Positive Definite Nonparametric Regression with Application to Covariance Function Estimation

by

Myeong-Jong Kang

A Dissertation
submitted in fulfillment of the requirement
for the degree of
Master of Science
in
Statistics

The Department of Statistics
College of Natural Sciences
Seoul National University
August, 2017

Abstract

Myeong-Jong Kang
The Department of Statistics
The Graduate School
Seoul National University

We propose a new nonparametric regression method subject to the positive definiteness constraint for estimating the covariance function of a stationary stochastic process. Our method ensures that the estimator can satisfy positive definiteness, as well as both isotropy and monotonicity. Also, unlike previous methods, our estimators can be computed automatically with cross validation. To achieve these advantages, we define our estimators by taking the integral transform of kernel distribution estimators and suggest universal estimation algorithm based on the framework of Estimation of Distribution Algorithms (EDAs). A small simulation study is performed that demonstrates the usefulness of our method to estimate the covariance function more robustly than other typical nonparametric regression estimators. An application to the sic.100 dataset is also presented.

Keywords : Nonparametric regression, kernel, positive definiteness, isotropy, monotonicity, covariance, variogram, estimation of distribution algorithm

Student Number : 2013-20209

Contents

Chapter 1

Introduction

Regression analysis is one of the most popular statistical methods used by both statisticians and practitioners. When applying regression techniques, it is commonly supposed that a set of observations $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_n, y_n)$ is observed. Note that $\mathbf{x}_i \in \mathbb{R}^d$, $d \geq 1$, and $y_i \in \mathbb{R}$ are called the explanatory variables and the response variables, respectively. Also, a relation between the explanatory variable and the response variable is modelled with the unknown regression function f, i.e.,

$$y_i = f(\mathbf{x}_i) + \epsilon_i\,, \quad i = 1, 2, \cdots, n \tag{1.1}$$

where ϵ_i's are observation errors. In this context, it is one of the goals to estimate the regression function f under reasonable assumptions.

It is well-known that regression techniques are used in a variety of fields of statistics. Especially in geostatistics, regression techniques are applied to estimate the covariance function or the semivariogram. In this paper, we focus our attention on the estimation of the covariance function, since the estimation of the semivariogram is directly derived from the relaton between the covariance function and the semivariogram. See Cressie (2015) for details.

In some regression problems, it is established that the regression function has special features and that the estimated functions are required to have the same features. In the case of the covariance function, *positive definiteness* has been considered as an essential condition since the covariance function is noteworthy not only in itself but for the foundation of prediction with kriging. Generally, positive definiteness of covariance function estimators is a necessary condition for applying kriging methods to predict values of the stochastic process at unobserved locations. See Cressie (2015) and Gorsich and Genton (2000) for more details.

The most widely used approach which forces estimated functions to have special features is a *parametric* approach. When the preselected parametric model is correct, there is a wide array of advantages, including nice efficiency properties of the corresponding inferences and possibilities of the interpretation in parameters (Eubank, 1999). On the other hand, when there is no information about the form of the regression function, it is possible that an irrelevant parametric model is preselected. In this case, most of the related inferences are spurious. The failure of the parametric approach usually arises when the parametric model is so restricted that the estimated functions cannot describe the relationship between the explanatory variable and the response variable enough.

A *Nonparametric* approach can be one of the solutions to overcome the inherent weakness with parametric approach. In general, the nonparametric approach only assumes that the regression function belongs to some infinite dimensional collection of functions (Eubank, 1999). It offers flexible tools in estimating the regression function without reference to parametric regression models fully described by a finite set of parameters (Härdle, 1990). However, typical nonparametric regression estimators cannot be forced to have special

features. Therefore, there have been numerous studies on nonparametric regression analysis under shape constraints such as monotonicity and convexity.

Unfortunately, positive definiteness have not received abundant attention compared to other shape constraints in regression research. On the other hand, there have been several related studies over the last four decades in the field of spatial statistics. After Armstrong and Jabin (1981) stresses the importance of using only positive definite functions as models for the covariance function, Christakos (1984) proposes several criteria that permit users to confirm whether a function which is a candidate for modelling the covariance function is positive definite. Although Christakos (1984) discusses only parametric estimation of the covariance function, this paper digests necessary and sufficient conditions of "permissible" models of the covariance function to suggest the criterions. Influenced by the ideas of the studies introduced in earlier, Shapiro and Botha (1991) proposes a nonparametric method for estimating the isotropic covariance function based on a non-negative regression technique. Meanwhile, Hall et al. (1994) suggests a different nonparametric method using a kernel-type regression technique. Both of these methods have the common mathematical background, Bochner's theorem.

Bochner's theorem is presented in Bochner (1933). This theorem states that a continuous positive definite function can be formulated as the Fourier transform of a finite non-negative Borel measure. Despite the intractability of definition of positive definiteness, Bochner's theorem allows continuous positive definite functions to be treated. Schoenberg (1938) introduces a modification of Bochner's theorem for continuous positive definite isotropic functions. See Cheney and Light (2009) and Fasshauer (2007) for more details.

It seems that the methods proposed by Shapiro and Botha (1991) and Hall et al. (1994) are considered to be major approaches to estimate the covariance

function nonparametrically. There are several studies on this issue based on either Shapiro-Botha method or Hall-Fisher-Hoffman method. Shapiro-Botha method is formally evaluated against typical parametric models by Cherry (1996). García-Soidán et al. (2003) and Garcìa-Soidán et al. (2004) propose nonparametric estimators of the semivariogram based on typical nonparametric regression estimators, like Hall et al. (1994). Besides, to guarantee the positive definiteness when estimating the semivariogram, Shapiro-Botha method is applied to the nonparametric estimators previously obtained. In addition, García-Soidán (2007) discusses the asymptotic properties of the Nadaraya-Watson type estimator of the semivariogram. A numerical study comparing several approaches is provided in Menezes et al. (2005). Gorsich and Genton (2000) proposes the procedure of parametric semivariogram model selection based on the derivative of Shapiro-Botha nonparametric estimator.

It is clear that both of the methods have various merits. Shapiro-Botha method is easily implementable and computationally flexible. Also, one can easily make additional assumptions such as smoothness and monotonicity when using Shapiro-Botha method. Meanwhile, Hall-Fisher-Hoffman method has nice asymptotic properties since it is based on the kernel method, one of the most widely used and known nonparametric methods.

On the other hand, both methods also have some downsides. One major disadvantage of Shapiro-Botha method is that there is no approved rule of node selection. Both Shapiro and Botha (1991) and Cherry (1996) consider only *ad hoc* ways to select nodes, e.g. 200 equispaced nodes. Unfortunately, it seems that selecting equispaced nodes causes spurious oscillations of the estimated function according to the simulation results in Shapiro and Botha (1991), Cherry (1996) and Genton and Gorsich (2002). In order to avoid suffering spurious oscillations, Genton and Gorsich (2002) claims that the nodes

should be chosen as zeros of Bessel functions. The node selection, however, remains as one of difficult tasks when using Shapiro-Botha method.

Meanwhile, those who use Hall-Fisher-Hoffman method need to select a proper positive rendering operator for the purpose of coercing the Fourier transform of kernel-type covariance function estimators into non-negative functions, since positive definiteness is equivalent to nonnegativity of the Fourier tranform by Bochner's theorem. Positive rendering operator selection seems to be a difficult task because it is undeterminable whether positive rendering affects the finite sample properties of kernel-type covariance function estimators. Furthermore, Hall et al. (1994) only deals with the case $d = 1$.

Overall, both of previous methods tend to remain subjective when the sample size is finite, since it is unmanageable to select nodes or positive rendering operators reasonably. In the present paper, we propose a new nonparametric regression estimator subject to the positive definiteness constraint.

