidths where the cross validation is computed.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Chen, S. X. (1999). Beta kernels estimators for density functions, *Computational Statistics and Data Analysis* **31**, 131 - 145.

Chen, S. X. (2000). Gamma kernels estimators for density functions, *Annals of the Institute of Statistical Mathematics* **52**, 471 - 480.

Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Igarashi, G. and Kakizawa, Y. (2015). Bias correction for some asymmetric kernel estimators, *Journal of Statistical Planning and Inference* **159**, 37 - 63.

Examples

```
V=rgamma(100,1.5,2.6)
## Not run:
hcvc.fun(V,NULL,"continuous","GA")

## End(Not run)
```

`hcvd.fun` *Cross-validation function for bandwidth selection in p.m.f. estimation*

Description

The (S3) generic function `hcvd.fun` computes the cross-validation bandwidth selector in p.m.f. estimation.

Usage

```
hcvd.fun(Vec, ...)
## Default S3 method:
hcvd.fun(Vec, seq_bws = NULL, ker = c("bino", "triang", "dirDU"), a = 1, c = 2,...)
```

Arguments

`Vec`	The data sample from which the estimate is to be computed.
`seq_bws`	The sequence of bandwidths where to compute the cross-validation. Default value is NULL.
`ker`	The associated kernel
`a`	The arm of the discrete triangular kernel. Default value is 1.
`c`	The number of categories in DiracDU kernel.Default value is 2.
`...`	Further arguments.

Details

The `hcvd.fun` function implements the choice of the bandwidth h using the cross-validation approach in p.m.f. estimate.

Value

Returns a list containing:

`hcv`	The optimal bandwidth parameter.
`CV`	The cross-validation function values.
`seq_h`	The sequence of bandwidths where the cross-validation is computed.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Chen, S. X. (1999). Beta kernels estimators for density functions, *Computational Statistics and Data Analysis* **31**, 131 - 145.

Chen, S. X. (2000). Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics* **52**, 471 - 480.

Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Igarashi, G. and Kakizawa, Y. (2015). Bias correction for some asymmetric kernel estimators, *Journal of Statistical Planning and Inference* **159**, 37 - 63.

Examples

```
## Data can be simulated data or real data
## We use real data
## and then compute the cross validation.
Vec<-c(10,0,1,0,4,0,6,0,0,0,1,1,1,2,4,4,5,6,6,6,6,7,1,7,0,7,7,
7,8,0,8,12,8,8,9,9,0,9,9,10,10,10,10,0,10,10,11,12,12,10,12,12,
13,14,15,16,16,17,0,12)
## Not run:
CV<-hcvd.fun(Vec,NULL,"bino")
CV$hcv

## End(Not run)
##The cross validation function can be also ploted.
## Not run:
plot.fun(CV$seq_bws,CV$CV, type="l")

## End(Not run)
```

`hcvreg.fun` *Cross-validation function for bandwidth selection in regresssion*

Description

The (S3) generic function `hcvreg.fun` computes the bandwidth by cross-validation for the regression. Its default method does so. It allows to compute the optimal bandwidth using the cross-validation method. The associated kernels available are: "BE" extended beta, "GA" gamma, "LN" lognormal and "RIG" reciprocal inverse Gaussian, DiracDU, binomial and discrete triangular; see Kokonendji and Senga Kiessé (2011), and also Kokonendji et al. (2009).

Usage

```
hcvreg.fun(Vec, ...)
## Default S3 method:
hcvreg.fun(Vec, y, type_data = c("discrete", "continuous"),
ker = c("bino", "triang", "dirDU", "BE", "GA", "LN", "RIG"),
 h = NULL, a0 = 0, a1 = 1, a = 1, c = 2, ...)
```

Arguments

`Vec`	The explanatory variable.
`y`	The response variable.
`type_data`	The data sample type. Data can be continuous or discrete.
`ker`	A character string giving the smoothing kernel to be used which is the associated kernel: "BE" extended beta, "GA" gamma, "LN" lognormal and "RIG" reciprocal inverse Gaussian, "dirDU" DiracDU,"bino" binomial, "triang" discrete triangular.
`h`	The bandwidth or smoothing parameter.the smoothing bandwidth to be used, can also be a character string giving a rule to choose the bandwidth.
`a0`	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`a1`	The right bound of the support used for extended beta kernel. Default value is 1 for beta kernel.
`a`	The arm of the discrete triangular kernel
`c`	The number of categories
`...`	Further arguments

Details

The selection of the bandwidth parameter is always crucial. If the bandwidth is small, we will obtain an undersmoothed estimator, with high variability. On the contrary, if the value is big, the resulting estimator will be very smooth and farther from the function that we are trying to estimate. The cross-validation function defined in the above sections is used to compute the optimal bandwidth for the associated kernels.

Value

Returns a list containing:

`kernel`	The associated kernel used to compute the optimal bandwidth.
`hcv`	The optimal bandwidth parameter obtained by cross-validation.
`CV`	The values of the cross-validation.
`seq_bws`	A sequence of bandwidths where the cross-validation is computed.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

Kokonendji, C.C., Senga Kiessé, T. and Demétrio, C.G.B. (2009). Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics* **12**, 99 - 125.

Examples

```
## Data can be simulated data or real data
## We use real data
## and then compute the cross validation.
data(milk)
x=milk$week
y=milk$yield
hcvreg.fun(x,y,"discrete",ker="triang",a=1)
```

`kef` *Continuous and discrete associated kernel function*

Description

This function computes the associated kernel function.

Usage

```
kef(x, t, h, type_data = c("discrete", "continuous"),
ker = c("bino", "triang", "dirDU", "BE", "GA", "LN", "RIG"),
a0 = 0, a1 = 1, a = 1, c = 2)
```

Arguments

x	The target.
t	A single value or the grid where the associated kernel function is computed.
h	The bandwidth or smoothing parameter.
type_data	The sample data type
ker	The associated kernel:"bino" Binomial, "triang" discrete triangular kernel, "BE" extended beta, "GA" gamma, "LN" lognormal and "RIG" reciprocal inverse Gaussian,"dirDU" DiracDU.
a0	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
a1	The right bound of the support used for extended beta kernel. Default value is 1 for beta kernel.
a	The arm in discrete triangular kernel. The default value is 1.
c	The number of categories in DiracDU kernel. The default value is 2.

Details

The associated kernel is one of the those which have been defined in the sections above : extended beta, gamma, lognormal, reciprocal inverse Gaussian, DiracDU, binomial and discrete triangular; see Kokonendji and Senga Kiessé (2011), and also Kokonendji et al. (2007).

Value

Returns the value of the associated kernel function at t according to the target and the bandwidth.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Chen, S. X. (1999). Beta kernels estimators for density functions, *Computational Statistics and Data Analysis* **31**, 131 - 145.

Chen, S. X. (2000). Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics* **52**, 471 - 480.

Igarashi, G. and Kakizawa, Y. (2015). Bias correction for some asymmetric kernel estimators, *Journal of Statistical Planning and Inference* **159**, 37 - 63.

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

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

Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*, Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France and Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.

Examples

```
x<-5
h<-0.2
t<-0:10
kef(x,t,h,"discrete","bino")
```

`kern.fun` *The associated kernel function*

Description

The (S3) generic function `kern.fun` computes the value of the associated kernel function. Its default method does so with a given kernel and bandwidth h.

