We consider the problem of estimation of a probability density function, by the method of delta sequences, for data with values in a semi-metric space.

Key words: Nonparametric density estimation; Method of delta sequences; Probability measure on a semi-metric space.

1 Introduction

Methods of nonparametric estimation of a density function are widely discussed in the literature starting from Prakasa Rao [15,16], Silverman [20] and more recently in Efromovich [8]. Modeling spatial data is of great importance and interest for applications in geology, soil science, ecology, geography, image processing, atmospheric science and other fields where spatial data are collected. Cressie [6], Ripley [19], Guyon [10] and Chiles and Delfiner [5], among others, discuss statistical methods for analysis of spatial data. Basse et al. [1] have recently studied kernel density estimation for spatial functional random variables. Spatial nonparametric estimation of the density of a real random field was studied in Tran [21], Tran and Yakowitz [22], Carbon et al. [3], Carbon et al. [4] and Hallin et al. [11] among others.
In the framework of either univariate or multivariate real-valued random variables, the probability density function is defined either with respect to the Lebesgue measure or the counting measure. For the spatial models under discussion here, there is no analogue of the Lebesgue measure on a semi-metric space. The density function of a random element here, if it exists, is related to the existence of a dominating measure of the family of probability measures with respect to which the density function or the Radon-Nikodym derivative is computed. The choice of the dominating measure for the family of probability measures, if it exists, is not crucial in view of the chain rule applicable for Radon-Nikodym derivatives. Any estimator of the density function with respect to one dominating measure can be transformed to an estimator of the corresponding probability density function with respect to another dominating measure by the chain rule for the Radon-Nikodym derivatives. In view of this, the existence of a dominating measure for the family of probability measures is sufficient for our discussion. Example of such a dominating measure for functional spaces is discussed later in this section. Problems involving the density estimation of a random element taking values in a metric space were earlier studied by Geffray [9]. Wertz [25] and Crosswell [7] investigated the properties of kernel type density estimators for random elements taking values in locally compact topological groups (cf. Prakasa Rao [15], p.226)). Bosq and Blanke [2] studied inference and prediction for models where the data and (or) parameters belong to a large or infinite dimensional space.

Our aim in this paper is to study density estimation for random elements taking values in a separable semi-metric space space which is possibly of infinite dimension. We study the problem of density estimation through the method of delta sequences generalizing the method of kernel density estimation in Basse et al. [1]. It is known that the method of delta sequences unifies the kernel method of density estimation, the histogram method and some other methods such as the method of orthogonal series for a suitable choice of orthonormal bases in the one-dimensional
and finite-dimensional cases. For a discussion of the method of delta sequences in the finite-dimensional cases, see Prakasa Rao [15], pp.136-143 and pp. 218-224. Nonparametric density estimation for functional data by delta sequences is studied in Prakasa Rao [18].

Let \((\Omega, \mathcal{F}, P)\) be a probability space and \(\{\mathcal{F}_t, t \geq 0\}\) be a nondecreasing family of sub-\(\sigma\)-algebras of \(\mathcal{F}\). Let \(\{W_t, t \geq 0\}\) be a standard Wiener process defined on \(\{\Omega, \mathcal{F}, P\}\) such that \(W_t\) is \(\mathcal{F}_t\)-measurable. Let \(C[0, T]\) be the space of real-valued continuous functions defined on the interval \([0, T]\) associated with supremum norm topology. It is known that the standard Wiener process induces a probability measure \(\mu_W\) on the space \(C[0, T]\) associated with Borel \(\sigma\)-algebra generated by the supremum norm topology. Consider a diffusion process \(\{X(t), 0 \leq t \leq T\}\) governed by the stochastic differential equation

\[
dX(t) = a(t, X(t))dt + b(t, X(t))dW(t), X(0) = x_0, 0 \leq t \leq T.
\]

Under some conditions on the functions \(a(., .)\) and \(b(., .)\), it can be shown that the probability measure \(\mu_X\) induced by the process \(X\) on the space \(C[0, T]\) is absolutely continuous with respect to the probability measure \(\mu_W\) and one can compute the Radon-Nikodym derivative of \(\mu_X\) with respect to \(\mu_W\) by using the Girsanov’s theorem. This can be considered as the probability density of the process \(X\) on the separable metric space \(C[0, T]\) under the supremum norm metric. More details on such a frame work and other examples are given in Prakasa Rao [17]. One of the motivations for analysis of spatial data in our view is inference for random fields. We are assuming here that the complete path of the process is observable for inferential purposes. However, if the process can be observed only at discrete times either on a fine grid or when the data is sparse, other methods have to be developed as in the case of parametric inference for discrete data, for instance, for diffusion processes (cf. Prakasa Rao [17]).
2 Preliminaries