The rest of this paper is organized as follows. Our new method is introduced in Chapter 2. In Section 2.1, three versions of our positive definite nonparametric regression estimator are described. Section 2.2 discusses the universal estimation algorithm to find a proper estimate of the regression function. Also, we illustrate our method with simulated data, as well as real data in Chapter 3. In Section 3.1, our method is evaluated with a simulation study related to the regression problem. Section 3.2 presents the results for estimating the covariance function with both simulated spatial data and real spatial data.

Chapter 2

Methodology

This chapter introduces our new method to estimate the regression function subject to the positive definiteness constraint. Like other methods, our positive definite nonparametric regression estimator is based on Bochner's theorem and Schoenberg's theorem, a modification of Bochner's theorem for continuous positive isotropic functions.

Our model has three different versions: (1) In the general version, we assume that the regression function f is continuous and positive definite. Also, for the sake of simplicity, we assume that $f(\mathbf{0}) = 1$. (2) In the isotropic version, it is additionally assumed in the general case that the regression function is isotropic. (3) In the monotone version, it is additionally assumed in the isotropic case that the regression function is monotone.

Our estimation algorithm is built on the framework of Estimation of Distribution Algorithms (EDAs). Based on the general Iterated Density Estimation Evolutionary Algorithm (IDEA), we propose the estimation algorithm which is applicable for every version of our model.

2.1 Models

First of all, we introduce the general version of our model. By Bochner's theorem, the regression function f can be represented in the form

$$f(\mathbf{x}) = \int_{\mathbb{R}^d} e^{-2\pi i \mathbf{x}\cdot\mathbf{u}} \, \mathrm{d}F(\mathbf{u}) \tag{2.1}$$

where F is a distribution function on $\mathbb{R}^d$. Therefore, it is obvious that estimating the regression function f is tantamount to estimating the distribution function F. Unfortunately, $\mathcal{D}(\mathbb{R}^d)$, the collection of distribution functions on $\mathbb{R}^d$, is too large to estimate F with finite observations. This motivates us to impose a proper restriction on $\mathcal{D}(\mathbb{R}^d)$.

There are several statistical methods to reduce $\mathcal{D}(\mathbb{R}^d)$, but, in this paper, we use the kernel smoothing approach. It is widely known that the kernel method is one of the best data-based procedures for estimating the unknown density or distribution function. Based on the kernel methods organized by Wand and Jones (1994), we can define an kernel-type estimator of the distribution function F by

$$\hat{F}_m(\mathbf{u}) = \frac{1}{m}\sum_{i=1}^{m} \int_{-\infty}^{\mathbf{u}} \mathbf{K}_{\mathbf{H}}(\tilde{\mathbf{u}} - \mathbf{v}_i)\mathrm{d}\tilde{\mathbf{u}}, \tag{2.2}$$

where $m \geq 1$, $\mathbf{v}_1, \cdots, \mathbf{v}_m \in \mathbb{R}^d$, $\mathbf{K}$ is a d-variate kernel function satisfying $\int \mathbf{K}(\mathbf{t})\mathrm{d}\mathbf{t} = 1$, $\mathbf{H}$ is a positive definite $d \times d$ matrix called the bandwidth matrix, and $\mathbf{K}_{\mathbf{H}}(\mathbf{t}) = |\mathbf{H}|^{-\frac{1}{2}}\mathbf{K}(\mathbf{H}^{-\frac{1}{2}}\mathbf{t})$. In most cases, it is supposed that a set of $\mathbf{v}_1, \cdots, \mathbf{v}_m$ is given as "data". On the other hand, in this case, it is just a collection of vectors in $\mathbb{R}^d$ existing for specifying an estimator of the distribution function F. We call a set of $\mathbf{v}_1, \cdots, \mathbf{v}_m$ pseudo-data. Also, we call m the size of pseudo-data.

Finally, our estimator of the regression function f can be expressed as

$$\hat{f}_m(\mathbf{x}) = \int_{\mathbb{R}^d} e^{-2\pi i \mathbf{x}\cdot\mathbf{u}} \, \mathrm{d}\hat{F}_m(\mathbf{u}) \tag{2.3}$$

$$= \frac{1}{m} \sum_{i=1}^{m} \cos(2\pi \mathbf{x} \cdot \mathbf{v}_i) \mathcal{F}(\mathbf{K_H})(\mathbf{x}) \tag{2.4}$$

where $\mathcal{F}(\mathbf{K_H})$ is the Fourier transform of $\mathbf{K_H}$. We call the estimator $\hat{f}_m$ the general version estimator. Although the estimator $\hat{f}_m$ is constructed with the familiar functions and operators, it is extremely computationally expensive to find a proper estimate since one need to determine not only hyperparameters m and $\mathbf{H}$ but a pseudo-data $\mathbf{v}_1, \cdots, \mathbf{v}_m$. Fortunately, this practical difficulty can be avoided when the isotropy of the regression function f is assumed.

In the isotropic version, we additionally assume that there is a function $g : [0, \infty) \to \mathbb{R}$ such that $f(\mathbf{x}) = g(|\mathbf{x}|)$ for all $\mathbf{x} \in \mathbb{R}^d$. Therefore, estimating the regression function f is equivalent to estimating the function g. Note that the function g is referred as a continuous positive definite isotropic function on $\mathbb{R}^d$ for convenience (Fasshauer, 2007). Also, $g(0) = 1$. Then, by Schoenberg's theorem, the function g can be represented in the form

$$g(r) = \int_0^\infty \Omega_d(ru) \mathrm{d}G(u) \tag{2.5}$$

where G is a distribution function on $\mathbb{R}_0^+ = [0, \infty)$,

$$\Omega_d(r) = \begin{cases} \cos(r) & \text{for } d = 1 \\ \mathbf{\Gamma}\left(\frac{d}{2}\right) \left(\frac{2}{r}\right)^{(d-2)/2} \mathbf{J}_{(d-2)/2}(r) & \text{for } d \geq 2, \end{cases}$$

$\mathbf{\Gamma}$ is the gamma function and $\mathbf{J}_\nu$ is the Bessel function of the first kind of order $\nu \in \mathbb{R}$. Similar to the gerneral version, we project the distribution function G on $\mathbb{R}_0^+$ into the finite dimensional function space with kernel methods.

There is an extensive literature on estimating density (or distribution) functions with bounded support that presents four well-known approaches: (1) Reflection method proposed by Schuster (1985), Silverman (1986) and Cline and

Hart (1991) (2) Transformation method proposed by Marron and Ruppert (1994) (3) Boundary kernel method proposed by Gasser and Müller (1979), Gasser et al. (1985) and Müller (1991) (4) Pseudo-data method proposed by Cowling and Hall (1996). Also, Hall and Park (2002) proposes new methods based on the translation. See Zhang et al. (1999) and Karunamuni and Alberts (2005) for more summaries.

In the present paper, reflection method is used to estimate the distribution function G on $\mathbb{R}_0^+$. There are mainly two advantages of the reflection method: one can use most of the kernels commonly used in typical kernel methods and simply derive an estimator of the function g. One disadvantage of the reflection method is that the bias is of order $O(h)$ near the endpoint. However, it is not likely to have a severe effect on defining our estimator.