Usage

```
kern.fun(x, ...)
## Default S3 method:
kern.fun(x, t, h, type_data = c("discrete", "continuous"),
 ker = c("bino", "triang", "dirDU", "BE", "GA", "LN", "RIG"),
a0 = 0, a1 = 1, a = 1, c = 2, ...)
```

Arguments

`x`	The target
`t`	A single value or the grid where the discrete associated kernel function is computed.
`h`	The bandwidth or smoothing parameter.
`type_data`	The sample data type
`ker`	The associated kernel: "dirDU" DiracDU,"bino" Binomial, "triang" Discrete Triangular kernel, "BE" extended beta, "GA" gamma, "LN" lognormal and "RIG" reciprocal inverse Gaussian.
`a0`	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`a1`	The right bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`a`	The arm in Discrete Triangular kernel. The default value is 1.
`c`	The number of categories in DiracDU kernel. The default value is 2.
`...`	Further arguments

Details

The associated kernel is one of the those which have been defined in the sections above : extended beta, gamma,lognormal, reciprocal inverse Gaussian, DiracDU, Binomial and Discrete Triangular; see Kokonendji and Senga Kiessé (2011), and also Kokonendji et al. (2007).

Value

Returns the value of the discrete associated kernel function at `t` according to the target and the bandwidth.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

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

Examples

```
x<-5
h<-0.2
t<-0:10
kern.fun(x,t,h,"discrete","bino")
```

`kpmfe.fun` *Function for associated kernel estimation of p.m.f.*

Description

The function estimates the p.m.f. in a single value or in a grid using discrete associated kernels. Three different associated kernels are available: DiracDU (for categorical data), binomial and discrete triangular (for count data).

Usage

```
kpmfe.fun(Vec,...)
## Default S3 method:
kpmfe.fun(Vec, h, type_data = c("discrete", "continuous"),
        ker = c("bino", "triang", "dirDU"), x = NULL, a = 1, c = 2, ...)
```

Arguments

`Vec`	the data sample from which the estimate is to be computed.
`h`	The bandwidth or smoothing parameter. The smoothing bandwidth to be used, can also be a character string giving a rule to choose the bandwidth.
`type_data`	The data sample type. Data type is "discrete" (categorical or count).
`ker`	The associated kernel: "dirDU" DiracDU,"bino" binomial, "triang" discrete triangular.

x	The points of the grid at which the density is to be estimated.
a	The arm in discrete triangular kernel. The default value is 1.
c	The number of categories in DiracDU. The default value is 2.
...	Further arguments.

Details

The associated kernel estimator $\widehat{f_n}$ of f is defined in the above sections. We recall that in general, the sum of the estimated values on the support is not equal to 1. In practice, we compute the global normalizing constant C_n before computing the estimated p.m.f. $\widetilde{f_n}$; see Kokonendji and Senga Kiessé (2011).

The bandwidth parameter in the function is obtained using the cross-validation technique for the three associated kernels. For binomial kernel, the local Bayesian approach is also implemented and is recommanded to select the bandwidth; see Zougab et al. (2012).

Value

Returns a list containing:

data	The number of observations.
n	The number of observations.
eval.points	The support of the estimated p.m.f.
h	The bandwidth
C_n	The global normalizing constant.
ISE_0	The integrated square error.
f_0	A vector of (x,f0(x)).
f_n	A vector of (x,fn(x)).
f0	The empirical p.m.f.
est.fn	The estimated p.m.f. containing estimated values after normalization.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

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

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics* **24**, 783 - 795.

Examples

```
## A sample data with n=60.
V<-c(10,0,1,0,4,0,6,0,0,0,1,1,1,2,4,4,5,6,6,6,6,7,1,7,0,7,7,
7,8,0,8,12,8,8,9,9,0,9,9,10,10,10,10,0,10,10,11,12,12,10,12,12,
13,14,15,16,16,17,0,12)


##The bandwidth can be the one obtained by cross validation.
h<-0.081
## We choose Binomial kernel.

est<-kpmfe.fun(Vec=V,h,"discrete","bino")
##To obtain the normalizing constant:
est
```

`milk` *Average daily fat yields.*

Description

This data is the average daily fat yields (kg/day) of milk from a single cow for each of 35 weeks; see Kokonendji et al. (2009).

