Statistics and Probability Letters 107 (2015) 272–279

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Theoretical investigation of an exploratory approach for log-density in scale-space

Jib Huh [a], Cheolwoo Park [b,*]

[a] *Department of Statistics, Duksung Women's University, Seoul 132-714, Republic of Korea*

[b] *Department of Statistics, University of Georgia, Athens, GA 30602, USA*

ARTICLE INFO

Article history:
Received 1 March 2015
Received in revised form 29 June 2015
Accepted 1 September 2015
Available online 8 September 2015

Keywords:
Asymptotic theory
Exploratory data analysis
Local linear smoothing
Log-density estimation
Scale-space

ABSTRACT

We develop an exploratory data analysis tool in scale-space to discover significant features of a log-density using a local likelihood approach with local polynomial regression estimators. We study its asymptotic properties at multiple locations and levels of resolution.

1. Introduction

Density estimation is one of the fundamental exploratory data analyses, and many methods using parametric or nonparametric approaches have been developed. A nonparametric smoothing approach is a useful and flexible tool for discovering features in data, and typically applies kernel (Silverman, 1986), spline (Gu and Qiu, 1993), or wavelet (Donoho et al., 1996) smoothing techniques depending on assumptions for the underlying distribution. In some cases, it is natural to estimate the logarithm of the density instead of the original function. For example, Kooperberg and Stone (1991) proposed the log-spline smoothing method and Loader (1996) studied the local likelihood density estimation.

As an effective exploratory data analysis tool, Chaudhuri and Marron (1999) developed a scale-space tool, called *SiZer* (SIgnificant ZERo crossing of the derivatives). SiZer utilizes a kernel smoothing approach to estimate the underlying distribution, but unlike traditional kernel smoothing methods, it does not attempt to select a single bandwidth to discover the true underlying structure. Instead, SiZer takes the philosophy of the scale-space approach (Lindeberg, 1994), in which the truth can change depending on which scale a data analyst investigates. Hence, SiZer thoroughly explores the data at multiple scales and reveals all the features found across those scales, some of which might be missing if one uses a single bandwidth. Another advantage of the SiZer tool is that it creates a colored map, *SiZer map*, which reports a statistical inference testing result on the behavior of the estimated function at a particular location and scale. Therefore, not only can a practitioner quickly learn the shape of the underlying structure, but also the practitioner can test whether the estimated shape is real or not at a given scale.

The scale-space approach has been adopted to various research problems and applications. They include multi-dimensional analysis (Godtliebsen et al., 2002, 2004; Ganguli and Wand, 2007; Vaughan et al., 2012), survival data analysis

* Corresponding author.
E-mail address: cpark@uga.edu (C. Park).

http://dx.doi.org/10.1016/j.spl.2015.09.003

(Marron and de Uña Álvarez, 2004), time series analysis (Park et al., 2004; Rondonotti et al., 2007; Olsen et al., 2008a,b; Park et al., 2009a,b), smoothing splines regression (Zhang and Marron, 2005), jump points detection (Kim and Marron, 2006), local likelihood (Li and Marron, 2005; Park and Huh, 2013b), Bayesian methods (Godtliebsen and Øigård, 2005; Erästö and Holmström, 2005; Øigård et al., 2006; Erästö and Holmström, 2007; Holmström and Pasanen, 2012), robust regression (Hannig and Lee, 2006), internet traffic analysis (Park et al., 2007), additive models (Ganguli and Wand, 2007; González-Manteiga et al., 2008), quantile regression (Park et al., 2010), and variance function estimation (Park and Huh, 2013a).

The objective of this study is to develop an exploratory data analysis tool for log-density function. We conduct SiZer inference using the local likelihood approach (Loader, 1996) and construct a visualization map based on the scale-space approach. Then, we study theoretical properties of the proposed SiZer tool for log-density under the frame of Chaudhuri and Marron (2000).

The remainder of the paper is organized as follows. Section 2 proposes a new SiZer tool for log-density estimation and demonstrates its performance in finite samples. Section 3 presents theoretical properties of the proposed SiZer. Appendix provides details on SiZer inference.

2. SiZer for log-density estimation

Section 2.1 proposes SiZer for log-density function and Section 2.2 investigates the performance of SiZer for log-density using simulated examples.

2.1. Proposed siZer

Suppose we have n independent observations $X_1, \ldots, X_n$ having unknown density f with support $(-\infty, \infty)$. A usual kernel density estimator of f is given as

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x),$$

where h is the bandwidth (scale), $K_h(\cdot) = K(\cdot/h)/h$, and K is a kernel function. In SiZer, the target is shifted from the true underlying curve f to the smoothed version of the underlying function $f_h(x) = \int f(u)K_h(x-u)du$, which is the basic idea behind a scale-space approach. Hence, the objective of the proposed method is to determine the significance of trends or modes in the function $f_h(x)$ at a particular location x and scale h and summarize all the information that is available at each scale.