Similar to the general version, we define a kernel-type estimator of the distribution function G by

$$\hat{G}_m(u) = \frac{1}{m}\sum_{i=1}^{m}\int_0^u [K_h(\tilde{u} - v_i) + K_h(\tilde{u} + v_i)]\mathrm{d}\tilde{u} \tag{2.6}$$

where $m \geq 1$, $v_1, \cdots, v_m \in \mathbb{R}_0^+$, K is a univariate kernel function satisfying $\int K(u)\mathrm{d}u = 1$, $h > 0$ is the bandwidth, and $K_h(t) = \frac{1}{h}K(\frac{t}{h})$. Finally, our estimator of the function g has a form as follows:

$$\hat{g}_m(r) = \frac{1}{m}\sum_{i=1}^{m}\left[\int_0^\infty \Omega_d(ru)K_h(u - v_i)\mathrm{d}u + \int_0^\infty \Omega_d(ru)K_h(u + v_i)\mathrm{d}u\right] \tag{2.7}$$

We call the estimator $\hat{g}_m$ the isotropic version estimator. To find a proper estimate $\hat{g}_m$, it is necessary to determine both the size of pseudo-data m, and the bandwidth h first. After the selection of m and h, one needs to determine a proper pseudo-data $v_1, \cdots, v_m$.

A reason why we propose the isotropic version estimator is that isotropy is one of the most frequently assumed properties of the covariance function

when estimating the covariance function. Furthermore, in some cases, it is additionally assumed that the covariance function is monotone, which is the reason why we propose the monotone version. Note that, in these cases, the covariance function is decreasing and convergent since every positive definite function with a finite value at zero is always bounded.

The monotone version is proposed based on the approach presented by Cherry (1996). Note that Schoenberg's theorem characterizes continuous positive definite isotropic functions on $\mathbb{R}^s$ for not only fixed $s \geq 1$, but all $s \geq 1$. Also, it is known that a continuous positive definite isotropic function on $\mathbb{R}^s$ for all $s \geq 1$ is monotone. Cherry (1996) suggests that using $\exp(-r^2u^2)$ instead of $\Omega_d(ru)$ in equation (2.7) guarantees the monotonicity of the estimate since the equation (2.7) with $\exp(-r^2u^2)$ characterizes continuous positive definite isotropic functions on $\mathbb{R}^s$ for all $s \geq 1$. Based on this approach, we define our monotone estimator by

$$\hat{p}_m(r) = \frac{1}{m} \sum_{i=1}^{m} \left[\int_0^\infty e^{-r^2u^2} K_h\left(u - v_i\right) \mathrm{d}u + \int_0^\infty e^{-r^2u^2} K_h\left(u + v_i\right) \mathrm{d}u \right]. \quad (2.8)$$

We call the estimator $\hat{p}_m$ the monotone version estimator. Similar to the isotropic version, it is necessary to determine both the size of pseudo-data m, and the bandwidth h first. After the selection of m and h, one needs to determine a proper pseudo-data $v_1, \cdots, v_m$.

Unfortunately, both equation (2.7) and equation (2.8) are formulated as unusual indefinite integrals. Therefore, computation of our estimators using the above equations may be difficult since the computation of indefinite integrals may be an intractable task in several cases. To overcome this practical difficulty, we derive other formulas in examples 2.1.1 - 2.1.3 to compute our estimators faster in typical cases.

Example 2.1.1. Assume that $d = 2$ and $r > 0$. Let the kernel K be the uniform kernel. Then the isotropic estimator $\hat{g}_m$ can be represented as

$$\hat{g}_m(r) = \frac{1}{mhr} \sum_{j=1}^{m} \left[\Lambda_0(r(v_j + h)) - \Lambda_0(r(v_j - h)) \right] \tag{2.9}$$

where $\mathbf{H}_\nu$ is the Sturve function of order ν and Λ_0 is defined as

$$\Lambda_0(r) = r\mathbf{J}_0(r) + \frac{\pi r}{2} \left[\mathbf{J}_1(r)\mathbf{H}_0(r) - \mathbf{J}_0(r)\mathbf{H}_1(r) \right].$$

Also, the monotone estimator $\hat{p}_m$ can be represented as

$$\hat{p}_m(r) = \frac{\sqrt{\pi}}{2mhr} \sum_{j=1}^{m} \left[\Phi\left(\sqrt{2} r(v_j + h) \right) - \Phi\left(\sqrt{2} r(v_j - h) \right) \right] \tag{2.10}$$

where Φ is the distribution function of standard normal distribution.

Example 2.1.2. Assume that $d = 2$ and $r > 0$. Let the kernel K be the gaussian kernel. Then the isotropic estimator $\hat{g}_m$ can be represented as

$$\hat{g}_m(r) = \frac{1}{m} e^{-\frac{1}{4}h^2 r^2} \sum_{j=1}^{m} \mathbf{J}_0^{(2)} \left(\frac{\mathrm{i}}{4} h^2 r^2, r v_j, \mathrm{i}^{-1} \right) \tag{2.11}$$

$$= \frac{1}{m\pi} e^{-\frac{1}{4}h^2 r^2} \sum_{j=1}^{m} \int_0^\pi \exp\left(\frac{1}{4} h^2 r^2 \cos(2\phi) \right) \cos(r v_j \sin(\phi)) \mathrm{d}\phi \tag{2.12}$$

where $\mathrm{i} = \sqrt{-1}$ and $\mathbf{J}_0^{(2)}$ is the $\left[{}^{(2)}_{0} \right]$ order generalized Bessel function (Dattoli et al., 1993). Also, the monotone estimator $\hat{p}_m$ can be represented as

$$\hat{p}_m(r) = \frac{1}{mh} \sum_{j=1}^{m} \frac{1}{\sqrt{2A(r,h)}} e^{C_j(r,h)} \tag{2.13}$$

where $A(r,h) = r^2 + \frac{1}{2h^2}$ and $C_j(r,h) = \frac{v_j^2}{4h^4 A(r,h)} - \frac{v_j^2}{2h^2}$ for $j = 1, 2, \cdots, m$.

Example 2.1.3. Assume that $d = 2$ and $r > 0$. Let the kernel K be the epanechnikov kernel. Then the isotropic estimator $\hat{g}_m$ can be represented as

$$\begin{aligned}\hat{g}_m(r) = \frac{3}{4mhr}\sum_{j=1}^{m}\Bigg[&\left(1-\frac{v_j^2}{h^2}\right)\left(\Lambda_0(r(v_j+h))-\Lambda_0(r(v_j-h))\right)\\ &-\left(1-\frac{v_j^2}{h^2}\right)\left(\mathbf{J}_1(r(v_j+h))-\mathbf{J}_1(r(v_j-h))\right)\\ &+\left(\frac{1}{r^2h^2}\right)\left(\lambda_0(r(v_j+h))-\lambda_0(r(v_j-h))\right)\Bigg] \quad (2.14)\end{aligned}$$

where

$$\lambda_0(r) = \frac{\pi r}{2}\left[\mathbf{J}_1(r)\mathbf{H}_0(r) - \mathbf{J}_0(r)\mathbf{H}_1(r)\right].$$

Also, the monotone estimator $\hat{p}_m$ can be represented as

$$\begin{aligned}\hat{p}_m(r) = \frac{3}{4mhr}\sum_{j=1}^{m}\Bigg[&\sqrt{\pi}\left(1-\frac{v_j^2}{h^2}-\frac{1}{2r^2h^2}\right)\left(\Phi(r(v_j+h))-\Phi(r(v_j-h))\right)\\ &+\left(\frac{1}{r^2h^2}\right)\left((v_j+h)e^{-r^2(v_j-h)^2}-(v_j-h)e^{-r^2(v_j+h)^2}\right)\Bigg]. \quad (2.15)\end{aligned}$$

To derive the formulae in the above examples, refer to Gradshteyn and Ryzhik (2014), Rosenheinrich (2016) and Dattoli et al. (1993). Both Gradshteyn and Ryzhik (2014) and Rosenheinrich (2016) provide voluminous tables of indefinite integrals of Bessel functions. Also, Dattoli et al. (1993) discusses the definition of the generalized Bessel function and its integral representation.

Unfortunately, our new estimators are so irregularly characterized that typical estimation approaches cannot be used. There are, however, optimization tools designed to search poorly characterized function spaces, and to be easily implemented on parallel computers. In the next section, we propose an estimation algorithm based on Estimation of Distribution Algorithms.

2.2 Estimation

According to Zhang and Muhlenbein (2004), Evolutionary Algorithms (EAs) are population based algorithms for optimization and search problems. The most popular implementations of EAs are genetic algorithms which are already introduced to statisticians by Chatterjee et al. (1996). For the last few decades, Estimation of Distribution algorithms (EDAs), also called *probabilistic model building genetic algorithms*, have attained great attention. Larranaga (2002) contains brief explanations about EDAs. Also, Zhang and Muhlenbein (2004) and Rastegar and Meybodi (2005) discuss the convergence of EDAs.

In this paper, an algorithmic framework based on IDEA proposed by Bosman and Thierens (1999, 2001) is applied to estimate the distribution function F. The general estimation algorithm is described in algorithm 1. Note that the algorithm for estimating G can be achieved by replacing F with G in algorithm 1. To implement algorithm 1, one need to determine six actions: (1) Initialization (2) Evaluation (3) Selection (4) Estimation (5) Replacement and (6) Termination. In this paper, we introduce these actions used to calculate the isotropic version estimator $\hat{g}_m$ for the illustration in this paper.

First, initialization is to generate l pseudo-datasets randomly from the initial distribution. Empirically, the exponential distribution with rate parameter 1 is a suitable initial distribution in views of stability.

Evaluation is to compute an objective function value for each pseudo-dataset. There are a variety of ways to define an objective function, but we apply the least square approach, that means, the objective function L of i-th pseudo-dataset $\Theta_i^{(k)}$ of k-th step is defined as

$$L(\Theta_i^{(k)}) = \sum_{j=1}^{n} \left(y_j - \hat{g}_m \left(|\mathbf{x}_j|; \Theta_i^{(k)} \right) \right)^2 \tag{2.16}$$

where $\hat{g}_m(|\mathbf{x}_j|;\Theta_i^{(k)})$ is defined by equation (2.7), $\Theta_i^{(k)} = \{v_{1,i}^{(k)}, \cdots, v_{m,i}^{(k)}\}$ and $v_{j,i}^{(k)}$ is the j-th element of i-th pseudo-dataset of k-th step for all $1 \leq j \leq m$, $1 \leq i \leq l$.

Selection is to select $\lfloor \tau l \rfloor$ best pseudo-datasets generated from the previous step with respect to the values of the objective function L. This action of k-step can be described as follows:

- Rearrange l pseudo-datasets $\Theta_1^{(k)}, \cdots, \Theta_l^{(k)}$ in ascending order of objective function values $L(\Theta_1^{(k)}), \cdots, L(\Theta_l^{(k)})$ and denote the ordered pseudo-datasets by $\Theta_{(1)}^{(k)}, \cdots, \Theta_{(l)}^{(k)}$.
- Choose $\lfloor \tau l \rfloor$ pseudo-datasets $\Theta_{(1)}^{(k)}, \cdots, \Theta_{(\lfloor \tau l \rfloor)}^{(k)}$ of which objective function values are smaller than those of other pseudo-datasets $\Theta_{(\lfloor \tau l \rfloor+1)}^{(k)}, \cdots, \Theta_{(l)}^{(k)}$.

Estimation is to estimate the distribution function G with the $\lfloor \tau l \rfloor$ selected pseudo-datasets. In k-th step, we obtain the k-step distribution estimator $\hat{G}^{(k)}$ using equation (2.6) with the merged pseudo-dataset $\Theta^{(k)}$ defined by

$$\Theta^{(k)} = \Theta_{(1)}^{(k)} \cup \cdots \cup \Theta_{(\lfloor \tau l \rfloor)}^{(k)}$$

Replacement is to replace $l - \lfloor \tau l \rfloor$ worst pseudo-datasets contained in the collection of pseudo-datasets with new pseudo-datasets. This action is based on the simulation approach proposed by Silverman (1986). Replacement of k-th step can be described as follows:

- Denote the merged pseudo-dataset $\Theta^{(k)}$ by $\{\tilde{v}_1, \cdots, \tilde{v}_{m\lfloor \tau l \rfloor}\}$.
- Choose $I_1, \cdots, I_{m(l-\lfloor \tau l \rfloor)}$ uniformly with replacement from $\{1, \cdots, m\lfloor \tau l \rfloor\}$.
- Generate $e_1, \cdots, e_{m(l-\lfloor \tau l \rfloor)}$ to have probability density function K.
- Set $\tilde{v}_i' = \mid \tilde{v}_{I_i} + e_i \mid$ for all $i = 1, 2, \cdots, m(l - \lfloor \tau l \rfloor)$.

- Divide $\{\tilde{v}'_1, \cdots, \tilde{v}'_{m(l-\lfloor \tau l \rfloor)}\}$ into $l - \lfloor \tau l \rfloor$ pseudo-datasets $\Theta^{(k+1)}_{\lfloor \tau l \rfloor + 1}, \cdots, \Theta^{(k+1)}_l$ uniformly.

- Set $\Theta^{(k+1)}_1 = \Theta^{(k)}_{(1)}, \cdots, \Theta^{(k+1)}_{\lfloor \tau l \rfloor} = \Theta^{(k)}_{(\lfloor \tau l \rfloor)}$.

Finally, $\Theta^{(k+1)}_1, \cdots, \Theta^{(k+1)}_l$ are $(k+1)$-step pseudo-datasets of algorithm 1.

Termination is based on the concept of Kullback-Leibler divergence proposed by Kullback and Leibler (1951). Let $\hat{\psi}^{(k)}$ be the kernel density estimator correspoding to the k-th step estimator $\hat{G}^{(k)}$ of the distribution G. Then, by Hershey and Olsen (2007), Kullback-Leibler divergence from $\hat{G}^{(k-1)}$ to $\hat{G}^{(k)}$ can be calculated by

$$\hat{D}_{KL}(\hat{G}^{(k)}, \hat{G}^{(k-1)}) = \frac{1}{m\lfloor \tau l \rfloor} \sum_{i=1}^{\lfloor \tau l \rfloor} \sum_{j=1}^{m} \log \left(\frac{\hat{\psi}^{(k)}_m(v^{(k-1)}_{j,i})}{\hat{\psi}^{(k-1)}_m(v^{(k-1)}_{j,i})} \right). \tag{2.17}$$

Therefore, algorithm 1 is terminated when $\hat{D}_{KL}$ is smaller than a certain preselected value. In the illustration, we terminate the algorithm when the stopping condition $\hat{D}_{KL} < 10^{-3}$ has been met five times consecutively for stability.

Finally, we can compute the estimate using equation (2.7) with the pseudo-data obtained by merging a collection of $\lfloor \tau l \rfloor$ selected pseudo-datasets of the last step of algorithm 1. To apply algorithm 1, one needs to preselect four hyperparameters: the bandwidth $h > 0$, the size m of pseudo-data, the size l of collection of pseudo-datasets, and the proportion τ of selection. Fortunately, it seems empirically that both $\frac{l}{m}$ and τ do not significantly affect the estimation result. We suggest selecting $l = 10m$ and $\tau = 0.1$ in algorithm 1. On the other hand, we recommend a grid-search using cross validation for determining the values of h and m since they have great influence on our estimators.

Algorithm 1 The general version of IDEA

procedure IDEA(h, m, l, τ)
  Generate a collection of l pseudo-datasets ▷ Initialization
  while the termination rule is true **do**
    Evaluate objective function values of pseudo-datasets ▷ Evaluation
    Select $\lfloor \tau l \rfloor$ best pseudo-datasets ▷ Selection
    Estimate F with $\lfloor \tau l \rfloor$ selected pseudo-datasets ▷ Estimation
    Check the termination rule ▷ Termination
    if the termination rule is false **then**
      Generate $l - \lfloor \tau l \rfloor$ pseudo-datasets ▷ Replacement
    end if
  end while
  return a collection of $\lfloor \tau l \rfloor$ selected pseudo-datasets
end procedure

Chapter 3

Illustration

In this chapter, to identify feasibility of our estimators, we provide some simulation results and applications focusing on isotropic covariance function estimation. To apply our estimation algorithm, we choose $l = 10m$, $\tau = 0.1$, and both m and h are chosen with a grid-search using 10-fold cross validation. Also, we terminate the algorithm when the stopping condition $\hat{D}_{KL} < 10^{-3}$ has been met five times consecutively.

3.1 Simulation Results

Suppose that the dimension d is 2. We generate simulated datasets by

$$y_i = g(r_i) + \epsilon_i\,, \quad i = 1, 2, \cdots, 100\,, \quad r_i \sim U(0, 10)\,, \quad \epsilon_i \sim N(0, 0.2)$$

and the ϵ_i's and r_i's are totally independent. Also, we consider four positive definite functions for g : (1) wave function with $c = 1$ (2) wave function with $c = 2$ (3) spherical function (4) exponential function. These functions are defined with reference to the isotropic variogram models presented in Cressie (2015) as follows.

· Wave function with a constant $c > 0$: $g(r) = \begin{cases} 1 & \text{for } r = 0 \\ \frac{\sin(r/c)}{r/c} & \text{for } r > 0 \end{cases}$

· Spherical function: $g(r) = \begin{cases} 1 & \text{for } r = 0 \\ \frac{1}{5}\left(3r - \frac{1}{4}r^3\right) & \text{for } 0 < r \leq 2 \\ \frac{4}{5} & \text{for } 2 < r \end{cases}$

· Exponential function: $g(r) = \exp\left(-\frac{r}{4}\right)$

We obtain the simulation results as shown in figure 3.1 and figure 3.2. To find the optimal values of h and m with the grid search, the values of m are considered from 3 to 20 in coarse increments, and the values of h are considered from 0.02 to 0.5 in increments of size 0.02. In wave cases, we compare three nonparametric regression estimators: our isotropic version estimator, local constant estimator, and local linear estimator. Similarly, in monotone cases, we compare three regression estimators: our monotone version estimator, monotone estimator with Bernstein polynomials proposed by McKay Curtis and Ghosh (2011) (Bernstein estimator), and the monotone estimator proposed by Dette et al. (2006) (Dette estimator).

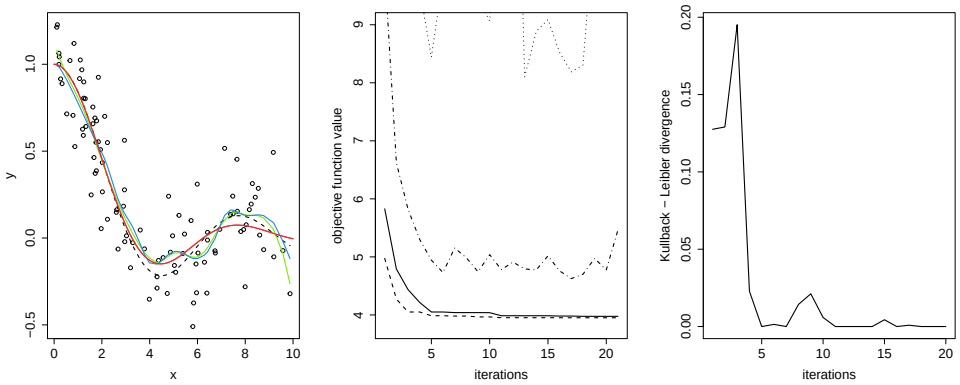
Figure 3.1 describes three nonparametric regression estimators, including our isotropic version estimator, and the associated true regression function. Both local constant estimator and local linear estimator are implemented by Hayfield and Racine (2008) in R-package **np**. In the case of figure 3.1(a), the cross-validated (h, m) of our isotropic version estimator are $(0.18, 5)$. Meanwhile, the cross-validated bandwidths of the local constant estimator and the local linear estimator are 0.422 and 0.488, respectively. Figure 3.1(a) shows that our isotropic version estimator produces substantially smoother curves than the other curves related to other nonparametric regression estimators. On the contrary, both local constant estimator and local linear estimator os-

cillate near the right-hand side of the plot. The stability of our estimator is stressed in figure 3.1(b). As shown in the left plot of figure 3.1(b), error variance is large in terms of the amplitude of the right side of the true regression function. In this case, our isotropic version estimator is likely to be more robust than others. The cross-validated (h, m) of our isotropic version estimator are $(0.24, 8)$. Meanwhile, the cross-validated bandwidths of the local constant estimator and the local linear estimator are 0.319 and 0.413, respectively.

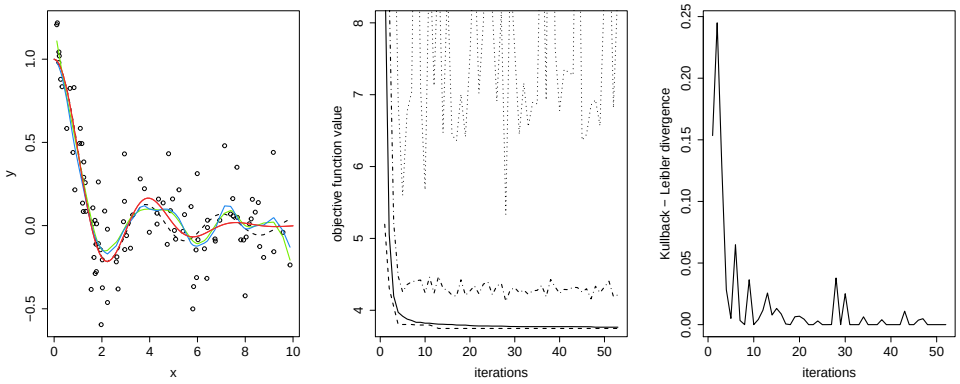
Figure 3.2 describes our monotone version estimator, Bernstein estimator, Dette estimator, and the associated the true regression function. Bernstein estimator is implemented by Curtis (2015) in R-package **bisoreg**, and Dette estimator is implemented by Pilz et al. (2015) in R-package **monreg**. In the case of figure 3.2(a), the cross-validated (h, m) of our monotone version estimator are $(0.08, 4)$. Meanwhile, the cross-validated (h_d, h_r) of Dette estimator are $(0.144, 0.38)$. As shown in figure 3.2(a), both our monotone version estimator and Bernstein estimator are likely to be more accurate than Dette estimator when the true regression function has a flat spot. Figure 3.2(b) shows that our monotone version estimator is closer to the true regression function than others near the centre of the left plot. The reason for this is to impose the constraint of positive definiteness. In this case, the cross-validated (h, m) of our monotone version estimator are $(0.48, 7)$. Meanwhile, the cross-validated (h_d, h_r) of Dette estimator are $(0.04, 0.2)$.