Usage

```
data(milk)
```

Format

A data frame with 35 observations on the following 2 variables.

`week` Number of the week

`yield` The yield quantity

Source

McCulloch, C.E. (2001). An Introduction to Generalized Linear Mixed Models, 46a Reuniao Anual da RBRAS - 9o SEAGRO, University of Sao Paulo - ESALQ, Piracicaba.

References

Kokonendji, C.C., Senga Kiessé, T. and Demétrio, C.G.B. (2009). Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics* **12**, 99 - 125.

Examples

```
data(milk)
```

`plot.dke.fun` *Plot of density function*

Description

The `plot.dke.fun` is to plot the associated kernel density estimation.

Usage

```
## S3 method for class 'dke.fun'
plot(x,main = NULL, sub = NULL, xlab = NULL,
ylab = NULL, type = "l", las = 1, lwd = 1, col = "blue", lty = 1, ...)
```

Arguments

`x`	An object class `dke.fun`
`main`	The main parameter
`sub`	The sub title
`xlab,ylab`	The axis label
`type`	the type parameter
`las`	Numeric in 0,1,2,3; the style of axis labels.
`lwd`	The line width, a positive number, defaulting to 1.
`col`	A specification for the default plotting color.
`lty`	The line type.
`...`	Futher arguments

Value

Plot of associated kernel density function is sent to graphics window.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

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

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics* **24**, 783 - 795.

`plot.hcvc.fun` *Plot of cross-validation function for bandwidth selection in density or p.m.f. estimation.*

Description

The functions allows to plot the cross-validation both in discrete `plot.hcvd.fun` and continuous `plot.hcvc.fun` cases.

Usage

```
## S3 method for class 'hcvc.fun'
plot(x,  ...)
## S3 method for class 'hcvd.fun'
plot(x,  ...)
```

Arguments

`x`	an object
`...`	Further arguments

Details

Plot a graphic for cross-validation function

Value

returns a graphics

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

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

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics* **24**, 783 - 795.

`plot.kern.fun` *Plot of associated kernel function*

Description

The `plot.kern.fun` function loops through calls to the `kern.fun` function.

Usage

```
## S3 method for class 'kern.fun'
plot(x, ...)
```

Arguments

`x`	an object of class `kern.fun` (output from `kern.fun`).
`...`	Other graphics parameters

Value

Plot of associated the kernel function is sent to graphics window.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

Kokonendji, C.C., Senga Kiessé, T. and Demétrio, C.G.B. (2009). Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics* **12**, 99 - 125.

`plot.kpmfe.fun` *Plot of the function for associated kernel estimation of the p.m.f.*

Description

The function plots the p.m.f. estimation in a single value or in a grid using discrete associated kernels. Three different associated kernels are available: DiracDU (for categorical data), binomial and discrete triangular (for count data).

Usage

```
## S3 method for class 'kpmfe.fun'
plot(x, ...)
```

Arguments

x	An object of class kpmfe.fun.
...	Further arguments

Details

Plot a graphic

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

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

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics* **24**, 783 - 795.

`plot.reg.fun` *Plot for associated kernel regression*

Description

Plot for associated kernel regression for univariate data. The `plot.reg.fun` function loops through calls to the `reg.fun` function.

Usage

```
## S3 method for class 'reg.fun'
plot(x,  ...)
```

Arguments

x	An object of class reg.fun
...	other graphics parameters

Details

The function allows to plot the regression

Value

Plot is sent to graphics window.

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

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

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics* **24**, 783 - 795.

`print.reg.fun` *Print for regression function*

Description

The function allows to print the result of computation in regression as a data frame.

Usage

```
## S3 method for class 'reg.fun'
print(x, digits = NULL, ...)
```

Arguments

`x`	object of class `reg.fun`.
`digits`	The number of digits
`...`	Further arguments

Details

The associated kernel estimator $\widehat{m}_n$ of m is defined in the above sections; see Kokonendji and Senga Kiessé (2011). The bandwidth parameter in the function is obtained using the cross-validation technique for the associated kernels.