Local likelihood was introduced by Tibshirani and Hastie (1987) as a method of smoothing by local polynomials in non-Gaussian regression models. Loader (1996) extended it to density estimation. We develop a SiZer tool for discovering significant features of a density based on a local likelihood approach with local polynomial regression fit.

Our focus of this study is to estimate $\log f_h$ instead of f_h. Let $g(x) = \log f(x)$ and $g_h(x) = \log f_h(x)$. Define $\widehat{g}_h(x) = \widehat{\beta}_0$ and $\widehat{g}'_h(x) = \widehat{\beta}_1$ as the estimators for $g_h(x)$ and $g'_h(x)$ where $(\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)$ maximizes the following kernel weighted local likelihood function (Loader, 1996):

$$\sum_{i=1}^{n} K\Big(\frac{X_i - x}{h}\Big)P(X_i - x) - n\int_0^1 K\Big(\frac{u-x}{h}\Big)\exp(P(u-x))du, \tag{2.1}$$

where $P(u-x) = \beta_0 + \beta_1(u-x) + \cdots + \beta_p(u-x)^p$. We note that no explicit solutions to the maximization (2.1) exist except when $p = 0$.

In our simulation, we use the local linear with $p = 1$. In particular, we solve the following local likelihood equations with respect to β_0 and β_1:

$$\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \dfrac{X_i - x}{h}\end{pmatrix} K\left(\frac{X_i - x}{h}\right) = \int_0^1 \begin{pmatrix} 1 \\ \dfrac{u-x}{h}\end{pmatrix} K\left(\frac{u-x}{h}\right)\exp(\beta_0 + \beta_1(u-x))du. \tag{2.2}$$

Then, $\hat{\beta}_1 = \hat{g}'_h(x)$ can be obtained by solving the following equation:

$$\int_0^1 (u-x)K\left(\frac{u-x}{h}\right)e^{\hat{\beta}_1 u}du = \frac{\sum_{i=1}^{n}(X_i - x)K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n}K\left(\frac{X_i - x}{h}\right)}\int_0^1 K\left(\frac{u-x}{h}\right)e^{\hat{\beta}_1 u}du,$$

and

$$\hat{g}_h(x) = \hat{\beta}_0 = \frac{e^{\hat{\beta}_1 x}\sum_{i=1}^{n}K\left(\frac{X_i - x}{h}\right)}{n\int_0^1 K\left(\frac{u-x}{h}\right)e^{\hat{\beta}_1 u}du}.$$

Since we attempt to find significant features in the smoothed curve at a given level of resolution, the null hypothesis in scale-space is given as $H_0 : g_h'(x) = 0$ for each location x and each scale h, i.e. no trend at (x, h). Then, the corresponding $100(1-\alpha)\%$ confidence interval for $g_h'(x)$ is given as

$$\hat{g}_h'(x) \pm q(h)\widehat{SD}(\hat{g}_h'(x)) \tag{2.3}$$

where the quantile $q(h)$ depending on h is given as

$$q(h) = \Phi^{-1}\left(\left(1-\frac{\alpha}{2}\right)^{1/(\theta n_p)}\right) \tag{2.4}$$

where Φ is the cumulative distribution function of the standard normal and n_p is the number of pixels in each row of a SiZer map. The θ, called the cluster index, measures the equivalent number of independent observations and is given as

$$\theta = 2\Phi\left(\sqrt{3\log n_p}\,\frac{\tilde{\Delta}}{2h}\right) - 1,$$

where $\tilde{\Delta}$ denotes the distance between the pixels of the SiZer map for a given scale. This quantile is derived by addressing the multiple testing problem across different locations for a given h. One can obtain the quantile which considers both location and scale into multiple testing adjustment, but it is known to be less powerful (Hannig and Marron, 2006).

One can show that the estimate of the standard deviation (SD) is given as

$$\widehat{SD}(\hat{g}_h'(x)) \approx \sqrt{\frac{1}{nh^3 f(x)}\int v^2 (K(v))^2 dv}.$$

We use the standard normal for the kernel function K due to the monotonicity property in scale-space maps (Chaudhuri and Marron, 2000), and the estimate of the standard deviation can be expressed as

$$\widehat{SD}(\hat{g}_h'(x)) \approx \sqrt{\frac{1}{4\sqrt{\pi}\,nh^3 f(x)}}. \tag{2.5}$$

The details of (2.5) are given in the Appendix.

