Scientific
Research
Publishing

On the Approximation of Maximum Deviation Spline Estimation of the Probability Density Gaussian Process

Mukhammadjon S. Muminov[1*], Kholiqjon S. Soatov[2]

[1]Institute of Mathematics, National University of Uzbekistan, Tashkent, Uzbekistan
[2]Tashkent University of Information Technologies, Tashkent, Uzbekistan
Email: [*]m.muhammad@rambler.ru

Abstract

In the paper, the deviation of the spline estimator for the unknown probability density is approximated with the Gauss process. It is also found zeros for the infimum of variance of the derivation from the approximating process.

Keywords

1. Introduction

The present work is a continuation of the work [1], that's why we use notations admitted in it. We shall not turn our attention to more detailed review because it is given [1].

Let $X_1, X_2, \cdots, X_n$ be a simple sample from the parent population with the probability density $f(t)$ concentrated and continuous on the segment $[0,1]$. Let $S_n(x)$ be a cubic spline interpolating values $y_k = F_n(t_k)$ at the points $x_k = kh$, $k = 0, \cdots, N$ with the boundary conditions

$$S_n'(0) = \frac{y_1 - y_0}{h}, \quad S_n'(1) = \frac{y_N - y_{N-1}}{h}$$

where $N = N(n)$, $h = 1/N$, $h \to 0$, $nh \to \infty$ as $n \to \infty$.

Remind that

---
[*]Corresponding author.

$$\xi_n = \sqrt{nh} \max_{0 \le x \le 1} \left| \frac{S_n'(x) - f(x)}{\sigma_N(x)\sqrt{f(x)}} \right|,$$

$$\xi_n^*(x) = \sqrt{nh} \max_{0 \le x \le 1} \frac{S_n'(x) - ES_n'(x)}{\sigma_N(x)\sqrt{f(x)}},$$

$$\eta_n(x) = \frac{1}{\sigma_N(x)\sqrt{h}} \int_0^1 W_N(x, y) \, d\omega_n(y),$$

where $\sigma_N^2(x) = \dfrac{1}{h} \int_0^1 W_N^2(x, y) \, dy$, $W_N(x, y)$ is the kernel of the spline, see [1], $\{\omega_n(t), \ t \in [0,1]\}$ is a sequence of Wiener processes.

Denote by $Q_n(y)$ the distribution function of the random variable

$$l_n \max_{0 \le x \le 1} \left| \xi_n^*(x) \right| - l_n^2,$$

and by $Q_n^{(1)}(y)$ the distribution functions of the random variable

$$l_n \max_{0 \le x \le 1} \left| \eta_n(x) \right| - l_n^2,$$

where

$$l_n = \sqrt{2 \log N} \left(1 + g_n\right), \quad \lim_{n \to \infty} g_n = 0. \tag{1}$$

In the second section of the work, Theorem 2 and 3 are proven:

$$l_n \max_{0 \le x \le 1} \left| \xi_n^*(x) \right| - l_n^2 \xrightarrow{D} l_n \max_{0 \le x \le 1} \left| \eta_n(x) \right| - l_n^2,$$

and

$$l_n \xi_n - l_n^2 \xrightarrow{D} l_n \max_{0 \le x \le 1} \left| \xi_n^*(x) \right| - l_n^2,$$

And it is also stated (Theorem 5) that

$$\inf_{0 \le x \le 1} Var \, \eta_n'(x) = 0.$$

2. Formulation and Proof the of Main Results

It holds the following

Theorem 1. Let ξ and η be random variables, in addition $P(|\xi - \eta| > \varepsilon) < \delta$ for some $\varepsilon > 0$, $\delta > 0$. Then for any x

$$P(\eta < x - \varepsilon) - \delta \le P(\xi < x) \le P(\eta < x + \varepsilon) + \delta.$$

The proof of this statement is easy, therefore we omit it.

Theorems 2 and Theorem 3 will be proved by the mthods given in [2].

Theorem 2. Let $h = n^{-\delta}$, $0 < \delta < 1$, and there exist a constant $\gamma > 0$ such that

$$n^{-\gamma} \ge \log n^{\frac{3}{2}} n^{-\frac{1-\delta}{2}} \vee \log n^{\frac{3}{2}} n^{-\frac{\delta}{2}}. \tag{2}$$

Then under our assumption a) and b) concerning $f(x)$, there exists a constant $C_{18} > 0$ such that for sufficiently large n

$$Q_n^{(1)}\left(x - C_{18}n^{-\gamma}\right) - C_{18}n^{-\gamma} < Q_n(x) \le Q_n^{(1)}\left(x + C_{18}n^{-\gamma}\right) + C_{18}n^{-\gamma}.$$