3.2 Applications

In this section, we propose how our estimators can be used to estimate the covariance function with spatial data. Consider a stationary spatial stochastic process $\{Z(\mathbf{s}) \mid \mathbf{s} \in D\}$, where the spatial domain D is a bounded subset of

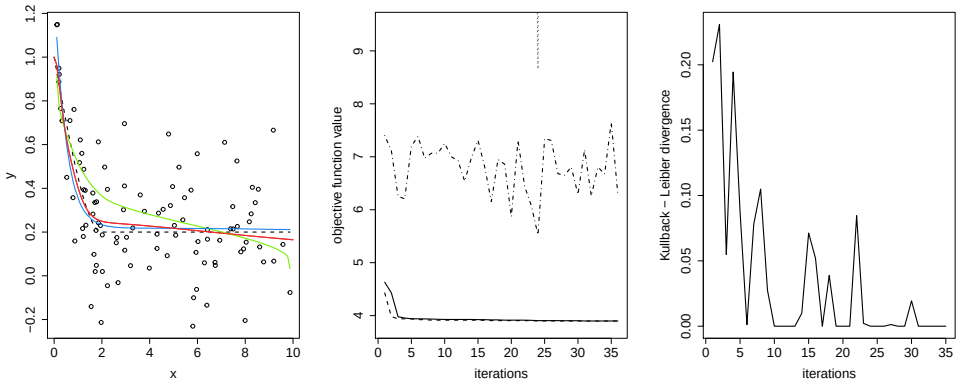


(a) Wave regression function with $c = 1$

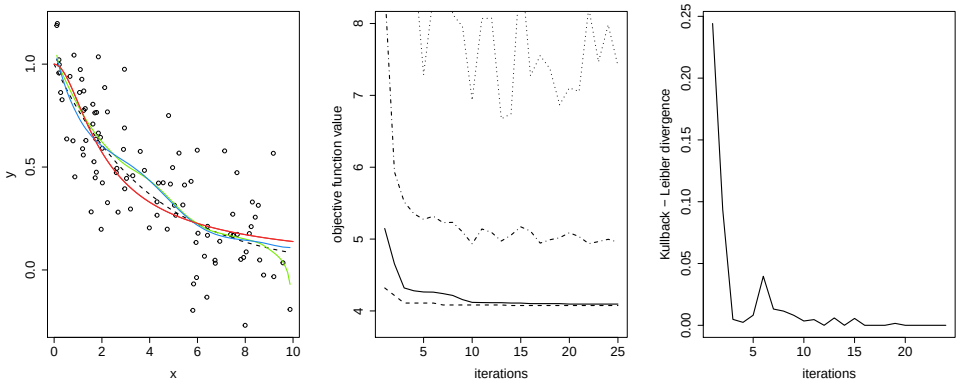


(b) Wave regression function with $c = 2$

Figure 3.1: Simulation study with regression setting comparing our isotropic version estimator, local constant estimator and local linear estimator with the gaussian kernel: **(Left)** The dot curve shows the true regression function. The red curve shows our isotropic version estimator. The blue curve shows local constant estimator. The green curve shows local linear estimator. **(Center)** Objective function values against iteration for a given run of the IDEA: Lowest curve shows the minimum value $L(\Theta_{(1)}^{(k)}), k \geq 1$. Second-lowest curve shows the worst selected value $L(\Theta_{(\lfloor \tau l \rfloor)}^{(k)}), k \geq 1$,. Second-highest curve shows the mean value of $L(\Theta_{(1)}^{(k)}), \cdots, L(\Theta_{(l)}^{(k)}), k \geq 1$. Highest curve shows the maximum value $L(\Theta_{(l)}^{(k)}), k \geq 1$. **(Right)** Kullback-Leibler divergence $\hat{D}_{KL}(\hat{G}^{(k)}, \hat{G}^{(k-1)}), k \geq 2$, against iteration for a given run of the IDEA.



(a) Spherical regression function



(b) Exponential regression function

Figure 3.2: Simulation study with regression setting comparing our monotone version estimator with the gaussian kernel, Bernstein estimator with the default setting and Dette estimator with the epanechnikov kernel : **(Left)** The dot curve shows the true regression function. The red curve shows our monotone version estimator. The blue curve shows Bernstein estimator with the default setting. The green curve shows Dette estimator with the epanechnikov kernel. A reason why we use Dette estimator with the epanechnikov kernel is that the gaussian kernel option is not implemented in R-package **monreg**. **(Center)** Objective function values against iteration for a given run of the IDEA: Lowest curve shows the minimum value $L(\Theta_{(1)}^{(k)}), k \geq 1$. Second-lowest curve shows the worst selected value $L(\Theta_{(\lfloor \tau l \rfloor)}^{(k)}), k \geq 1$,. Second-highest curve shows the mean value of $L(\Theta_{(1)}^{(k)}), \cdots, L(\Theta_{(l)}^{(k)}), k \geq 1$. Highest curve shows the maximum value $L(\Theta_{(l)}^{(k)}), k \geq 1$. **(Right)** Kullback-Leibler divergence $\hat{D}_{KL}(\hat{G}^{(k)}, \hat{G}^{(k-1)}), k \geq 2$, against iteration for a given run of the IDEA.

$\mathbb{R}^d$. In this setting, the covariance function $C : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$C(\mathbf{s} - \mathbf{s}') = \text{cov}\left(Z(\mathbf{s}), Z(\mathbf{s}')\right) \quad \text{for all } \mathbf{s}, \mathbf{s}' \in D. \tag{3.1}$$

Let $Z(\mathbf{s}_1), Z(\mathbf{s}_2), \cdots, Z(\mathbf{s}_w)$ be a realization of the stationary spatial stochastic process Z. Based on the classical estimator of the semivariogram proposed by Matheron (1962), we can estimate the covariance function by

$$\hat{c}_{i,j} = \hat{C}(\mathbf{s}_i - \mathbf{s}_j) = \left(Z(\mathbf{s}_i) - \bar{Z}\right)\left(Z(\mathbf{s}_j) - \bar{Z}\right) \tag{3.2}$$

where $\bar{Z} = \frac{1}{w}\sum_{i=1}^{w} Z(\mathbf{s}_i)$. We call them the classical estimators of the covariance function. Therefore, this setting can be viewed as a regression problem with observations $(\mathbf{r}_{i,j}, \hat{c}_{i,j})$ where $\mathbf{r}_{i,j} = \mathbf{s}_i - \mathbf{s}_j$. According to Hall et al. (1994), the set of observations may either include or exclude the diagonal terms corresponding to $i = j$. In this paper, we include diagonal terms.

Now we propose the procedure to estimate the covariance function C with our estimators. Note that we assume that the regression function f is one at zero in Chapter 2. Therefore, before applying our method, we need to normalize the covariance function C. We suggest the following approach. First, estimate $C(\mathbf{0})$ with the common estimator $\hat{V} = \frac{1}{w}\sum_{i=1}^{w}\left(Z(\mathbf{s}_i) - \bar{Z}\right)^2$. Next, estimate the function $C(\mathbf{r})/\hat{V}$ with our estimators and observations $(\mathbf{r}_{i,j}, \hat{c}_{i,j})$ where $i \neq j$. Let the resulting function be represented by $\hat{C}_0(\mathbf{r})$. Finally, the function $\hat{C}(\mathbf{r}) = \hat{V}\hat{C}_0(\mathbf{r})$ is our covariance function estimator.