SiZer visually reports the results of statistical inference using the confidence intervals in (2.3) as a color map, called SiZer map. In a map, the horizontal axis represents location x and the vertical axis the logarithm of the scale, $\log h$. Each pixel (x, h) displays one of the following colors: black, white, intermediate or dark gray. It is colored black if the confidence interval in (2.3) is greater than zero. This implies that the estimated function $\hat{g}_h(x)$ is increasing at the corresponding location x and scale h. Similarly, it is colored white if the confidence interval is less than zero, which implies that the estimated function $\hat{g}_h(x)$ is decreasing at (x, h). It is colored intermediate gray if the confidence interval includes zero, implying that the null hypothesis $H_0 :\ g_h'(x) = 0$ is not rejected at (x, h). Finally, the pixel is colored dark gray if no decision can be made due to the insufficient data points. Chaudhuri and Marron (1999) proposed to use the estimated effective sample size (ESS) for each (x, h):

$$\mathrm{ESS}(x, h) = \frac{\sum_{i=1}^{n} K_h(X_i - x)}{K_h(0)}$$

to determine the dark gray region. If ESS $(x, h) < 5$, then the test is not carried out and the corresponding pixel is colored dark gray.

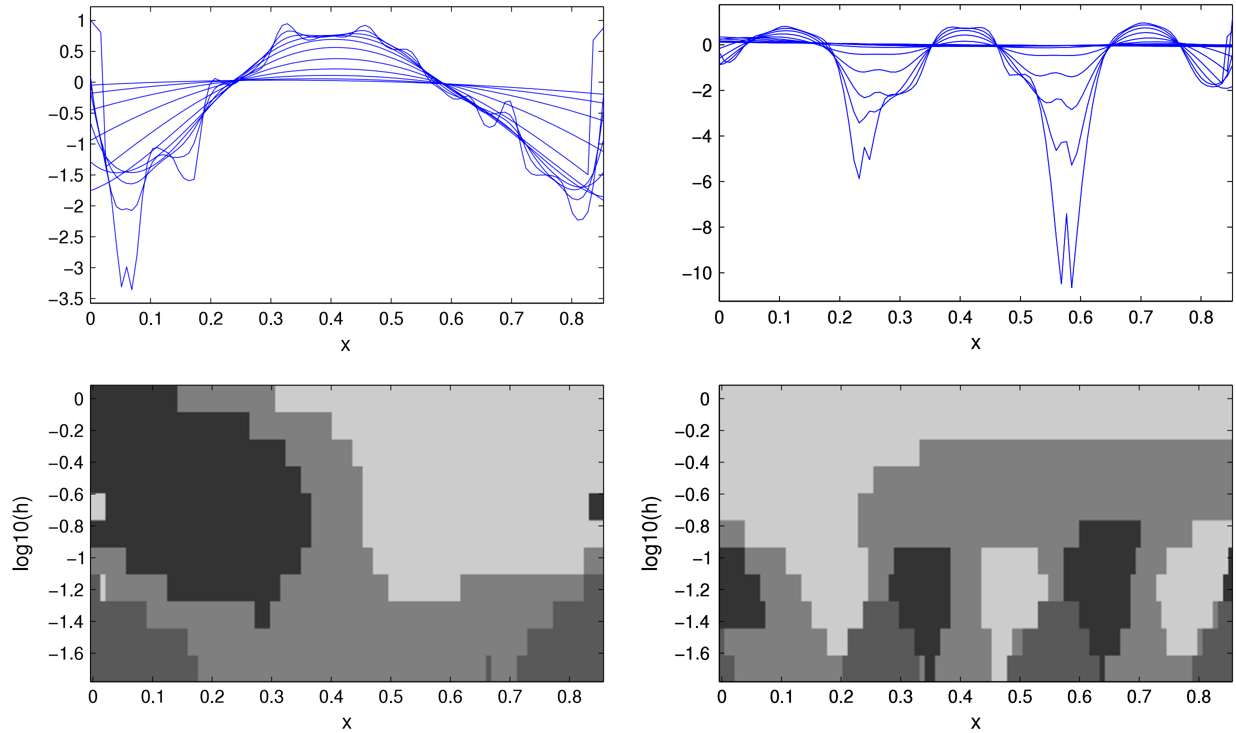
2.2. Simulation

We perform a simulation study to demonstrate the performance of the proposed SiZer. We generate simulated data from one of the following densities:

(i) truncated normal: $p_1(x) = c_1\phi\{6(x-0.5)\}I_{[0,1]}(x)$
(ii) tri-modal: $p_2(x) = c_2\left[\frac{7}{20}\phi\{20(x-0.2)\} + \frac{6}{20}\phi\{20(x-0.5)\} + \frac{7}{20}\phi\{20(x-0.8)\}\right]I_{[0,1]}(x)$

with the sample size $n = 200$. Here, c_1 and c_2 are scaling constants. Note that the first function is uni-modal while the second has three modes. For h, 11 equally spaced values are chosen between $2(X_{\max} - X_{\min})/(n_p - 1)$ and $(X_{\max} - X_{\min})$ on a logarithmic scale with base 10.

Fig. 1 displays SiZer plots for two simulated examples. The top panels, called family of smooths, consist of 11 smoothed curves $\hat{g}_h(x)$ with 11 different bandwidth values. If h is small, the estimated curve shows many fluctuations due to the undersmoothing effect. On the other hand if h is large, the estimated curve is close to a straight line due to the oversmoothing effect. There are some bandwidths that reveal a mode in Fig. 1(a) and three modes in Fig. 1(b), but it should be noted that the



(a) Truncated normal.

(b) Tri-modal.

Fig. 1. We apply the proposed SiZer to two simulated examples: (a) truncated normal and (b) tri-modal densities.

goal of SiZer analysis does not attempt to find the best scale. Instead, it explores all scales and finds features on those scales. The bottom panels display SiZer maps that return the result of hypothesis testing $H_0 : g_h'(x) = 0$ at each (x, h) using the 95% confidence interval defined in (2.3). In the map of Fig. 1(a), black is colored in the first half region and white in the second half at large and medium scales. This describes the significance of the single mode (overall increasing and decreasing trend) in the underlying density on those scales. For small scales, the features are statistically insignificant or no decision is made due to lack of neighboring data points. In the map of Fig. 1(b), three modes are found statistically significant at medium and small scales. For large scales, only one mode is detected due to oversmoothing. In both cases, small boundary effects are found on both edges when medium scales are used, which might be a drawback of the local likelihood approach.

3. Asymptotic results

In this section, we present asymptotic properties of the proposed SiZer for log-density, which provides its theoretical justification. In particular, we describe statistical convergence of the difference between the empirical and the theoretical scale-space surfaces $(\widehat{g}_h^{(m)}(x), E\{\widehat{g}_h^{(m)}(x)\})$ for fixed $m = 0, 1, \dots,$ under the frame of Chaudhuri and Marron (2000) who addressed these issues for the estimation of the original density f. We derive asymptotic properties of the proposed SiZer using local polynomial regression with degree $p \geq m$. A similar theoretical study of SiZer for log-variance function estimation and generalized linear models is done in Park and Huh (2013a) and Park and Huh (2013b), respectively.

Before presenting our main results, we provide Lemma 1, which describes an approximation form of $\widehat{g}_{h(x)}^{(m)}$. Define $\boldsymbol{u} = (1, (u-x)/h, \dots, (u-x)^p/h^p)^T$, $\boldsymbol{\mu}_{h,x} = \int_0^1 \boldsymbol{u} K((u-x)/h)du$ and $S_{h,x} = \int_0^1 \boldsymbol{u}\boldsymbol{u}^T K((u-x)/h)du$. Let $\boldsymbol{e}_j$ be the $(p+1) \times 1$ vector with 1 appearing at the jth position and 0 otherwise. Also let

$$W_{n,m}(h, x, X_j) = h^m \boldsymbol{e}_{m+1}^T S_{h,x}^{-1}\left(\frac{1}{f(x)} K\left(\frac{X_j - x}{h}\right)\boldsymbol{Z}_j - \boldsymbol{\mu}_{h,x}\right)$$

where $\boldsymbol{Z}_j = (1, (X_j - x)/h, \dots, (X_j - x)^p/h^p)^T, j = 1, \dots, n$. Let I and H be compact subintervals of $(-\infty, \infty)$ and $(0, \infty)$, respectively. In order to obtain the approximation form of $\widehat{g}_h^{(m)}(x)$ uniformly in $h \in H$ and $x \in I$ in Lemma 1, we need the following assumptions.

(A.1) The function $g^{(p+1)}$ is uniformly continuous over I.
(A.2) $\inf_{x \in I} f(x) > 0$.
(A.3) There exists $S_{h,x}^{-1}$ for any $h \in H$ and $x \in I$.

We note that (A.1) is needed to obtain the bias term using the local pth polynomial fit, and to guarantee the asymptotic form in Lemma 1 uniformly in $x \in I$ and $h \in H$ (Fan et al., 1995; Park et al., 2009a; Park and Huh, 2013a,b). The assumptions (A.2) and (A.3) are needed to ensure that the denominator in $W_{n,m}(h, x, X_j)$ is away from 0 at x in the compact set I.