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(E, d, \mathcal{B})$ be a separable semi-metric space with the associated Borel $\sigma$-algebra $\mathcal{B}$. Suppose $X$ is a random element defined on $(\Omega, \mathcal{F}, P)$ taking values in $(E, \mathcal{B})$ and that it has a probability density function $f$ with respect to a $\sigma$-finite and diffuse measure $\mu$ on $(E, \mathcal{B})$ such that $0 < \mu(A) < \infty$ for every open ball $A \subset E$. Let $\mathcal{I}_n = \{i \in \mathcal{N}^N : 1 \leq i_k \leq n_k, k = 1, \ldots, N\}$ where $n = (n_1, \ldots, n_N)$. Here $\mathcal{N} = \{1, 2, 3, \ldots\}$. We write $n \to \infty$ if

$$\min_{k=1,\ldots,N} n_k \to \infty$$

and

$$\left|\frac{n_j}{n_k}\right| < C$$

for some constant $0 < C < \infty$. Let $\mathbf{n} = n_1 \times \ldots \times n_N$. We now consider the problem of estimation of the probability density function $f$ based on a spatial sample obtained over the rectangular region $\mathcal{I}_n$. Let $\{X_i, i \in \mathcal{I}_n\}$ be independent and identically distributed random elements as $X$ with the probability density function $f$ with respect to the measure $\mu$. Let $x \in E$. We assume that

(G1) the density function $f$ is continuous at $x \in E$ and $\sup_{y \in E} |f(y)| \leq M < \infty$.

Definition: A family of nonnegative measurable functions $\{\delta_n(y, x), n \in \mathcal{N}^N\}$ defined on $E \times E$ is said to be a delta family with respect to the measure $\mu$ if the following conditions hold:

(G2) for every $\gamma, 0 < \gamma \leq \infty$,

$$\lim_{m \to \infty} \left| \int_{|y-d(x, y)| \leq \gamma} \delta_m(x, y) \mu(dy) - 1 \right| = 0;$$

(G3) for any $\gamma > 0$,
\[
\lim_{m \to \infty} \sup_{y:d(y,x) > \gamma} \delta_m(x, y) d(y, x) = 0;
\]

(G4) further suppose that

\[
S_m^x = \sup_{y \in E} \delta_m(y, x) < \infty.
\]

Let

\[
f_n(x) = \frac{1}{n} \sum_{i \in I_n} \delta_m(x, X_i).
\]

The choice of \( m \) might depend on \( n \) such that \( m \to \infty \) as \( n \to \infty \).

We now prove the following result leading to mean square consistency of the estimator \( f_n(x) \) as an estimator of \( f(x) \).

**Theorem**: Let \( x \in E \). Suppose that \( \lim_{n \to \infty} (\hat{n}/S_n^x) = \infty \). Then, under the conditions (G1)-(G4),

\[
\lim_{n \to \infty} E[(f_n(x) - f(x))^2] = 0.
\]

**Proof**: Let \( \gamma > 0 \) and \( x \in E \). Define

\[
I_1(x) = \int_{[y:d(y,x) \leq \gamma]} \delta_m(x, y)(f(y) - f(x)) \mu(dy)
\]

and

\[
I_2(x) = \int_{[y:d(y,x) > \gamma]} \delta_m(x, y)(f(y) - f(x)) \mu(dy).
\]

Observe that

\[
E[f_n(x)] - f(x) = \int_E \delta_m(x, y) f(y) \mu(dy) - f(x)
\]
and hence
\[
E[f_n(x)] - f(x) - I_1(x) - I_2(x) = \int_E \delta_m(x, y) f(y) \mu(dy) - f(x) \\
- \int_E \delta_m(x, y)(f(y) - f(x)) \mu(dy) \\
= f(x) \int_E \delta_m(x, y) \mu(dy) - 1.
\]

Note that
\[
\lim_{m \to \infty} |\int_E \delta_m(x, y) \mu(dy) - 1| = 0
\]
by (G2). Hence
\[
\lim_{m \to \infty} |E[f_n(x)] - f(x) - I_1(x) - I_2(x)| = 0
\]
by the condition (G2). Furthermore
\[
|I_2(x)| \leq \int_{[y:d(y,x) > \gamma]} \delta_m(x, y) f(y) \mu(dy) + f(x) \int_{[y:d(y,x) > \gamma]} \delta_m(x, y) \mu(dy) \\
= \int_{[y:d(y,x) > \gamma]} \delta_m(x, y) \frac{d(y,x)}{d(y,x)} f(y) \mu(dy) + f(x) \int_{[y:d(y,x) > \gamma]} \delta_m(x, y) \mu(dy) \\
\leq \frac{1}{\gamma} \sup_{[y:d(y,x) > \gamma]} [\delta_m(x, y)d(y,x)] \int_{[y:d(y,x) > \gamma]} f(y) \mu(dy) \\
+ f(x) \int_{[y:d(y,x) > \gamma]} \delta_m(x, y) \mu(dy) \\
\leq \frac{1}{\gamma} \sup_{[y:d(y,x) > \gamma]} [\delta_m(x, y)d(y,x)] \\
+ f(x) \int_{[y:d(y,x) > \gamma]} \delta_m(x, y) \mu(dy).
\]

which implies that
\[
|I_2(x)| \leq \frac{1}{\gamma} \sup_{[y:d(y,x) > \gamma]} [\delta_m(x, y)d(y,x)] + f(x) \int_{[y:d(y,x) > \gamma]} \delta_m(x, y) \mu(dy).
\]