Now we conduct a numerical study to evaluate the performance of the proposed method. To find the optimal values of h and m with the grid search, values of m are considered from 4 to 7 in unit increments, and values of h are considered from 0.02 to 0.2 in increments of size 0.02. First, we provide a simulation result of covariance function estimation. In the simulation, the uniform distribution on $[0, 20/3] \times [0, 20/3]$ is assumed for spatial locations

$\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{30} \in \mathbb{R}^2$. Also, datasets $Z(\mathbf{s}_1), Z(\mathbf{s}_2), \cdots, Z(\mathbf{s}_{30})$ are generated with gaussian random field using the wave function with $c = 1$ as the covariance function. This simulation data is generated with R-package **geoR** implemented by Ribeiro Jr and Diggle (2016). For simplicity, we assume that the variance of the stochastic process Z is known.

Figure 3.3 describes three nonparametric estimators, including our isotropic version estimator, and the associated true covariance function. The cross-validated (h, m) of our isotropic version estimator are $(0.06, 6)$. Meanwhile, the cross-validated bandwidths of the local constant estimator and the local linear estimator are 0.355 and 0.433, respectively. It seems that, especially in the right side of the left plot, our isotropic version estimator provides an estimate that is near the true covariance function than others.

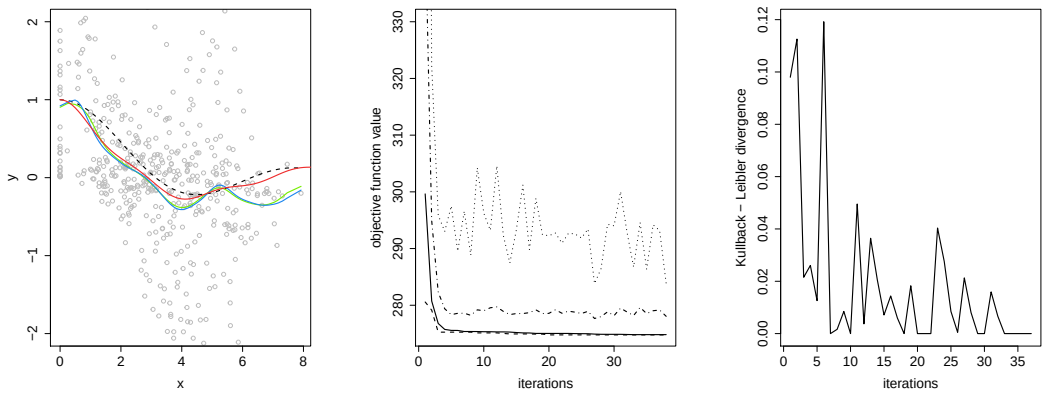


Figure 3.3: Simulation study with the stationary spatial stochastic process comparing our isotropic version estimator, local constant estimator and local linear estimator with the gaussian kernel: **(Left)** The dot curve shows the true regression function. The red curve shows our isotropic version estimator. The blue curve shows local constant estimator. The green curve shows local linear estimator. The gray points shows the classical estimates of the covariance function with equation (3.2). **(Center)** Objective function values against iteration for a given run of the IDEA: Lowest curve shows the minimum value $L(\Theta^{(k)}_{(1)}), k \geq 1$. Second-lowest curve shows the worst selected value $L(\Theta^{(k)}_{(\lfloor \tau l \rfloor)}), k \geq 1$,. Second-highest curve shows the mean value of $L(\Theta^{(k)}_{(1)}), \cdots, L(\Theta^{(k)}_{(l)}), k \geq 1$. Highest curve shows the maximum value $L(\Theta^{(k)}_{(l)}), k \geq 1$. **(Right)** Kullback-Leibler divergence $\hat{D}_{KL}(\hat{G}^{(k)}, \hat{G}^{(k-1)}), k \geq 2$, against iteration for a given run of the IDEA.

Second, we use the real data **sic.100** from the SIC-97 project to illustrate our covariance function estimator. Data **sic.100** is 100 daily rainfall observations made in Switzerland on the 8^{th} of May 1986 which were randomly selected from a dataset of 467 observations. Circle plot of data **sic.100** and border of Switzerland is described in Figure 3.4. This dataset is contained in the R-package **geoR**.[1] In this paper, we concentrate on the estimation of the covariance function of 100 rainfall observations.

Figure 3.5 shows three nonparametric estimators, including our isotropic version estimator of the covariance function of data **sic.100**. Note that, for stability, distances between the observations are normalized to a maximum value of 8 when we estimate the covariance function, and retrieved when we describe figure 3.5. Since the cross-validated (h, m) of our isotropic version estimator are $(0.1, 7)$, there is a possibility that the estimate of our method is improved by using more wider range of the grid search. Nevertheless, our estimator is likely to provide a more reasonable result than typical nonparametric regression estimators, especially in the right side of the left plot. Meanwhile, the cross-validated bandwidths of the local constant estimator and the local linear estimator are 0.196 and 0.210, respectively.

[1]See `https://wiki.52north.org/bin/view/AI_GEOSTATS/EventsSIC97` for details.

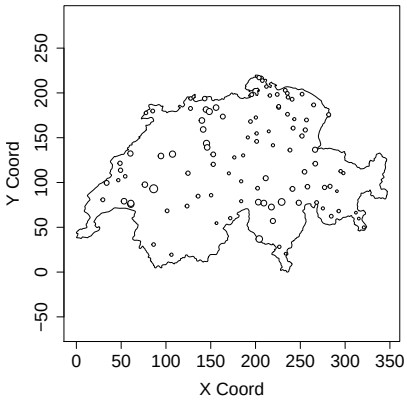


Figure 3.4: Circle plot of data **sic.100** and border of Switzerland

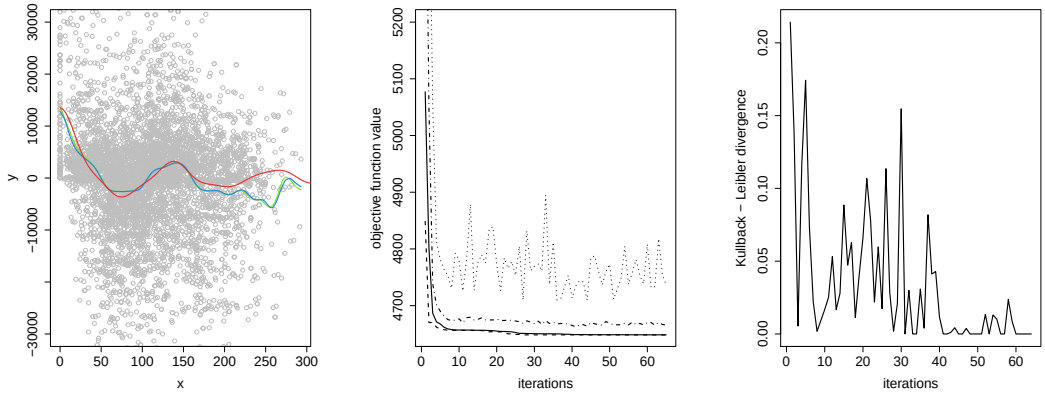


Figure 3.5: Application to covariance function estimation with real data **sic.100 (Left)** The red curve shows our isotropic version estimator. The blue curve shows local constant estimator. The green curve shows local linear estimator. The gray points shows the classical estimates of the covariance function with equation (3.2). **(Center)** Objective function values against iteration for a given run of the IDEA: Lowest curve shows the minimum value $L(\Theta_{(1)}^{(k)}), k \geq 1$. Second-lowest curve shows the worst selected value $L(\Theta_{(\lfloor \tau l \rfloor)}^{(k)}), k \geq 1$,. Second-highest curve shows the mean value of $L(\Theta_{(1)}^{(k)}), \cdots, L(\Theta_{(l)}^{(k)}), k \geq 1$. Highest curve shows the maximum value $L(\Theta_{(l)}^{(k)}), k \geq 1$. **(Right)** Kullback-Leibler divergence $\hat{D}_{KL}(\hat{G}^{(k)}, \hat{G}^{(k-1)}), k \geq 2$, against iteration for a given run of the IDEA.