Lemma 1. *Let m be a non-negative integer. Suppose that the conditions* (A.1)–(A.3) *are satisfied.*

(i) *If* $m = 0$, *then*

$$\widehat{g}_h(x) - g(x) = \frac{1}{n}\sum_{i=1}^{n} W_{n,0}(h, x, X_i)(1 + o_P(1))$$

uniformly in $h \in H$ *and* $x \in I$ *as* $n \to \infty$.

(ii) *If* $m > 0$, *then*

$$\widehat{g}_h^{(m)}(x) = \frac{1}{n}\sum_{i=1}^{n} W_{n,m}(h, x, X_i)(1 + o_P(1))$$

uniformly in $h \in H$ *and* $x \in I$ *as* $n \to \infty$.

Proof. Define $\bar{\boldsymbol{\beta}} = \sqrt{n}(\widehat{\beta}_0 - g(x), h\widehat{\beta}_1, \ldots, h^p\widehat{\beta}_p)^T$ where $(\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)$ is the maximizer of (2.1). Note that $\sum_{i=0}^{p} \widehat{\beta}_i(u - x)^i = g(x) + \bar{\boldsymbol{\beta}}^T \boldsymbol{u}/\sqrt{n}$. According to (2.1), $\bar{\boldsymbol{\beta}}$ maximizes

$$\sum_{i=1}^{n} K\Big(\frac{X_i - x}{h}\Big)\{g(x) + \boldsymbol{\beta}^{*T}\boldsymbol{Z}_i/\sqrt{n}\} - n\int_0^1 K\Big(\frac{u-x}{h}\Big)\exp(g(x) + \boldsymbol{\beta}^{*T}\boldsymbol{u}/\sqrt{n})du$$

as a function of $\boldsymbol{\beta}^*$. Loader (1996) showed that if the maximizer of (2.1) exists, it is the solution of the following system of local likelihood equation:

$$\frac{1}{n}\sum_{j=1}^{n} K\Big(\frac{X_j - x}{h}\Big)\boldsymbol{Z}_j = \int_0^1 \boldsymbol{u}K\Big(\frac{u-x}{h}\Big)\exp(g(x) + \bar{\boldsymbol{\beta}}^T\boldsymbol{u}/\sqrt{n})du. \tag{3.1}$$

Using a Taylor series expansion with the assumption (A.1), the right-hand side in (3.1) is

$$\int_0^1 \boldsymbol{u}K\Big(\frac{u-x}{h}\Big)e^{g(x)}du + \int_0^1 \boldsymbol{u}K\Big(\frac{u-x}{h}\Big)e^{g(x)}\frac{1}{\sqrt{n}}\bar{\boldsymbol{\beta}}^T\boldsymbol{u}du(1 + o_P(1))$$

uniformly in $h \in H$ and $x \in I$. The asymptotic expression of Eq. (3.1) becomes

$$\frac{1}{n}\sum_{j=1}^{n} K\Big(\frac{X_j - x}{h}\Big)\boldsymbol{Z}_j = \bar{\boldsymbol{\beta}}\frac{f(x)}{\sqrt{n}}\int_0^1 \boldsymbol{u}\boldsymbol{u}^T K\Big(\frac{u-x}{h}\Big)du + f(x)\int_0^1 \boldsymbol{u}K\Big(\frac{u-x}{h}\Big)du. \tag{3.2}$$

Then, the solution of (3.2) is given by

$$\bar{\boldsymbol{\beta}} = \frac{1}{\sqrt{n}}S_{h,x}^{-1}\Big(\frac{1}{f(x)}K\Big(\frac{X_j - x}{h}\Big)\boldsymbol{Z}_j - n\boldsymbol{\mu}_{h,x}\Big)$$

uniformly in $h \in H$ and $x \in I$ under the assumptions (A.2) and (A.3). Therefore, we obtain the result. ■

We define the approximation form of $\widehat{g}_h(x) - g(x)$ or $\widehat{g}_h^{(m)}(x)$ as

$$\tilde{g}_h^{(m)}(x) = \frac{1}{n}\sum_{i=1}^{n} W_{n,m}(h, x, X_i). \tag{3.3}$$

In what follows we present two main theorems. Theorem 1 shows that the empirical scale-space surfaces weakly converge to their theoretical scale-space surfaces. Theorem 2 states the uniform convergence of the difference between the empirical and the theoretical scale-space surfaces under the supremum norm. In order to prove the two theorems, the following four additional assumptions for the weight $W_{n,m}(h, x, u)$ in (3.3) are necessary. Chaudhuri and Marron (2000) point out that these assumptions are satisfied by standard kernels such as the Gaussian kernel.