Proof. By the main Theorem from [1],

$$l_n \left| \max_{0 \le x \le 1} \left| \xi_n^{(1)}(x) + \xi_n^{(2)}(x) \right| - \max_{0 \le x \le 1} \left| \xi_n^{(1)}(x) \right| \right| \le l_n \max_{0 \le x \le 1} \left| \xi_n^{(2)}(x) \right|, \tag{3}$$

and for any $\varepsilon > 0$

$$P\left(l_n \left| \max_{0 \le x \le 1} \left| \eta_n(x) + \eta_n^{(1)} \right| - \max_{0 \le x \le 1} \left| \eta_n(x) \right| \right| > \varepsilon \right) \le \frac{C_4 h}{\log N} + C_5 \varepsilon^{-1} \sqrt{h} \exp\left\{ -C_6 \varepsilon^2 h^{-1} l_n^{-2} \right\}. \tag{4}$$

Set $\varepsilon = \varepsilon_n = C_{20} \sqrt{n \log N} \cdot l_n$.

Theorem 2 follows now from Theorem 1, relations (2) from [1], inequalities (3) and (4), and the fact that the random variables $l_n \max_{0 \le x \le 1} \left| \xi_n^{(1)}(x) \right| - l_n^2$ and $l_n \max_{0 \le x \le 1} \left| \eta_n(x) + \eta_n^{(1)}(x) \right| - l_n^2$ have the same distribution.

Theorem 3. If conditions of Theorem 2 hold and $\dfrac{1}{3} < \delta < 1$, then for sufficiently large n

$$Q_n \left(x - C_1^* n^{-\gamma} \right) - C_1^* n^{-\gamma} < G_n(x) \le Q_n \left(x + C_1^* n^{-\gamma} \right) + C_1^* n^{-\gamma},$$

where $C_1^* > 0$ is a constant, $G_n(x) = P\left(l_n \xi_n - l_n^2 < x \right)$, γ is defined in (2).

Proof. From the interpolation condition

$$S_n(x_k) = F_n(x_k), \ k = \overline{0, N}$$

we have

$$ES_n(x_k) = F(x_k), \ k = \overline{0, N}.$$

One can easily note that $MS_n(x)$ is a cubical spline interpolating of

$$y_k = F(x_k), \ k = 0, 1, \cdots, N,$$

in the points of interpolation $x_k = kh$, $k = \overline{0, N}$. On the other hand $ES_n'(x) = \left(ES_n(x) \right)'$. By Theorem 9 from the monograph [3] we get

$$\left| ES_n'(x) - f(x) \right| \le 5h\beta_n, \tag{5}$$

where

$$\beta_n = \frac{2}{5} \omega(f', h) + \max\left\{ \frac{3}{5} f'(0), \frac{3}{5} f'(1) \right\},$$

$$\omega(f', h) = \max_{|x - y| \le h} \left| f'(x) - f'(y) \right|, \ 0 \le x, y \le 1.$$

The relation (5) implies that for arbitrary $\varepsilon > 0$

$$P\left(l_n \left| \xi_n - \max_{0 \le x \le 1} \left| \xi_n^*(x) \right| \right| > \varepsilon \right)$$

$$= P\left(l_n \left| \sqrt{nh} \max_{0 \le x \le 1} \left| \frac{S_n'(x) - f(x)}{\sigma_N(x) \sqrt{f(x)}} \right| - \sqrt{nh} \max_{0 \le x \le 1} \left| \frac{S_n'(x) - ES_n'(x)}{\sigma_N(x) \sqrt{f(x)}} \right| \right| > \varepsilon \right)$$

$$\le P\left(l_n \sqrt{nh} \max_{0 \le x \le 1} \left| \frac{ES_n'(x) - f(x)}{\sigma_N(x) \sqrt{f(x)}} \right| > \varepsilon \right) \le P\left(l_n \sqrt{nh} C_{20} h > \varepsilon \right).$$

It remains to choose $\varepsilon = 2C_{20} l_n \sqrt{nh^3}$ and using Theorem 1 [1]. Theorem 3 is proved.

Relations $\lim_{t \uparrow x_i} E_{i,j}'(t) = \lim_{t \downarrow x_i} E_{i+1,j}'(t)$, $\lim_{t \uparrow x_i} E_{i,j}''(t) \ne \lim_{t \downarrow x_i} E_{i+1,j}''(t)$, $i = 1, \cdots, N-1$ imply

Theorem 4. First order mean square derivations of the Gauss process $\eta_n(x) = 0$ are continuous in [0, 1], and second order mean square derivations do not have discontinuity in the points of the spline interpolation.