Value

Returns a list containing:

`data`	The explanatory variable, printed as a data frame
`y`	The response variable, printed as a data frame
`n`	The size of the sample
`kernel`	The associated kernel
`h`	The smoothing parameter
`eval.points`	The grid where the regression is computed, printed as data frame
`m_n`	The estimated values, printed as data frame
`Coef_det`	The Coefficient of determination

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

Kokonendji, C.C., Senga Kiessé, T. and Demétrio, C.G.B. (2009). Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics* **12**, 99 - 125.

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2014). Bayesian approach in nonparametric count regression with Binomial Kernel, *Communications in Statistics - Simulation and Computation* **43**, 1052 - 1063.

Examples

```
data(milk)
x=milk$week
y=milk$yield
##The bandwidth is the one obtained by cross validation.
h<-0.10
## We choose binomial kernel.
m_n<-reg.fun(x, y, "discrete",ker="bino", h)
print.reg.fun(m_n)
```

`reg.fun` *Function for associated kernel estimation of regression*

Description

The function estimates the discrete and continuous regression in a single value or in a grid using associated kernels. Different associated kernels are available: extended beta, gamma, lognormal, reciprocal inverse Gaussian (for continuous data), DiracDU (for categorical data), binomial and also discrete triangular (for count data).

Usage

```
reg.fun(Vec, ...)
## Default S3 method:
reg.fun(Vec, y, type_data = c("discrete", "continuous"),
ker = c("bino", "triang", "dirDU", "BE", "GA", "LN", "RIG"),
 h, x = NULL, a0 = 0, a1 = 1, a = 1, c = 2, ...)
```

Arguments

`Vec`	The explanatory variable.
`y`	The response variable.
`type_data`	The sample data type.
`ker`	The associated kernel: "dirDU" DiracDU,"bino" binomial, "triang" discrete triangular, etc.
`h`	The bandwidth or smoothing parameter.
`x`	The single value or the grid where the regression is computed.
`a0`	The left bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`a1`	The right bound of the support used for extended beta kernel. Default value is 0 for beta kernel.
`a`	The arm in Discrete Triangular kernel. The default value is 1.
`c`	The number of categories in DiracDU. The default value is 2.
`...`	Further arguments

Details

The associated kernel estimator $\widehat{m}_n$ of m is defined in the above sections; see also Kokonendji and Senga Kiessé (2011). The bandwidth parameter in the function is obtained using the cross-validation technique for the seven associated kernels. For binomial kernel, the local Bayesian approach is also implemented; see Zougab et al. (2014).

Value

Returns a list containing:

`data`	The data sample, explanatory variable
`y`	The data sample, response variable
`n`	The size of the sample
`kernel`	The asociated kernel
`h`	The bandwidth
`eval.points`	The grid where the regression is computed
`m_n`	The estimated values
`Coef_det`	The coefficient of determination

Author(s)

W. E. Wansouwé, S. M. Somé and C. C. Kokonendji

References

Kokonendji, C.C. and Senga Kiessé, T. (2011). Discrete associated kernel method and extensions, *Statistical Methodology* **8**, 497 - 516.

Kokonendji, C.C., Senga Kiessé, T. and Demétrio, C.G.B. (2009). Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics* **12**, 99 - 125.

Zougab, N., Adjabi, S. and Kokonendji, C.C. (2014). Bayesian approach in nonparametric count regression with binomial kernel, *Communications in Statistics - Simulation and Computation* **43**, 1052 - 1063.

Examples

```
data(milk)
x=milk$week
y=milk$yield
##The bandwidth is the one obtained by cross validation.
h<-0.10
## We choose binomial kernel.
## Not run:
m_n<-reg.fun(x, y, "discrete",ker="bino", h)

## End(Not run)
```

Index