(A.4) For a non-negative integer $m \geq 0$, assume that as $n \longrightarrow \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}[\{W_{n,m}(h_1, x_1, X_i) - E(W_{n,m}(h_1, x_1, X_i))\} - \{W_{n,m}(h_2, x_2, X_i) - E(W_{n,m}(h_2, x_2, X_i))\}]$$

converges in probability to a covariance function $cov(h_1, x_1, h_2, x_2)$ for all (h_1, x_1) and $(h_2, x_2) \in H \times I$.

(A.5) Assume that as $n \longrightarrow \infty$,

$$n^{-(1+\rho/2)}\left\{\max_{1\le i\le n}|W_{n,m}(h,x,X_i)|^{\rho}\right\}\sum_{i=1}^{n}\{W_{n,m}(h,x,X_i)\}^2$$

converges in probability to 0 for all $(h,x)\in H\times I$ and for some $\rho>0$.

(A.6) Assume that as h varies in H and x varies in I, $\{\partial^2 W_{n,m}(h,x,X_i)/(\partial h\partial x)\}^2$ is uniformly dominated by a positive function $M(X_i)$ such that $E(M(X_i))<\infty$.

(A.7) Assume that as h varies in H and x varies in I, $\{\partial W_{n,m}(h,x,X_i)/\partial h\}^2$ and $\{\partial W_{n,m}(h,x,X_i)/\partial x\}^2$ are uniformly dominated by a positive function $M^*(X_i)$ such that $E(M^*(X_i))<\infty$.

Theorem 1. *Suppose that the assumptions* (A.1)–(A.6) *are satisfied. Then, the two-parameter stochastic process*

$$U_n(h,x)=n^{1/2}\left[\hat{g}_h^{(m)}(x)-E\{\hat{g}_h^{(m)}(x)\}\right]$$

with $(h,x)\in H\times I$ converges weakly to a Gaussian process on $H\times I$ with zero mean and covariance function $cov(h_1,x_1,h_2,x_2)$ in (A.4) *as $n\longrightarrow\infty$.*

Proof. First, we show that all the finite dimensional distributions of the process weakly converge to a normal distribution. Let us fix $(h_1,x_1),(h_2,x_2),\ldots,(h_k,x_k)\in H\times I$ and $t_1,t_2,\ldots,t_k\in(-\infty,\infty)$. Define

$$Z_n=n^{1/2}\sum_{i=1}^{k}t_i\left[\widehat{g}_{h_i}^{(m)}(x_i)-E\{\widehat{g}_{h_i}^{(m)}(x_i)\}\right]\quad\text{and}\quad \tilde{Z}_n=n^{1/2}\sum_{i=1}^{k}t_i\left[\tilde{g}_{h_i}^{(m)}(x_i)-E\{\tilde{g}_{h_i}^{(m)}(x_i)\}\right].$$

Note that $Z_n=\tilde{Z}_n(1+o_P(1))$ for any fixed $h_i\in H$ and $x_i\in I$ by Lemma 1. Then, the mean of $\tilde{Z}_n$ is 0, and its variance is given as

$$\sum_{i=1}^{k}\sum_{j=1}^{k}t_it_j\frac{1}{n}\sum_{l=1}^{n}E\Big[\{W_{n,m}(h_i,x_i,X_l)-E(W_{n,m}(h_i,x_i,X_l))\}\{W_{n,m}(h_j,x_j,X_l)-E(W_{n,m}(h_j,x_j,X_l))\}\Big].$$

We note that the variance converges to $\sum_{i=1}^{k}\sum_{j=1}^{k}t_it_jcov(h_i,x_i,h_j,x_j)$ in probability as $n\to\infty$ by the assumption (A.4). Since the Lindeberg's condition holds for $\tilde{Z}_n$ by the assumption (A.5), the limiting distribution is normal. Consider the following sequence of two parameter process

$$\tilde{U}_n(h,x)=n^{1/2}\left[\tilde{g}_h^{(m)}(x)-E\{\tilde{g}_h^{(m)}(x)\}\right].$$

Then, we have

$$U_n(h,x)=\tilde{U}_n(h,x)(1+o_P(1))$$

uniformly in $h\in H$ and $x\in I$. Therefore, by the application of Cramer–Wold device, the joint limiting distribution of $\tilde{U}_n(h_i,x_i)$ for $1\le i\le k$ as $n\to\infty$ is obtained as multivariate normal with zero mean and the covariance matrix with (i,j)th entry $cov(h_i,x_i,h_j,x_j)$ for $1\le i,j\le k$.

Second, we show that the process has the tightness property. Let us fix $h_1<h_2$ in H and $x_1<x_2$ in I. Then,

$$\begin{aligned}E\left\{\tilde{U}_n(h_2,x_2)-\tilde{U}_n(h_2,x_1)-\tilde{U}_n(h_1,x_2)+\tilde{U}_n(h_1,x_1)\right\}^2&\le C(h_2-h_1)^2(x_2-x_1)^2E\left\{n^{-1}\sum_{i=1}^{n}M(X_i)\right\}\\&\le C(h_2-h_1)^2(x_2-x_1)^2.\end{aligned}$$

Here, C is a generic positive constant which varies from place to place. Using the same argument in Bickel and Wichura (1971), one can show that the sequence of two parameter process $\tilde{U}_n(h,x)$ on $H\times I$ satisfies the tightness condition, which completes the proof. ∎

Theorem 2. *Suppose that the assumptions* (A.1)–(A.7) *are satisfied. Then as $n\longrightarrow\infty$,*

$$\sup_{x\in I,h\in H}n^{1/2}\left|\widehat{g}_h^{(m)}(x)-E\{\widehat{g}_h^{(m)}(x)\}\right|$$

converges weakly to a random variable whose distribution is the same as that of $\sup_{x\in I,h\in H}|Z(h,x)|$ where $Z(h,x)$ is a Gaussian process with zero mean and covariance function $cov(h_1,x_1,h_2,x_2)$ in Theorem 1 so that

$$Pr\left(\sup_{x\in I,h\in H}|Z(h,x)|<\infty\right)=1.$$

Proof. For (h_1, x_1) and (h_2, x_2) in $H \times I$, since

$$E\{\tilde{U}_n(h_2, x_2) - \tilde{U}_n(h_1, x_1)\}^2 \le C\{(h_2 - h_1)^2 + (x_2 - x_1)^2\}$$

by (A.4), one obtains

$$E\{Z(h_2, x_2) - Z(h_1, x_1)\}^2 \le C\{(h_2 - h_1)^2 + (x_2 - x_1)^2\}.$$

The rest of proof is similar to that of Theorem 3.3 in Chaudhuri and Marron (2000). ■

Acknowledgment

The research of the first author was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science, ICT & Future Planning (No. NRF-2011-0010856).

Appendix

In this section, we show the standard deviation of $\widehat{g}_h'(x)$ in (2.5) in the case of the local linear fit with $p = 1$. Let $\boldsymbol{V}$ be a random vector defined in Lemma 1 of Loader (1996). According to Theorem 2 in Loader (1996), the asymptotic variance of $h(\widehat{g}_h'(x) - g'(x))$ can be written as

$$\frac{1}{h^2 f^2(x)} \boldsymbol{e}_2^T M_1^{-1} \mathrm{Var}(\boldsymbol{V}) M_1^{-1} \boldsymbol{e}_2 \tag{A.4}$$

when $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Here $\boldsymbol{e}_2 = (0, 1)^T$,

$$M_1 = \begin{pmatrix} \mu_0(K) & \mu_1(K) \\ \mu_1(K) & \mu_2(K) \end{pmatrix} \quad \text{and} \quad \mathrm{Var}(\boldsymbol{V}) = \frac{h}{n} f(x) \begin{pmatrix} \mu_0(K^2) & \mu_1(K^2) \\ \mu_1(K^2) & \mu_2(K^2) \end{pmatrix}$$

where $\mu_j(K)$ and $\mu_j(K^2)$ are the jth moments of K and K^2, respectively. In the case of using the Gaussian kernel, M_1 is the identity matrix and

$$\mathrm{Var}(\boldsymbol{V}) = \frac{h}{n} f(x) \begin{pmatrix} 1 & 0 \\ 0 & (4\sqrt{\pi})^{-1} \end{pmatrix}.$$

Then, (A.4) becomes the standard deviation in (2.5).