Let now $x_k = kh, k = 0, 1, \cdots, N$ be points of the cubical spline interpolation, and $T_k = [x_{k-1}, x_k], k = 1, \cdots, N$ be a uniform partition of the interval [0, 1]. Is is valid the following

Theorem 5. 1) The variance of mean square derivations of the Gauss process

$$\eta_n(x) = \frac{1}{\sigma_N(x)\sqrt{h}} \int_0^1 W_N(x,y) d\omega_n(t) \text{ vanishes in the intervals } T_1 \text{ and } T_N \text{ at the points } t_1 = -\frac{2\sigma h}{(1-\sigma)^2} \text{ and}$$

$$t_N = x_{N-1} + \frac{(1+\sigma^2)h}{(1-\sigma)^2}, \text{ respectively;}$$

2) If the variance vanishes also in intervals $T_2, \cdots, T_{N-1}$, then there will be not more than two roots in each interval.

Proof. At the beginning of the proof of the theorem, we proceed as in [2]. Let $ht, hs \in T_1$. Then using the relation ([4], p. 28)

$$\left(\sum_j a_j b_j \right)^2 = \sum_j a_j^2 \sum_j b_j^2 - \frac{1}{2} \sum_{k,j} \left(a_k b_j - a_j b_k \right)^2,$$

we get for $(t,s) \in [0,1]^2$

$$\rho_n^2(th, sh) = \left(\text{cov}\left(\eta_n(th), \eta_n(sh) \right) \right)^2$$

$$= 1 - \frac{\sum_{k,j=0}^{N-1} \left[E_{1,k}(th) E_{1,j}(sh) - E_{1,j}(th) E_{1,k}(sh) \right]^2}{2\sigma_N^2(th)\sigma_N^2(sh)}.$$

(6)

Substituting into (6)

$$E_{1,0}(th) E_{1,1}(sh) - E_{1,1}(th) E_{1,0}(sh)$$
$$= D_{1,0}(th) D_{1,1}(sh) - D_{1,0}(sh) D_{1,1}(th) + D_{1,1}(sh) - D_{11}(th)$$

and taking into account that $D_{1,0}(th) D_{1,1}(sh) - D_{1,0}(sh) D_{1,1}(th) = 0$, we obtain

$$E_{1,0}(th) E_{1,1}(sh) - E_{1,0}(sh) E_{1,1}(th) = D_{1,1}(sh) - D_{1,1}(th)$$

or

$$\left[E_{1,0}(th) E_{1,1}(sh) - E_{1,0}(sh) E_{1,1}(th) \right]^2$$
$$= \frac{9}{4}(s-t)^2 \left[(2-t-s)\left(A_{0,1}^{-1} - A_{0,2}^{-1} \right) + (s+t)\left(A_{1,1}^{-1} - A_{1,2}^{-1} \right) \right]^2.$$

We find analogously

$$\left[E_{1,k}(th) E_{1,j}(sh) - E_{1,k}(sh) E_{1,j}(th) \right]^2$$
$$= \frac{9}{4}(s-t)^2 \left[(2-t-s)\left(A_{0,j}^{-1} - A_{0,j+1}^{-1} \right) + (t+s)\left(A_{1,j}^{-1} - A_{1,j+1}^{-1} \right) \right]^2,$$

and also

$$\left[E_{1,N-1}(th) E_{1,N-2}(sh) - E_{1,N-1}(sh) E_{1,N-2}(th) \right]^2$$
$$= \frac{9}{4}(s-t)^2 \left[(2-t-s) A_{0,N-1}^{-1} + (t+s) A_{1,N-1}^{-1} \right]^2.$$

Generalizing the obtained results, we have

$$1 - \rho_n^2(th, sh) = \frac{9}{2} \left\{ \sum_{j=1}^{N-2} \left[(2-s-t)\left(A_{0,j}^{-1} - A_{0,j+1}^{-1}\right) + (t+s)\left(A_{1,j}^{-1} - A_{1,j+1}^{-1}\right) \right]^2 \right.$$

$$\left. + \left[(2-t-s)A_{0,N-1}^{-1} + (t+s)A_{1,N-1}^{-1} \right]^2 \right\} \cdot \left(\sigma_N(th) \cdot \sigma_N(sh)\right)^{-2} \cdot (s-t)^2.$$

Denote $\tilde{K}_n(t,\tau) = W_N(t,\tau) \big/ \left(\sqrt{h}\sigma_N(t)\right)$. The equality

$$1 - \rho_n(t,s) = \frac{1}{2} \int_0^1 \left[\tilde{K}_n(t,\tau) - \tilde{K}_n(s,\tau) \right]^2$$

implies

$$\rho_n(t,t) = 1, \quad \left. \frac{\partial \rho_n(t,s)}{\partial t} \right|_{t=s} = \left. \frac{\partial \rho_n(t,s)}{\partial s} \right|_{t=s} = 0.$$

On the other hand,

$$\operatorname{Var} \eta_n'(th) = \left. \frac{\partial^2 \rho_n(th, sh)}{\partial t \partial s} \right|_{s=t} = \frac{B_{N,1}(t,t)}{2\sigma^4(th)}$$

where

$$B_{N,1}(t,t) = 18 \left\{ \sum_{j=1}^{N-2} \left[(1-t)\left(A_{0,j}^{-1} - A_{0,j+1}^{-1}\right) + t\left(A_{1,j}^{-1} - A_{1,j+1}^{-1}\right) \right]^2 + \left[(1-t)A_{0,N}^{-1} + tA_{0,N-1}^{-1} \right]^2 \right\}.$$

Obviously, $\sigma_N^2(th) = \sum_{j=0}^{N-1} E_{1,j}^2(th) > 0$. The point $t_1 = -h \cdot \dfrac{2\sigma}{(1-\sigma)^2} \in T_1$ will be a solution of the equation $\operatorname{Var} \eta_n'(th) = 0$. Recall that $\sigma = \sqrt{3} - 2$. Like the case of $i = 1$, we can act analogously in the case of $i = N$, i.e. at $ht + x_{N-1} \in T_N$, $\operatorname{Var} \eta_n'(th) = 0$ when $t_N = h\dfrac{1+\sigma^2}{(1-\sigma)^2} + x_{N-1} \in T_N$.

The first part of Theorem 5 is proved.

Let pass to the proof of the second part. Both in the case of $i = 1$, i.e. when ht, $hs \in T_1$, and in the case of $i = 2, \cdots, N-1$, the equality

$$\rho_n^2(th + x_{i-1}, sh + x_{i-1}) = 1 - (t-s)^2 \frac{B_{N,i}(t,s)}{4\sigma_N^4(th + x_{i-1})}$$

is valid for $ht + x_{i-1}$, $hs + x_{i-1} \in T_i$.

The explicit form of $B_{N,i}(t,s)$ is given in Muminov (1987), and it is very cumbersome.

Note, in this case $\operatorname{Var} \eta_n'(th) = \dfrac{B_{N,i}(t,t)}{2\sigma^4(th + x_{i-1})}$ also.

One can easily see that $B_{n,i}(t,t)$ is the sum of second powers of quadratic trinomials with respect to $t \in [0,1]$, and it has not more than two real roots if they exist in [0, 1].

The first part of Theorem 5 is proved.

At last, Theorems 2 and 3 imply that limit distributions of the random variables $l_n \xi_n - l_n^2$ and $l_n \max\limits_{0 \le x \le 1} |\eta_n(x)| - l_n^2$ coincide. However, the Gauss process $\eta_n(x)$ does not have second order mean square derivatives in the interpolation points for the spline, and $\inf\limits_{0 \le x \le 1} \operatorname{Var} \eta_n'(x) = 0$. Therefore one can not apply results of the works [5]-[7] to investigate the distribution of the maximum of $\eta_n(x)$. This deficiency has been removed in [8].

References

[1] Muminov, M.S. and Soatov, Kh. (2011) A Note on Spline Estimator of Unknown Probability Density Function. *Open*

Journal of Statistics, 157-160. http://dx.doi.org/10.4236/ojs.2011.13019

[2] Khashimov, Sh.A. and Muminov, M.S. (1987) The Limit Distribution of the Maximal Deviation of a Spline Estimate of a Probability Density. *Journal of Mathematical Sciences*, **38**, 2411-2421.

[3] Stechkin, S.B. and Subbotin, Yu.N. (1976) Splines in Computational Mathematics. Nauka, Moscow, 272p.

[4] Hardy G.G., Littlewood, J.E. and Polio, G. (1948) Inequalities. Moscow. IL, 456 p. http://dx.doi.org/10.1007/BF01095085

[5] Berman, S.M. (1974) Sojourns and Extremes of Gaussian Processes. *Annals of Probability*, **2**, 999-1026. http://dx.doi.org/10.1214/aop/1176996495

[6] Rudzkis, R.O. (1985) Probability of the Large Outlier of Nonstationary Gaussian Process. *Lit. Math. Sb.*, **XXV**, 143-154.

[7] Azais, J.-M. and Wschebor, M. (2009) Level Sets and Extrema of Random Processes and Fields, Wiley, Hoboken, 290p.

[8] Muminov, M.S. (2010) On Approximation of the Probability of the Large Outlier of Nonstationary Gauss Process. *Siberian Mathematical Journal*, **51**, 144-161. http://dx.doi.org/10.1007/s11202-010-0015-6