Chapter 4

Conclusion

In this paper, we propose the nonparametric regression method subject to the positive definiteness constraint. The proposed method provides a regression-type solution to a problem that have not received enough attention from statisticians, namely to ensure that nonparametric covariance function estimators are not only positive definite but also smooth. To treat various preconditions including isotropy and monotonicity, we suggest three different versions of our estimator and the estimation algorithm which is applicable to any versions. Based on the illustration above, it seems that our estimators are more robust than typical nonparametric estimators. This finite sample property makes our estimators be adequate to estimate the covariance function.

Furthermore, our method can be modified for various purposes. There are several properties of functions related to integral representations. For instance, completely monotone functions and multiply monotone functions can be formulated as integral transform of non-negative Borel measures. Further investigation is necessary to generalize our approach to general properties.

References

Armstrong, M. and Jabin, R. (1981). Variogram models must be positive-definite. *Mathematical Geology*, 13(5):455–459.

Bochner, S. (1933). Integration von funktionen, deren werte die elemente eines vektorraumes sind. *Fundamenta Mathematicae*, 20(1):262–176.

Bosman, P. A. and Thierens, D. (1999). *An algorithmic framework for density estimation based evolutionary algorithms*. Univ.

Bosman, P. A. and Thierens, D. (2001). Continuous iterated density estimation evolutionary algorithms within the idea framework.

Chatterjee, S., Laudato, M., and Lynch, L. A. (1996). Genetic algorithms and their statistical applications: an introduction. *Computational Statistics & Data Analysis*, 22(6):633–651.

Cheney, E. W. and Light, W. A. (2009). *A course in approximation theory*, volume 101. American Mathematical Soc.

Cherry, S. (1996). An evaluation of a non-parametric method of estimating semi-variograms of isotropic spatial processes. *Journal of Applied Statistics*, 23(4):435–449.

Christakos, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research*, 20(2):251–265.

Cline, D. B. H. and Hart, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics: A Journal of Theoretical and Applied Statistics*, 22(1):69–84.

Cowling, A. and Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 551–563.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Curtis, S. M. (2015). *bisoreg: Bayesian Isotonic Regression with Bernstein Polynomials*. R package version 1.4.

Dattoli, G., Chiccoli, C., Lorenzutta, S., Maino, G., Richetta, M., and Torre, A. (1993). Advances on the theory of generalized bessel functions and applications to multiphoton processes. *Journal of Scientific Computing*, 8(1):69–109.

Dette, H., Neumeyer, N., Pilz, K. F., et al. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3):469–490.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.

Fasshauer, G. E. (2007). *Meshfree approximation methods with MATLAB*, volume 6. World Scientific.

García-Soidán, P. (2007). Asymptotic normality of the nadaraya–watson semivariogram estimators. *Test*, 16(3):479–503.

Garcìa-Soidán, P. H., Febrero-Bande, M., and González-Manteiga, W. (2004). Nonparametric kernel estimation of an isotropic variogram. *Journal of Statistical Planning and Inference*, 121(1):65–92.

García-Soidán, P. H., González-Manteiga, W., and Febrero-Bande, M. (2003). Local linear regression estimation of the variogram. *Statistics & probability letters*, 64(2):169–179.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation*, pages 23–68. Springer.

Gasser, T., Muller, H.-G., and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 238–252.

Genton, M. G. and Gorsich, D. J. (2002). Nonparametric variogram and covariogram estimation with fourier–bessel matrices. *Computational Statistics & Data Analysis*, 41(1):47–57.

Gorsich, D. J. and Genton, M. G. (2000). Variogram model selection via nonparametric derivative estimation. *Mathematical Geology*, 32(3):249–270.

Gradshteyn, I. S. and Ryzhik, I. M. (2014). *Table of integrals, series, and products*. Academic press.

Hall, P., Fisher, N. I., and Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *The Annals of Statistics*, pages 2115–2134.

Hall, P. and Park, B. U. (2002). New methods for bias correction at endpoints and boundaries. *Annals of Statistics*, pages 1460–1479.

Härdle, W. (1990). *Applied nonparametric regression*. Number 19. Cambridge university press.

Hayfield, T. and Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5).

Hershey, J. R. and Olsen, P. A. (2007). Approximating the kullback leibler divergence between gaussian mixture models. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–317. IEEE.

Karunamuni, R. J. and Alberts, T. (2005). On boundary correction in kernel density estimation. *Statistical Methodology*, 2(3):191–212.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

Larranaga, P. (2002). A review on estimation of distribution algorithms. In *Estimation of distribution algorithms*, pages 57–100. Springer.

Marron, J. S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 653–671.

Matheron, G. (1962). *Traité de géostatistique appliquée. 1 (1962)*, volume 1. Editions Technip.

McKay Curtis, S. and Ghosh, S. K. (2011). A variable selection approach to monotonic regression with bernstein polynomials. *Journal of Applied Statistics*, 38(5):961–976.

Menezes, R., Garcia-Soidán, P., and Febrero-Bande, M. (2005). A comparison of approaches for valid variogram achievement. *Computational statistics*, 20(4):623–642.

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, pages 521–530.

Pilz, K., developements by Holger Dette, S. T. E., and Pilz., K. (2015). *monreg: Nonparametric Monotone Regression*. R package version 0.1.3.

Rastegar, R. and Meybodi, M. (2005). A study on the global convergence time complexity of estimation of distribution algorithms. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 441–450.

Ribeiro Jr, P. J. and Diggle, P. J. (2016). *geoR: Analysis of Geostatistical Data*. R package version 1.7-5.2.

Rosenheinrich, W. (2016). Tables of some indefinite integrals of bessel functions. http://www.eah-jena.de/ rsh/Forschung/Stoer/besint.pdf.

Schoenberg, I. J. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics*, pages 811–841.

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics-Theory and methods*, 14(5):1123–1136.

Shapiro, A. and Botha, J. D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics & Data Analysis*, 11(1):87–96.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. Crc Press.

Zhang, Q. and Muhlenbein, H. (2004). On the convergence of a class of estimation of distribution algorithms. *IEEE Transactions on evolutionary computation*, 8(2):127–136.

Zhang, S., Karunamuni, R., and Jones, M. (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, 94(448):1231–1240.

국문초록

우리는 정상확률과정(stationary stochastic process)의 공분산 함수를 적절히 추정하기 위한 추정량의 양정치성(positive definiteness)을 보장하는 새로운 비모수 회귀분석 방법을 제안하였다. 우리의 방법은 추정량의 양정치성 뿐 만 아니라 등방성(isotropy)과 단조성(monotonicity)도 보장할 수 있다. 또한, 기존의 방법들과 다르게 교차검증법을 통해 자동적으로 추정량을 얻을 수 있다. 이런 장점들을 얻기 위해 우리는 커널 분포 추정량(kernel distribution estimator)의 적분 변환을 통해 추정량을 정의하였으며, 분포 추정 알고리즘(estimation of distribution algorithm)에 기반한 범용적인 알고리즘을 제안하였다. 또한 간단한 모의실험을 통해 우리의 방법이 통상적인 비모수회귀모형 방법들에 비해 강인하게 공분산 함수를 추정하는데 유용함을 보여주었다. 더불어 실제 공간자료 sic.100에 우리의 방법을 적용한 결과를 제시하였다.

주요어 : 비모수 회귀분석, 커널, 양정치성, 등방성, 단조성, 공분산,
베리오그램, 분포 추정 알고리즘

학 번 : 2013-20209