References

Bickel, P.J., Wichura, M.J., 1971. Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist. 42, 1656–1670.
Chaudhuri, P., Marron, J.S., 1999. SiZer for exploration of structures in curves. J. Amer. Statist. Assoc. 94, 807–823.
Chaudhuri, P., Marron, J.S., 2000. Scale space view of curve estimation. Ann. Statist. 28, 408–428.
Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D., 1996. Density estimation by wavelet thresholding. Ann. Statist. 24, 508–539.
Erästö, P., Holmström, L., 2005. Bayesian multiscale smoothing for making inferences about features in scatter plots. J. Comput. Graph. Statist. 14, 569–589.
Erästö, P., Holmström, L., 2007. Bayesian analysis of features in a scatter plot with dependent observations and errors in predictors. J. Stat. Comput. Simul. 77, 421–431.
Fan, J., Heckman, N.E., Wand, M.P., 1995. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. J. Amer. Statist. Assoc. 90, 141–150.
Ganguli, B., Wand, M.P., 2007. Feature significance in generalized additive models. Stat. Comput. 17, 179–192.
Godtliebsen, F., Marron, J.S., Chaudhuri, P., 2002. Significance in scale space for bivariate density estimation. J. Comput. Graph. Statist. 11, 1–21.
Godtliebsen, F., Marron, J.S., Chaudhuri, P., 2004. Statistical significance of features in digital images. Image Vis. Comput. 22, 1093–1104.
Godtliebsen, F., Øigård, T.A., 2005. A visual display device for significant features in complicated signals. Comput. Statist. Data Anal. 48, 317–343.
González-Manteiga, W., Martinez-Miranda, M.D., Raya-Miranda, R., 2008. SiZer map for inference with additive models. Stat. Comput. 18, 297–312.
Gu, C., Qiu, C., 1993. Smoothing spline density estimation: Theory. Ann. Statist. 21, 217–234.
Hannig, J., Lee, T., 2006. Robust SiZer for exploration of regression structures and outlier detection. J. Comput.Graph. Statist. 15, 101–117.
Hannig, J., Marron, J.S., 2006. Advanced distribution theory for SiZer. J. Amer. Statist. Assoc. 101, 484–499.
Holmström, L., Pasanen, L., 2012. Bayesian scale space analysis of differences in images. Technometrics 54, 16–29.
Kim, C.S., Marron, J.S., 2006. Sizer for jump detection. J. Nonparametr. Stat. 18, 13–20.
Kooperberg, C., Stone, C.J., 1991. A study of logspline density estimation. Comput. Statist. Data Anal. 12, 327–347.
Li, R., Marron, J.S., 2005. Local likelihood SiZer map. Sankhya 67, 476–498.
Lindeberg, T., 1994. Scale-Space Theory in Computer Vision. Kluwer, Boston.
Loader, C.R., 1996. Local likelihood density estimation. Ann. Statist. 24, 1602–1618.
Marron, J.S., de Uña Álvarez, J., 2004. SiZer for length biased, censored density and hazard estimation. J. Statist. Plann. Inference 121, 149–161.
Øigård, T.A., Rue, H., Godtliebsen, F., 2006. Bayesian multiscale analysis for time series data. Comput. Statist. Data Anal. 51, 1719–1730.
Olsen, L.R., Chaudhuri, P., Godtliebsen, F., 2008a. Multiscale spectral analysis for detecting short and long range change points in time series. Comput. Statist. Data Anal. 52, 3310–3330.
Olsen, L.R., Sorbye, S.H., Godtliebsen, F., 2008b. A scale-space approach for detecting non-stationarities in time series. Scand. J. Stat. 35, 119–138.
Park, C., Godtliebsen, F., Taqqu, M., Stoev, S., Marron, J.S., 2007. Visualization and inference based on wavelet coefficients, SiZer and SiNos. Comput. Statist. Data Anal. 51, 5994–6012.
Park, C., Hannig, J., Kang, K., 2009a. Improved SiZer for time series. Statist. Sinica 19, 1511–1530.
Park, C., Huh, J., 2013a. Nonparametric estimation of a log-variance function in scale-space. J. Statist. Plann. Inference 143, 1766–1780.
Park, C., Huh, J., 2013b. Statistical inference and visualization in scale-space using local likelihood. Comput. Statist. Data Anal. 57, 336–348.

Park, C., Lee, T., Hannig, J., 2010. Multiscale exploratory analysis of regression quantiles using quantile SiZer. J. Comput. Graph. Statist. 19, 497–513.
Park, C., Marron, J.S., Rondonotti, V., 2004. Dependent SiZer: goodness of fit tests for time series models. J. Amer. Statist. Assoc. 31, 999–1017.
Park, C., Vaughan, A., Hannig, J., Kang, K., 2009b. Sizer analysis for the comparison of time series. J. Statist. Plann. Inference 139, 3974–3988.
Rondonotti, V., Marron, J.S., Park, C., 2007. SiZer for time series: a new approach to the analysis of trends. Electron. J. Stat. 1, 268–289.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.
Tibshirani, R., Hastie, T., 1987. Local likelihood estimation. J. Amer. Statist. Assoc. 82, 559–567.
Vaughan, A., Jun, M., Park, C., 2012. Statistical inference and visualization in scale-space for spatially dependent images. J. Korean Stat. Soc. 41, 115–135.
Zhang, J.-T., Marron, J.S., 2005. SiZer for smoothing splines. Comput. Statist. 20, 481–502.