Assumptions (G2) and (G3) imply that the two terms on the right side of the above inequality tend to zero as \(m \to \infty\).
Note that, for every $\epsilon > 0$, there exists $\gamma > 0$ such that
\[ |f(y) - f(x)| \leq \epsilon \quad \text{if} \quad d(y, x) \leq \gamma, \quad y \in E \]
by the condition (G1). Then there exists $\gamma > 0$ such that
\[ |I_1(x)| \leq \epsilon \int_{|y - x| \leq \gamma} \delta_m(x, y) \mu(dy). \]
Hence
\[ |I_1(x)| \leq \epsilon \sup_{x \in C} \int_{|y - x| \leq \gamma} \delta_m(x, y) \mu(dy) \]
and the term on the right hand side can be made smaller than $2\epsilon$ as $m \to \infty$ by
the condition (G2). Therefore
\[ \lim_{n \to \infty} |E[f_n(x)] - f(x)| = 0. \tag{2.1} \]
Furthermore
\[ E(f_n(x) - f(x))^2 = Var(f_n(x)) + (E[f_n(x)] - f(x))^2. \tag{2.2} \]
In view of (2.1) and (2.2), it is sufficient to prove that
\[ \lim_{n \to \infty} Var(f_n(x)) = 0. \]
Observe that
\[ Var(f_n(x)) = [\hat{n}]^{-2} \sum_{i \in I_n} Var(\delta_m(x, X_i)) \]
\[ = [\hat{n}]^{-1} Var(\delta_m(x, X)) \]
\[ \leq [\hat{n}]^{-1} E[(\delta_m(x, X))^2] \]
\[ = [\hat{n}]^{-1} \int_E \delta_m^2(x, y) f(y) \mu(dy) \]
\[ \leq \frac{S_x}{\hat{n}} \int_E \delta_m(x, y) f(y) \mu(dy) \]
\[ \leq M \frac{S_x}{\hat{n}} \int_E \delta_m(x, y) \mu(dy) \quad \text{(by (G1))} \]
\[ \leq C \frac{S_x}{\hat{n}} \quad \text{(by (G2))} \]
for some constant $C > 0$. The last term tends to zero as $n \to \infty$ by hypothesis. Hence

$$\lim_{n \to \infty} Var(f_n(x)) = 0.$$  

This completes the proof of the theorem.

**Remarks:** Examples of delta sequences in the one dimensional and finite dimensional cases are discussed in Prakasa Rao [15], pp. 136-143 and pp. 218-224 respectively following the work of Walter and Blum [24]. Density estimation for Markov processes using delta sequences is studied in Prakasa Rao [12, 13] and sequential nonparametric estimation of density in the univariate case via delta sequences is investigated in Prakasa Rao [14]. Nonparametric density estimation for functional data by delta sequences is discussed in Prakasa Rao [18]. We have studied the mean square consistency of the estimator $f_n(x)$ as an estimator of $f(x)$. The rate of convergence of the mean square error of a density estimator of this type of estimator is not good in the general case. However, in some special cases, Vieu has observed that the same rate as in the finite dimensional case can be obtained. Related work on the rates of convergence for mean square error is discussed in Vieu [23].

Let $C_0[0,1]$ be the space of real-valued continuous functions $x(.)$ on the interval $[0,1]$ with $x(0) = 0$. Suppose the space $C_0[0,1]$ is equipped with uniform metric induced by the supremum norm

$$||x|| = \sup_{t \in [0,1]} |x(t)|.$$  

Let $\mu(.)$ denote the Wiener measure on the space $C_0[0,1]$ induced by the standard Wiener process and $B_r^x$ be the closed ball with center $x \in C_0[0,1]$ and radius $\frac{1}{r}$. 

8
Define
\[ \delta_m(x, y) = \frac{1}{\mu(B_m^x)} I(y \in B_m^x) \]
where \( I(A) \) denotes the indicator function of the set \( A \). It is easy to check that the corresponding density estimator \( f_n(x) \) is the naive kernel estimator proposed in Basse et al. [1]. This delta family satisfies the condition (G2).

Suppose \( a_n^x \) is a sequence of positive numbers. Let
\[ \delta_m(x, y) = \frac{1}{a_n^x} K_n(d(y, x)) \]
where \( K_n(.) \) is sequence of functions satisfying the conditions (H1)-(H3) in Basse et al. [1]. As was indicated earlier, the choice of \( m \) may depend on \( n \) such that \( m \to \infty \) as \( n \to \infty \). Then we get the kernel estimator proposed in Basse et al. [1]. Other estimators using delta sequences can be constructed following the ideas in Example 2.8.3 in Prakasa Rao [15], p.136 depending on the space \( E \) and the measure \( \mu \). The result proved above generalizes the results in Basse et al. [1], for the kernel type estimators in the i.i.d. case, to estimators obtained through the method of delta families. Results in this paper can also be extended to stationary spatial processes which satisfy a mixing condition and a local dependency condition as in Basse et al. [1]. We will not go into the details here.

References:


