CONDITIONAL PROBABILITY DENSITY AND REGRESSION FUNCTION ESTIMATIONS WITH TRANSFORMATION OF DATA

Moriyama, Taku
Dept. of Management of Social Systems and Civil Eng., Tottori University

Maesono, Yoshihiko
Department of Mathematics, Chuo University

https://doi.org/10.5109/2558879

出版情報：Bulletin of informatics and cybernetics. 52 (1), pp.1-25, 2020. Research Association of Statistical Sciences
バージョン：

CONDITIONAL PROBABILITY DENSITY AND REGRESSION FUNCTION ESTIMATIONS WITH TRANSFORMATION OF DATA

by

Taku Moriyama and **Yoshihiko Maesono**

Reprinted from the Bulletin of Informatics and Cybernetics Research Association of Statistical Sciences, Vol.52

FUKUOKA, JAPAN
2020

Bulletin of Informatics and Cybernetics, Vol. 52, 2020

CONDITIONAL PROBABILITY DENSITY AND REGRESSION FUNCTION ESTIMATIONS WITH TRANSFORMATION OF DATA

By

Taku Moriyama* and **Yoshihiko Maesono†**

Abstract

This study extends 'empirical transformation' in kernel smoothing to conditional probability density and regression function estimations. We propose nonparametric conditional probability density and regression function estimators, which are based on an extension of Ćwik and Mielniczuk (1989)'s method. We derived asymptotic properties of the proposed estimators and conducted a simulation that showed mean squared errors of conditional probability density and regression function estimators.

Key Words and Phrases: conditional probability density, mean squared error, nonparametric estimation, regression function.

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be n pairs of independent random variables (r.v.s.) with the continuous joint density function $f(x, y)$ and $g(x)$ be the marginal density of X. Throughout the paper, we assume that the support of f covers $\mathbb{R}^2$ to avoid what we call the 'boundary bias problem'. The conditional probability density of Y given $X = x_0$ (fixed) is given by

$$f_{Y|x_0}(y) := f_Y(y|X = x_0) = \frac{f(x_0, y)}{g(x_0)}.$$

Conditional density estimation provides not only its mean but also its mode, multimodality, skewness, quantile points and so on. Rosenblatt (1969) first considered nonparametric kernel-type estimation. The estimator at $Y = y_0$ (fixed) is

$$\widehat{f}_{Y|x_0}(y_0) = \frac{\widehat{f}(x_0, y_0)}{\widehat{g}(x_0)},$$

where $\widehat{f}$ and $\widehat{g}$ are the 'naïve' kernel density estimators:

$$\widehat{f}(x_0, y_0) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} K\left(\frac{x_0 - X_i}{h_1}\right) K\left(\frac{y_0 - Y_i}{h_2}\right)$$

* Dept. of Management of Social Systems and Civil Eng., Tottori University, 4-101 Koyama-Minami Tottori 680-8550 Japan. tel +81–857–31–5335 moriyama@tottori-u.ac.jp

† Department of Mathematics, Chuo University, 1-13-27 Kasuga, Bunkyo-ku Tokyo 112-8551 Japan. tel +81–03–3817–1762 maesono@math.chuo-u.ac.jp

and

$$\widehat{g}(x_0) = \frac{1}{nh_3}\sum_{i=1}^{n} K\left(\frac{x_0 - X_i}{h_3}\right),$$

where K is a symmetric probability density function and h_j $(j = 1, 2, 3)$ are the bandwidths. We assume that all the bandwidths in the estimators are the same, i.e., $h_j = h$ $(j = 1, 2, 3)$, where $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Let us use the following notation,

$$A_{i,j} := \int K^i(u)u^j \mathrm{d}u.$$

We also assume that all $A_{1,j}$ for $j = 0, \cdots, 4$ and $A_{2,k}$ for $k = 0, 1$ are bounded.

Hyndman et al. (1996) investigated the mean squared error of the kernel-type conditional density estimator and proposed an improved estimator.
Bashtannyk and Hyndman (2001) discussed methods of choosing the bandwidth of the conditional density estimator from the view of asymptotic results. Hall et. al. (2004) proved that a cross-validation method is quite effective for automatically shrinking and removing irrelevant explanatory components in multivariate settings. Hall et. al. (1999) introduced and investigated local logistic and adjusted the Nadaraya-Watson method for nonparametric conditional density estimation. Otneim and Tjøstheim (2017) proposed a method for tackling the curse of dimensionality in nonparametric settings using locally Gaussian approximations.

As pointed out by Faugeras (2009), the ratio-type estimator $\widehat{f}_{Y|x_0}$ can be numerically unstable when the denominator $\widehat{g}(x_0)$ is close to zero. Faugeras (2009) proposed a 'product-shaped estimator':

$$\widehat{f}^{\dagger}_{Y|x_0}(y_0) = \widehat{c}(G_n(x_0), L_n(y_0))\widehat{\ell}(y_0),$$

where $\widehat{c}$ is a copula density estimator and G_n, L_n are the empirical distribution functions of the marginal distribution functions of X and Y, respectively. $\widehat{\ell}$ is the kernel density estimator of the r.v. Y. The idea is based on the property of copula C; for any bivariate cumulative distribution function (c.d.f.) F, with marginal c.d.f. G of X and L of Y, there exists some function $C : [0, 1]^2 \to [0, 1]$. Copula C is called the copula function, which satisfies

$$F(x_0, y_0) = C(G(x_0), L(y_0))$$

(known as Sklar's theorem). If G and L are continuous, this representation is unique. Copula C is itself a c.d.f. on $[0, 1]^2$ with uniform marginal distributions. By differentiating the formula, we have the following representation:

$$f(x_0, y_0) = c(G(x_0), L(y_0))g(x_0)\ell(y_0),$$

where c is the copula density function and ℓ is the marginal density of Y. The above estimator $\widehat{f}^{\dagger}_{Y|x_0}$ comes from this relation. Faugeras (2009) investigated some asymptotic properties of this estimator.

By extending the idea of Ćwik and Mielniczuk (1989), we propose the following conditional density estimator:

$$\widehat{f}^{\star}_{Y|x_0}(y_0) := \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{Y_i - y_0}{h}\right) K\left(\frac{G_n(X_i) - G_n(x_0)}{h}\right).$$

The proposed estimator $\widehat{f}^{\star}_{Y|x_0}$ can be seen as an estimator with 'empirical transformation' $(X \mapsto G_n(X))$ (see Ruppert and Cline (1994)). Since $\widehat{f}^{\star}_{Y|x_0}$ is not a ratio, we also expect it to have less variance, as Faugeras (2009) did.

Since both Faugeras (2009)'s estimator $\widehat{f}^{\dagger}_{Y|x}$ and our $\widehat{f}^{\star}_{Y|x}$ use the empirical distribution function G_n, they are discontinuous functions as x. Let us replace G_n with the kernel-type estimator $\widehat{G}$. Then, we propose the following conditional density estimator:

$$\widehat{f}^{\diamond}_{Y|x_0}(y_0) := \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{Y_i - y_0}{h}\right) K\left(\frac{\widehat{G}(X_i) - \widehat{G}(x_0)}{h}\right).$$

We assume that the bandwidth of $\widehat{G}$ coincides with h, which is used in this estimator $\widehat{f}^{\diamond}_{Y|x_0}$. Though the convergence rate of $\widehat{f}^{\diamond}_{Y|x_0}$ is not improved, as shown later, the kernel estimator $\widehat{f}^{\diamond}_{Y|x}(y_0)$ is a smooth function of x unlike $\widehat{f}^{\star}_{Y|x}(y_0)$. The asymptotic mean squared errors ($AMSEs$) of $\widehat{f}^{\star}_{Y|x_0}$ and $\widehat{f}^{\diamond}_{Y|x_0}$ are given in the following theorems.

Theorem 1 *Let us assume that* (i) *f and g are three-times differentiable at (x_0, y_0) and at x_0 respectively,* (ii) *$f^{(i,0)}(x_0, y_0)$, $g^{(i)}(x_0)$ and $f^{(0,j)}(x_0, y_0)$ are bounded for $i = 0, 1, 2, 3$ and $j = 1, 2, 3$,* (iii) *K is three-times continuously differentiable,* (iv) *is the support of K is given by $[-d, d]$ for some positive constant $d > 0$, and* (v) *$\sup_v |K^{(3)}(v)|$ is bounded. Then, the AMSE of $\widehat{f}^{\star}_{Y|x_0}(y_0)$ is given by*

$$\begin{aligned} & E\left[\widehat{f}^{\star}_{Y|x_0}(y_0) - f_{Y|x_0}(y_0)\right]^2 \\ =& \frac{h^4 A_{1,2}^2}{4}\Bigg[f^{(0,2)}(x_0, y_0) - f_{Y|x_0}(y_0)\frac{g(x_0)g''(x_0) - 3(g'(x_0))^2}{g^4(x_0)} \\ &\quad + \frac{1}{g^4(x_0)}\Big(3g'(x_0)f^{(1,0)}(x_0, y_0) - g(x_0)f^{(2,0)}(x_0, y_0)\Big)\Bigg]^2 \\ &+ \frac{A_{2,0}^2}{nh^2} f_{Y|x_0}(y_0) + O(h^6) + O\left(\frac{1}{nh}\right), \end{aligned}$$

where

$$f^{(i,j)}(x_0, y_0) = \left(\frac{\partial}{\partial w}\right)^i \left(\frac{\partial}{\partial z}\right)^j f(w, z)\Bigg|_{w=x_0, z=y_0}.$$

Proof: *See the appendix.*

Theorem 2 *Under the assumptions of Theorem 1, the AMSE of $\widehat{f}^{\diamond}_{Y|x_0}(y_0)$ is given by*

$$\begin{aligned}
&E\left[\widehat{f}^{\diamond}_{Y|x_0}(y_0) - f_{Y|x_0}(y_0)\right]^2 \\
=&\frac{h^4 A_{1,2}^2}{4}\left[\frac{2f(x_0,y_0)g''(x_0)}{g'(x_0)} + f^{(0,2)}(x_0,y_0) - f_{Y|x_0}(y_0)\frac{g(x_0)g''(x_0)-3(g'(x_0))^2}{g^4(x_0)}\right. \\
&\left. + \frac{1}{g^4(x_0)}\left(3g'(x_0)f^{(1,0)}(x_0,y_0) - g(x_0)f^{(2,0)}(x_0,y_0)\right)\right]^2 \\
&+ \frac{A_{2,0}^2}{nh^2} f_{Y|x_0}(y_0) + O(h^6) + O\left(\frac{1}{nh}\right).
\end{aligned}$$

Proof: *See the appendix.*

Remark 1 As seen above, the difference between the *AMSEs* of $\widehat{f}^{\star}_{Y|x_0}$ and $\widehat{f}^{\diamond}_{Y|x_0}$ is whether the first term in the bias terms exists. Under some regularity conditions, the *AMSE* of $\widehat{f}_{Y|x_0}(y_0)$ is given by

$$\begin{aligned}
&E\left[\widehat{f}_{Y|x_0}(y_0) - f_{Y|x_0}(y_0)\right]^2 \\
=&\frac{h^4 A_{1,2}^2}{4}\frac{1}{g^2(x_0)}\left[f^{(2,0)}(x_0,y_0) + f^{(0,2)}(x_0,y_0) - f_{Y|x_0}(y_0)g''(x_0)\right]^2 \\
&+ \frac{A_{2,0}^2}{nh^2}\frac{f_{Y|x_0}(y_0)}{g(x_0)} + O(h^6) + O\left(\frac{1}{nh}\right).
\end{aligned}$$

The difference between the asymptotic variances of the naïve $\widehat{f}_{Y|x_0}$ and $\widehat{f}^{\star}_{Y|x_0}$ is whether the first term is divided by $g(x_0)$.

Remark 2 Under some regularity conditions, the *AMSE* of $\widehat{f}^{\dagger}_{Y|x_0}(y_0)$ is given by

$$\begin{aligned}
&E\left[\widehat{f}^{\dagger}_{Y|x_0}(y_0) - f_{Y|x_0}(y_0)\right]^2 \\
=&\frac{h^4 A_{1,2}^2}{4}\ell^2(y_0)\left[c^{(0,2)}(G(x_0),L(y_0)) + c^{(2,0)}(G(x_0),L(y_0))\right]^2 \\
&+ \frac{A_{2,0}^2}{nh^2} f_{Y|x_0}(y_0)\ell(y_0) + O(h^6) + O\left(\frac{1}{nh}\right),
\end{aligned}$$

where

$$c^{(i,j)}(x_0,y_0) = \left(\frac{\partial}{\partial w}\right)^i \left(\frac{\partial}{\partial z}\right)^j c(w,z)\Big|_{w=x_0,z=y_0}.$$

As the theorems and remarks show, the $AMSEs$ of the conditional density function estimators are of the order $(h^4 + (nh^2)^{-1})$. Therefore, the theoretically optimal bandwidth is given by $n^{-1/6} \times C$ where $C > 0$ is constant. For the following $AMSE$:

$$\frac{B}{4}h^4 + \frac{V}{nh^2},$$

the constant C is given by $(2V/B)^{-1/6}$. Determining the bandwidth needs to estimate the optimal one; however, it is quite difficult for conditional density estimators, as explained by Faugeras (2009). We postpone this for future work.

We present some results of a numerical study on conditional density estimation. We first compared the accuracy of $\widehat{f}_{Y|x_0}(y)$, $\widehat{f}^{\dagger}_{Y|x_0}(y)$, $\widehat{f}^{\star}_{Y|x_0}(y)$, and $\widehat{f}^{\diamond}_{Y|x_0}(y)$ at some fixed points $y = y_0$. Let us suppose that (X, Y) is distributed normally, i.e., $f \sim N_2(\mathbf{0}, \Sigma)$. Set

$$\Sigma = \begin{pmatrix} 1 & \rho\sigma_Y \\ \rho\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

It follows that g is also the normal density function, where the mean is 0 and variance is 1. We simulated the mean squared error (MSE), i.e.,

$$\left(\widetilde{f}_{Y|x_0}(y_0) - f_{Y|x_0}(y_0)\right)^2$$

and its standard deviation (SD) for the conditional density estimator $\widetilde{f}_{Y|x_0}(y_0)$. Tables 1-3 show 100 times the obtained values. In the tables, 'Naïve', 'Product', 'Proposed', and 'Smoothed' denote $\widetilde{f}_{Y|x_0}(y_0) = \widehat{f}_{Y|x_0}(y_0)$, $\widehat{f}^{\dagger}_{Y|x_0}(y_0)$, $\widehat{f}^{\star}_{Y|x_0}(y_0)$, and $\widehat{f}^{\diamond}_{Y|x_0}(y_0)$ respectively. We simulated MSE values (and their SD) of the conditional density estimators 100,000 times. The term $G^{-1}(\epsilon)$ denotes $x_0 = G^{-1}(\epsilon)$, where $G^{-1}(\epsilon)$ is the ϵth quantile of the distribution G (of X, i.e., the standard normal distribution). The term $L^{-1}(\epsilon)$ denotes $y_0 = L^{-1}(\epsilon)$, where $L^{-1}(\epsilon)$ is that of the distribution L (of Y). In Tables 1-3, all the sample sizes are $(n =)$100 and the bandwidths of the conditional density estimators are $n^{-1/6}$. From the results, we can first see that the smoothed version usually outperforms the proposed estimator. Table 1 suggests that the smoothed version of the proposed estimator gives almost best estimates especially when $\rho = 0$ and $\sigma_Y \leq \sigma_X$. The naïve estimator is almost worst especially for $\sigma_Y \geq \sigma_X$. Table 2 also suggests that the naïve estimator is almost worst especially when $\rho = 1/3$ and $\sigma_Y \geq \sigma_X$. As σ_Y tends to be comparatively large, the product-shaped estimator becomes best. Table 3 suggests that the naïve one is most accurate especially when $\rho = 2/3$ and $\sigma_Y < \sigma_X$. For $\sigma_Y > \sigma_X$, the product-shaped estimator is almost best. We summarized the obtained simulation results in Table 4. The table shows a recommendation on nonparametric conditional density estimation for bivariate normal distributions. To sum up, the smoothed version of the proposed estimator seem to be superior to others especially when $\rho \approx 0$ and $\sigma_Y \leq \sigma_X$. When the correlation is high, the proposed estimators with $h = n^{-1/6}$ have large variance in some cases. Including the method of choosing bandwidth, we need to continue studying the proposed estimators.

Table 1: MSE and SD values (×100) with $(n = 100, \rho = 0, h = n^{-1/6})$ for the conditional density estimators

$G^{-1}(\cdot)$	$L^{-1}(\cdot)$	Naïve MSE	Naïve SD	Product MSE	Product SD	Proposed MSE	Proposed SD	Smoothed MSE	Smoothed SD
$\sigma_Y = 1/2$									
1/4	1/4	2.130	2.941	3.298	2.497	2.753	2.437	2.020	2.047
	1/2	2.249	3.128	3.150	2.363	3.208	2.658	2.356	2.232
	3/4	2.162	3.046	3.472	2.614	2.780	2.491	2.057	2.101
1/2	1/4	1.772	2.465	1.167	1.357	0.881	1.150	0.637	0.886
	1/2	1.823	2.522	0.919	1.125	0.976	1.263	0.646	0.909
	3/4	1.717	2.415	1.178	1.364	0.832	1.092	0.592	0.826
3/4	1/4	2.115	2.892	3.555	2.615	3.000	2.588	2.035	2.077
	1/2	2.226	3.066	3.468	2.592	3.536	2.891	2.393	2.311
	3/4	2.111	2.927	3.697	2.686	3.013	2.569	2.040	2.057
$\sigma_Y = 1$									
1/4	1/4	1.326	1.895	0.920	0.817	0.600	0.742	0.514	0.668
	1/2	1.504	2.094	0.710	0.786	0.838	1.000	0.684	0.871
	3/4	1.308	1.867	0.975	0.862	0.597	0.746	0.510	0.665
1/2	1/4	1.062	1.542	0.510	0.599	0.403	0.584	0.460	0.700
	1/2	1.229	1.754	0.380	0.529	0.471	0.656	0.530	0.780
	3/4	1.050	1.473	0.535	0.619	0.405	0.573	0.460	0.679
3/4	1/4	1.306	1.911	0.958	0.834	0.636	0.771	0.515	0.669
	1/2	1.519	2.217	0.743	0.801	0.858	1.002	0.654	0.831
	3/4	1.315	1.885	1.011	0.866	0.626	0.752	0.507	0.652
$\sigma_Y = 2$									
1/4	1/4	0.420	0.719	0.158	0.126	0.127	0.160	0.127	0.167
	1/2	0.901	1.337	0.253	0.312	0.337	0.423	0.308	0.401
	3/4	0.411	0.703	0.165	0.126	0.130	0.163	0.129	0.170
1/2	1/4	0.324	0.506	0.126	0.118	0.125	0.188	0.137	0.216
	1/2	0.742	1.115	0.223	0.330	0.280	0.405	0.315	0.484
	3/4	0.337	0.537	0.129	0.118	0.126	0.186	0.138	0.216
3/4	1/4	0.403	0.653	0.160	0.125	0.128	0.167	0.128	0.179
	1/2	0.935	1.368	0.260	0.311	0.350	0.435	0.314	0.409
	3/4	0.424	0.741	0.166	0.126	0.128	0.164	0.127	0.174

Table 2: MSE and SD values ($\times 100$) with ($n = 100, \rho = 1/3, h = n^{-1/6}$) for the conditional density estimators

$G^{-1}(\cdot)$	$L^{-1}(\cdot)$	Naïve		Product		Proposed		Smoothed	
		MSE	SD	MSE	SD	MSE	SD	MSE	SD
$\sigma_Y = 1/2$									
1/4	1/4	2.427	3.289	4.788	3.050	7.691	4.547	6.242	4.061
	1/2	2.329	3.226	3.902	2.736	3.855	3.050	2.866	2.580
	3/4	1.991	2.753	3.369	2.518	0.964	1.260	0.698	0.970
1/2	1/4	1.826	2.522	1.680	1.726	1.144	1.406	0.735	1.013
	1/2	2.070	2.797	1.463	1.520	1.428	1.604	0.821	1.110
	3/4	1.881	2.610	1.753	1.773	1.221	1.518	0.750	1.075
3/4	1/4	2.021	2.853	3.458	2.585	1.077	1.361	0.717	0.976
	1/2	2.291	3.205	4.076	2.791	4.156	3.123	2.825	2.508
	3/4	2.318	3.171	5.135	3.200	8.157	4.639	6.225	4.030
$\sigma_Y = 1$									
1/4	1/4	1.476	2.084	1.528	1.200	2.393	1.736	2.046	1.611
	1/2	1.546	2.181	0.831	0.878	0.922	1.072	0.741	0.924
	3/4	1.160	1.726	0.629	0.587	0.596	0.840	0.727	0.980
1/2	1/4	1.051	1.470	0.596	0.667	0.438	0.600	0.471	0.679
	1/2	1.251	1.702	0.447	0.615	0.520	0.713	0.517	0.748
	3/4	1.071	1.541	0.613	0.679	0.440	0.599	0.471	0.688
3/4	1/4	1.166	1.672	0.638	0.597	0.562	0.811	0.717	0.979
	1/2	1.548	2.152	0.883	0.914	1.006	1.134	0.752	0.932
	3/4	1.456	2.057	1.616	1.251	2.492	1.775	2.024	1.614
$\sigma_Y = 2$									
1/4	1/4	0.528	0.824	0.304	0.213	0.381	0.285	0.350	0.278
	1/2	0.940	1.386	0.291	0.349	0.374	0.459	0.339	0.432
	3/4	0.282	0.519	0.051	0.050	0.282	0.412	0.307	0.439
1/2	1/4	0.302	0.501	0.105	0.105	0.129	0.198	0.143	0.228
	1/2	0.773	1.100	0.239	0.325	0.295	0.402	0.315	0.453
	3/4	0.314	0.517	0.108	0.105	0.131	0.194	0.147	0.229
3/4	1/4	0.287	0.538	0.050	0.051	0.283	0.408	0.314	0.439
	1/2	0.959	1.467	0.295	0.347	0.374	0.453	0.326	0.416
	3/4	0.525	0.866	0.316	0.213	0.393	0.292	0.352	0.281

Table 3: MSE and SD values (×100) with ($n = 100, \rho = 2/3, h = n^{-1/6}$) for the conditional density estimators

$G^{-1}(\cdot)$	$L^{-1}(\cdot)$	Naïve *MSE*	Naïve *SD*	Product *MSE*	Product *SD*	Proposed *MSE*	Proposed *SD*	Smoothed *MSE*	Smoothed *SD*
$\sigma_Y = 1/2$									
1/4	1/4	3.763	4.633	13.48	5.180	27.35	10.20	23.96	9.560
	1/2	2.441	3.427	6.344	3.687	6.019	4.371	4.539	3.718
	3/4	1.784	2.578	1.850	1.732	1.756	1.849	2.498	2.294
1/2	1/4	2.788	3.499	6.409	3.615	4.142	3.443	2.826	2.844
	1/2	3.684	4.047	7.510	3.850	6.401	3.735	4.454	3.227
	3/4	2.686	3.418	6.419	3.707	4.343	3.669	2.818	2.884
3/4	1/4	1.798	2.517	1.864	1.736	1.576	1.759	2.495	2.309
	1/2	2.415	3.299	6.424	3.656	6.557	4.623	4.538	3.742
	3/4	3.756	4.530	13.92	5.287	28.61	10.46	23.96	9.582
$\sigma_Y = 1$									
1/4	1/4	1.752	2.373	4.611	2.444	10.70	3.929	9.750	3.825
	1/2	1.610	2.277	1.310	1.174	1.343	1.461	1.054	1.254
	3/4	0.877	1.417	0.146	0.216	4.357	2.770	4.849	2.978
1/2	1/4	1.097	1.529	0.872	0.855	0.614	0.828	0.600	0.842
	1/2	1.484	2.073	1.569	1.466	1.267	1.406	0.953	1.184
	3/4	1.113	1.590	0.905	0.888	0.670	0.901	0.626	0.885
3/4	1/4	0.909	1.432	0.149	0.221	4.160	2.726	4.768	2.976
	1/2	1.610	2.250	1.341	1.180	1.447	1.488	1.029	1.199
	3/4	1.782	2.457	4.782	2.499	11.13	3.961	9.810	3.845
$\sigma_Y = 2$									
1/4	1/4	0.641	1.004	0.501	0.330	1.129	0.397	1.070	0.404
	1/2	1.014	1.536	0.406	0.437	0.470	0.549	0.410	0.501
	3/4	0.088	0.240	0.025	0.038	1.050	0.914	1.091	0.943
1/2	1/4	0.230	0.445	0.045	0.068	0.172	0.308	0.193	0.336
	1/2	0.917	1.268	0.520	0.588	0.482	0.589	0.424	0.546
	3/4	0.233	0.429	0.043	0.064	0.163	0.296	0.195	0.340
3/4	1/4	0.092	0.271	0.024	0.037	1.030	0.899	1.065	0.923
	1/2	1.003	1.488	0.411	0.441	0.502	0.578	0.414	0.511
	3/4	0.638	0.994	0.510	0.326	1.145	0.390	1.063	0.403

Table 4: Recommendation on nonparametric conditional density estimation for $\sigma_X = 1$

$G^{-1}(\cdot)$	$L^{-1}(\cdot)$	$\sigma_Y = 1/2$	$\sigma_Y = 1$	$\sigma_Y = 2$
$\rho = 0$				
	1/4	Smoothed		Proposed
1/4	1/2	Naïve	Smoothed	Product
	3/4	Smoothed		Smoothed
	1/4		Proposed	Proposed
1/2	1/2	Smoothed	Product	Product
	3/4		Proposed	Proposed
	1/4	Smoothed		Proposed
3/4	1/2	Naïve	Smoothed	Product
	3/4	Smoothed		Smoothed
$\rho = 1/3$				
	1/4	Naïve	Naïve	
1/4	1/2	Naïve	Smoothed	Product
	3/4	Smoothed	Proposed	
	1/4		Proposed	
1/2	1/2	Smoothed	Product	Product
	3/4		Proposed	
	1/4	Smoothed	Proposed	
3/4	1/2	Naïve	Smoothed	Product
	3/4	Naïve	Naïve	
$\rho = 2/3$				
	1/4	Naïve	Naïve	
1/4	1/2	Naïve	Smoothed	Product
	3/4	Proposed	Product	
	1/4			Product
1/2	1/2	Naïve	Smoothed	Smoothed
	3/4			Product
	1/4	Proposed	Product	
3/4	1/2	Naïve	Smoothed	Product
	3/4	Naïve	Naïve	

Next, we consider to construct new regression function estimators. The regression function of Y on $X = x_0$ is defined as the following integral:

$$m(x_0) := E[Y|X = x_0] = \int y f_{Y|x_0}(y)\mathrm{d}y,$$

that is, the expectation of Y given $X = x_0$. It holds that

$$\begin{aligned}\int y\widehat{f}_{Y|x_0}(y)\mathrm{d}y =& \frac{1}{nh h\widehat{g}(x_0)} \int \sum_{i=1}^{n} yK\left(\frac{x_0 - X_i}{h}\right) K\left(\frac{y - Y_i}{h}\right) \mathrm{d}y \\ =& \frac{1}{nh\widehat{g}(x_0)} \sum_{i=1}^{n} K\left(\frac{x_0 - X_i}{h}\right) \int (Y_i + hu)K(u)\,\mathrm{d}u \\ =& \frac{1}{nh\widehat{g}(x_0)} \sum_{i=1}^{n} Y_i K\left(\frac{x_0 - X_i}{h}\right) \\ =&: \widehat{m}(x_0),\end{aligned}$$

where $\widehat{m}$ is the Nadaraya-Watson estimator. Asymptotic properties of the Nadaraya-Watson estimator have been studied and summarized, for example, by Wand and Jones (1995). Some asymptotic properties in multivariate settings were given by Georgiev (1988). Lei and Wasserman (2014) studied nonparametric prediction bands in a finite sample, and Burman and Chen (1989) introduced a nonparametric estimator for the heteroscedastic regression model. We fix the point $X = x_0$ throughout the paper; thus, we do not consider heteroscedasticity.

By the following integral of $\widehat{f}^{\star}_{Y|x_0}(y)$ and $\widehat{f}^{\diamond}_{Y|x_0}(y)$:

$$\begin{aligned}\int y\widehat{f}^{\star}_{Y|x_0}(y)\mathrm{d}y =& \frac{1}{nh^2} \int \sum_{i=1}^{n} yK\left(\frac{Y_i - y}{h}\right) K\left(\frac{G_n(X_i) - G_n(x_0)}{h}\right) \mathrm{d}y \\ =& \frac{1}{nh} \sum_{i=1}^{n} \int (Y_i - hu)K(u)\,K\left(\frac{G_n(X_i) - G_n(x_0)}{h}\right) \mathrm{d}u \\ =& \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{G_n(X_i) - G_n(x_0)}{h}\right),\end{aligned}$$

we also propose the following regression function estimators:

$$\widehat{m}^{\star}(x_0) := \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{G_n(X_i) - G_n(x_0)}{h}\right)$$

and

$$\widehat{m}^{\diamond}(x_0) := \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{\widehat{G}(X_i) - \widehat{G}(x_0)}{h}\right).$$

The next theorems show the *AMSEs* of these regression function estimators.

Theorem 3 *Let us assume the assumptions of Theorem 1 and that $V[Y|X = x_0]$ exists. Then, the AMSE of $\widehat{m}^{\star}(x_0)$ is given by*

$$\begin{aligned}
&E\left[\widehat{m}^{\star}(x_0) - m(x_0)\right]^2 \\
=&\frac{h^4 A_{1,2}^2}{4}\Bigg[-m(x_0)\frac{g(x_0)g''(x_0) - 3(g'(x_0))^2}{g^4(x_0)} \\
&\quad + \int z\left(\frac{3g'(x_0)}{g^4(x_0)}f^{(1,0)}(x_0,z) - \frac{1}{g^3(x_0)}f^{(2,0)}(x_0,z)\right)\mathrm{d}z\Bigg]^2 \\
&+ \frac{A_{2,0}}{nh}V[Y|X = x_0] + O(h^6) + O\left(\frac{1}{n}\right).
\end{aligned}$$

Proof: *The proof follows from Theorem 1.*

Theorem 4 *Under the assumptions of Theorem 3, the AMSE of $\widehat{m}^{\diamond}(x_0)$ is given by*

$$\begin{aligned}
&E\left[\widehat{m}^{\diamond}(x_0) - m(x_0)\right]^2 \\
=&\frac{h^4 A_{1,2}^2}{4}\Bigg[m(x_0)\left(\frac{2g(x_0)g''(x_0)}{g'(x_0)} - \frac{g(x_0)g''(x_0) - 3(g'(x_0))^2}{g^4(x_0)}\right) \\
&\quad + \int z\left(\frac{3g'(x_0)}{g^4(x_0)}f^{(1,0)}(x_0,z) - \frac{1}{g^3(x_0)}f^{(2,0)}(x_0,z)\right)\mathrm{d}z\Bigg]^2 \\
&+ \frac{A_{2,0}}{nh}V[Y|X = x_0] + O(h^6) + O\left(\frac{1}{n}\right).
\end{aligned}$$

Proof: *The proof follows from Theorem 2.*

Remark 3 As seen above, the difference between the $AMSEs$ of $\widehat{m}^{\star}(x_0)$ and $\widehat{m}^{\diamond}(x_0)$ is whether the first term in the bias terms exists. Under some regularity conditions, the $AMSE$ of the Nadaraya-Watson estimator $\widehat{m}(x_0)$ is given by

$$\begin{aligned}
&E\left[\widehat{m}(x_0) - m(x_0)\right]^2 \\
=&\frac{h^4 A_{1,2}^2}{4}\frac{1}{g^2(x_0)}\left[-m(x_0)g''(x_0) + \int z f^{(2,0)}(x_0,z)\mathrm{d}z\right]^2 \\
&+ \frac{A_{2,0}}{nh}\frac{1}{g(x_0)}V[Y|X = x_0] + O(h^6) + O\left(\frac{1}{n}\right).
\end{aligned}$$

The difference between the asymptotic variances of $\widehat{m}(x_0)$ and $\widehat{m}^{\star}(x_0)$ is whether the first term is divided by $g(x_0)$.

The $AMSEs$ of the regression function estimators are of the order $(h^4 + (nh)^{-1})$, so the optimal bandwidth is of the order $n^{-1/5}$. For the following $AMSE$

$$\frac{B}{4}h^4 + \frac{V}{nh},$$

the optimal bandwidth is given by $h = n^{-1/5} \times (V/B)^{-1/5}$.

Suppose that F is the bivariate normal distribution $N_2(\mathbf{0}, \Sigma_0)$, where

$$\Sigma_0 = \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}.$$

From the results, we can see that

$$E\left[\widehat{m}^{\star}(x_0) - m(x_0)\right]^2 \approx \frac{h^4 A_{2,1}^2}{4}\frac{1}{\phi^4(x_0)}\left\{\frac{\rho x_0(2\rho^2+1)}{1-\rho^2}\right\}^2 + \frac{A_{0,2}}{nh}(1-\rho^2)$$

and

$$E\left[\widehat{m}(x_0) - m(x_0)\right]^2 \approx \frac{h^4 A_{2,1}^2}{4}(2\rho x_0)^2 + \frac{A_{0,2}}{nh}\frac{(1-\rho^2)}{\phi(x_0)},$$

where $\phi(x_0)$ is the probability density of the standard normal distribution. Theoretically optimal bandwidths minimize $AMSE$ values, which equal

$$n^{-4/5}\frac{5}{4}\left[A_{2,1}\left(\frac{A_{0,2}}{4\phi(x_0)}\right)^2(1-\rho^2)(\rho x_0(2\rho^2+1))\right]^{2/5}$$

$$n^{-4/5}\frac{5}{4}\left[A_{2,1}\left(\frac{A_{0,2}}{4\phi(x_0)}\right)^2(1-\rho^2)^2(2\rho x_0)\right]^{2/5}.$$

It follows that the proposed estimator is theoretically better for $-\frac{1}{2} < \rho < \frac{1}{2}$ regardless of x_0 when $F \sim N_2(\mathbf{0}, \Sigma)$. The optimal bandwidths are

$$n^{-1/5}\left[\frac{A_{0,2}}{A_{2,1}^2}\phi^4(x_0)(1-\rho^2)\left\{\frac{1-\rho^2}{\rho x_0(2\rho^2+1)}\right\}^2\right]^{1/5}$$

$$n^{-1/5}\left[\frac{A_{0,2}}{A_{2,1}^2}\frac{(1-\rho^2)(2\rho x_0)^2}{\phi(x_0)}\right]^{1/5}$$

respectively. We recommend choosing each bandwidth h of the proposed estimators by a usual cross-validation procedure like the Nadaraya-Watson estimator.

We next compared the accuracy of $\widehat{m}(x_0)$, $\widehat{m}^{\star}(x_0)$, and $\widehat{m}^{\diamond}(x_0)$. Suppose $f \sim N_2(\mathbf{0}, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 1 & \rho\sigma_Y \\ \rho\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

We simulated the following mean squared error (MSE):

$$(\widetilde{m}(x_0) - m(x_0))^2$$

and its standard deviation (SD) for the regression estimator $\widetilde{m}(x_0)$. Tables 5 and 6 show 100 times the obtained values. The difference in the results between Tables 5 and 6 comes from the way to choose each bandwidth. In Table 5, the bandwidths of all regression estimators were fixed and $n^{-1/5}$. In Table 6, they were data-driven and calculated by the cross-validation. $\widehat{h}$ denotes the cross-validation estimator. In the tables, 'Naïve', 'Proposed', and 'Smoothed' denote $\widetilde{m}(x_0) = \widehat{m}(x_0)$, $\widehat{m}^{\star}(x_0)$, and $\widehat{m}^{\diamond}(x_0)$ respectively. All the sample sizes were ($n =$)100, and we simulated MSE values (and their SD) of the regression estimators 100,000 times. The cross-validated bandwidth in Table 6 was calculated at every time (100,000) step. The term $G^{-1}(\epsilon)$ denotes $x_0 = G^{-1}(\epsilon)$ defined before. We first see that the cross-validation is effective for estimating each optimal bandwidth by comparing Table 6 with Table 5. Compared with 'Smoothed', 'Proposed' is usually better. As seen from Table 6, 'Proposed' outperforms the others for $\rho = 0$ regardless of both x_0 and σ_Y. Thus, the tables supports the above theoretical result that the proposed estimator theoretically outperforms the Nadaraya-Watson estimator for $\rho \approx 0$ regardless of x_0. Although the naïve Nadaraya-Watson estimator is better for $\rho = 1/3$ and $2/3$, the difference is not so much. Therefore, we recommend 'Proposed' when the underlying probability distribution is bivariate normal and its correlation seems negligible.

Table 5: MSE and SD values with $(n = 100, h = n^{-1/5})$ for the regression estimators

$G^{-1}(\cdot)$	Naïve MSE	Naïve SD	Proposed MSE	Proposed SD	Smoothed MSE	Smoothed SD
$\rho = 0, \quad \sigma_Y = 1/2$						
1/4	0.178	0.196	0.137	0.109	0.138	0.112
1/2	0.062	0.091	0.024	0.033	0.025	0.036
3/4	0.180	0.197	0.138	0.109	0.139	0.112
$\rho = 0, \quad \sigma_Y = 1$						
1/4	0.767	0.880	0.545	0.426	0.548	0.433
1/2	0.248	0.357	0.095	0.134	0.101	0.143
3/4	0.765	0.878	0.546	0.424	0.550	0.434
$\rho = 0, \quad \sigma_Y = 2$						
1/4	4.316	5.701	2.088	1.447	2.088	1.446
1/2	0.990	1.433	0.378	0.537	0.403	0.572
3/4	4.363	5.787	2.081	1.434	2.089	1.456
$\rho = 1/3, \quad \sigma_Y = 1/2$						
1/4	0.056	0.080	0.031	0.044	0.032	0.046
1/2	0.097	0.128	0.271	0.171	0.266	0.172
3/4	0.228	0.233	0.879	0.282	0.872	0.287
$\rho = 1/3, \quad \sigma_Y = 1$						
1/4	0.241	0.350	0.531	0.464	0.545	0.475
1/2	0.239	0.347	0.540	0.466	0.527	0.467
3/4	0.407	0.593	2.540	0.868	2.544	0.885
$\rho = 1/3, \quad \sigma_Y = 2$						
1/4	1.893	2.698	4.874	2.409	4.890	2.433
1/2	0.987	1.423	1.403	1.447	1.368	1.449
3/4	4.966	7.697	7.691	2.566	7.735	2.592
$\rho = 2/3, \quad \sigma_Y = 1/2$						
1/4	0.041	0.059	0.462	0.272	0.446	0.269
1/2	0.060	0.083	1.404	0.373	1.386	0.380
3/4	0.095	0.126	2.470	0.411	2.470	0.423
$\rho = 2/3, \quad \sigma_Y = 1$						
1/4	0.140	0.202	0.216	0.301	0.215	0.299
1/2	0.349	0.434	2.929	1.066	2.877	1.086
3/4	1.046	1.127	6.030	1.153	6.077	1.178
$\rho = 2/3, \quad \sigma_Y = 2$						
1/4	3.044	2.775	7.670	3.451	7.711	3.497
1/2	2.939	2.721	7.672	3.422	7.511	3.481
3/4	24.99	21.55	16.47	3.354	16.47	3.369

Table 6: MSE and SD values with $(n = 100, h = \widehat{h})$ for the regression estimators

$G^{-1}(\cdot)$	Naïve		Proposed		Smoothed	
	MSE	SD	MSE	SD	MSE	SD
$\rho = 0, \quad \sigma_Y = 1/2$						
1/4	0.115	0.022	0.114	0.020	0.117	0.040
1/2	0.001	0.003	0.001	0.002	0.004	0.007
3/4	0.115	0.022	0.115	0.020	0.117	0.040
$\rho = 0, \quad \sigma_Y = 1$						
1/4	0.459	0.088	0.458	0.078	0.468	0.160
1/2	0.004	0.014	0.003	0.010	0.014	0.029
3/4	0.459	0.090	0.458	0.078	0.469	0.160
$\rho = 0, \quad \sigma_Y = 2$						
1/4	1.838	0.380	1.832	0.319	1.864	0.598
1/2	0.017	0.058	0.013	0.039	0.056	0.117
3/4	1.839	0.377	1.833	0.314	1.868	0.599
$\rho = 1/3, \quad \sigma_Y = 1/2$						
1/4	0.001	0.004	0.001	0.002	0.004	0.007
1/2	0.108	0.020	0.114	0.020	0.125	0.041
3/4	0.433	0.044	0.457	0.039	0.486	0.076
$\rho = 1/3, \quad \sigma_Y = 1$						
1/4	0.111	0.038	0.123	0.043	0.152	0.090
1/2	0.106	0.037	0.118	0.042	0.146	0.090
3/4	0.952	0.136	1.044	0.117	1.146	0.232
$\rho = 1/3, \quad \sigma_Y = 2$						
1/4	0.916	0.244	1.097	0.253	1.322	0.518
1/2	0.110	0.072	0.132	0.096	0.209	0.223
3/4	2.552	0.469	2.954	0.402	3.319	0.800
$\rho = 2/3, \quad \sigma_Y = 1/2$						
1/4	0.101	0.019	0.112	0.021	0.126	0.037
1/2	0.419	0.048	0.455	0.039	0.486	0.067
3/4	0.962	0.077	1.027	0.057	1.076	0.098
$\rho = 2/3, \quad \sigma_Y = 1$						
1/4	0.003	0.010	0.004	0.011	0.014	0.031
1/2	0.397	0.087	0.468	0.082	0.534	0.143
3/4	1.664	0.200	1.856	0.156	1.993	0.275
$\rho = 2/3, \quad \sigma_Y = 2$						
1/4	0.382	0.145	0.519	0.180	0.675	0.329
1/2	0.365	0.139	0.497	0.179	0.648	0.321
3/4	3.588	0.576	4.254	0.484	4.691	0.876

Acknowledgement

The authors thank the editor's and referees' valuable comments, which helped us improve this manuscript significantly. The authors gratefully acknowledge the Japan Society for the Promotion of Science, JSPS KAKENHI Grant Numbers JP16H02790, JP19K20223.

References

Bashtannyk, D. M. and Hyndman, R. J. (2001). Estimating and visualizing conditional densities, *Computational Statistics & Data Analysis*, 36(3), 279–298.

Burman, P. and Chen, K. W. (1989). Nonparametric estimation of a regression function, *The Annals of Statistics*, 1567–1596.

Ćwik, J. and Mielniczuk, J. (1989). Estimating density ratio with application to discriminant analysis, *Communications in Statistics-Theory and Methods*, 18(8), 3057–3069.

Faugeras, O. P. (2009). A quantile-copula approach to conditional density estimation, *Journal of Multivariate Analysis*, 100(9), 2083–2099.

Georgiev, A. A. (1988). Asymptotic properties of the multivariate nadaraya-watson regression function estimate: the fixed design case, *Statistics & probability letters*, 7(1), 35–40.

Hall, P., Wolff, R. C., and Yao, Q. (1999). Methods for estimating a conditional distribution function, *Journal of the American Statistical Association*, 94(445), 154–163.

Hall, P., Racine, J., and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities, *Journal of the American Statistical Association*, 99(468), 1015–1026.

Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. (1996). Reaction and Control I. Mixing Additive and Multiplicative Network Algebras, *Journal of Computational and Graphical Statistics*, 5(4), 315–336.

Lei, J. and Wasserman, L. (2014). Distribution-free prediction bands for nonparametric regression, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 71–96.

Otneim, H. and Tjøstheim, D. (2017). Conditional density estimation using the local gaussian correlation, *Statistics and Computing*, 28(2), 301–321.

Rosenblatt, M. (1969). Conditional probability density and regression estimators, in *Multivariate analysis II* (P. R. Krishnaiah, ed.), 25–31. Academic Press, New York.

Ruppert, D. and Cline, D. B. H. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations, *The Annals of Statistics*, 22(1), 185–210.

Vieu, P. (1991). Nonparametric regression: Optimal local bandwidth choice. *Journal of the Royal Statistical Society Series B (Methodological)*, 53(2), 453–464.

Wand, M. P. and Jones M. C. (1995). *Kernel Smoothing*, Chapman and Hall, London.

Appendix: Proofs of Theorems

Proof of Theorem 1

Suppose that the sample size n is large enough, which ensures

$$\min\left\{\frac{G(x_0)}{h}, \frac{1-G(x_0)}{h}\right\} > d.$$

The following expansion holds:

$$\begin{aligned}
\widehat{f}^{\star}_{Y|x_0}(y_0) =& \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K\left(\frac{G(X_i)-G(x_0)}{h}\right) \\
&+ \frac{1}{nh^3}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K'\left(\frac{G(X_i)-G(x_0)}{h}\right) \\
&\qquad \{[G_n(X_i)-G(X_i)]-[G_n(x_0)-G(x_0)]\} \\
&+ \frac{1}{nh^4}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K''\left(\frac{G(X_i)-G(x_0)}{h}\right) \\
&\qquad \{[G_n(X_i)-G(X_i)]-[G_n(x_0)-G(x_0)]\}^2 \\
&+ \frac{1}{nh^5}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K^{(3)}\left(\frac{{G_n}^*(X_i)-{G_n}^*(x_0)}{h}\right) \\
&\qquad \{[G_n(X_i)-G(X_i)]-[G_n(x_0)-G(x_0)]\}^3 \\
=:& J_1+J_2+J_3+{J_4}^*
\end{aligned}$$

where ${G_n}^*(X_i)$ is an r.v. between $G_n(X_i)$ and $G(X_i)$ with probability 1 and ${G_n}^*(x_0)$ is between $G_n(x_0)$ and $G(x_0)$ with probability 1.

First, we evaluate J_1, which is the sum of *i.i.d.* r.v.s. Let us use $\psi(u) := G^{-1}(G(x_0)+hu)$. The expectation of J_1 is obtained as follows:

$$\begin{aligned}
&E[J_1] = \frac{1}{h^2}\iint K\left(\frac{z-y_0}{h}\right) K\left(\frac{G(w)-G(x_0)}{h}\right) f(w,z)\mathrm{d}z\mathrm{d}w \\
=&\frac{1}{h}\int K\left(\frac{z-y_0}{h}\right)\mathrm{d}z \int K(u)\frac{f(\psi(u),z)}{g(\psi(u))}\mathrm{d}u \\
=&\frac{1}{h}\int K\left(\frac{z-y_0}{h}\right)\left[\frac{f(x_0,z)}{g(x_0)} - \left\{f(x_0,z)\frac{g(x_0)g''(x_0)-3(g'(x_0))^2}{2g^5(x_0)} - f^{(1,0)}(x_0,z)\frac{g'(x_0)}{g^4(x_0)}\right.\right. \\
&\qquad \left.\left. + \frac{1}{2g^4(x_0)}\Big(g(x_0)f^{(2,0)}(x_0,z) - g'(x_0)f^{(1,0)}(x_0,z)\Big)\right\}h^2A_{1,2}\right]\mathrm{d}z + O(h^4) \\
=&\int K(v)\left[\left\{\frac{1}{g(x_0)}f(x_0,y_0) + f^{(0,2)}(x_0,y_0)\frac{(hv)^2}{2}\right\} - \left\{f(x_0,y_0)\frac{g(x_0)g''(x_0)-3(g'(x_0))^2}{2g^5(x_0)}\right.\right. \\
&\quad \left.\left. - f^{(1,0)}(x_0,y_0)\frac{g'(x_0)}{g^4(x_0)} + \frac{1}{2g^4(x_0)}\Big(g(x_0)f^{(2,0)}(x_0,y_0) - g'(x_0)f^{(1,0)}(x_0,y_0)\Big)\right\}h^2A_{1,2}\right]\mathrm{d}v \\
&\quad + O(h^4).
\end{aligned}$$

From the calculation of the following second moment:

$$\begin{aligned}
&\frac{1}{nh^4}\iint K^2\left(\frac{z-y_0}{h}\right)K^2\left(\frac{G(w)-G(x_0)}{h}\right)f(w,z)\mathrm{d}z\mathrm{d}w\\
=&\frac{1}{nh^3}\int K^2\left(\frac{z-y_0}{h}\right)K^2(u)\frac{f(\psi(u),z)}{g(\psi(u))}\mathrm{d}z\mathrm{d}u\\
=&\frac{A_{2,0}^2}{nh^2}f_{Y|x_0}(y_0)+O\left(\frac{1}{nh}\right)
\end{aligned}$$

we can see that

$$V[J_1]=E[(J_1)^2]-E[J_1]^2=\frac{A_{2,0}^2}{nh^2}f_{Y|x_0}(y_0)+O\left(\frac{1}{nh}\right).$$

By applying a conditional expectation to J_2, we can obtain

$$\begin{aligned}
E[J_2]=&\frac{1}{n^2h^3}\sum_{i=1}^n\sum_{j=1}^n E\Bigg[K\left(\frac{Y_i-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&\qquad\{[I(X_j\le X_i)-G(X_i)]-[I(X_j\le x_0)-G(x_0)]\}\Bigg]\\
=&\frac{1}{nh^3}E\Bigg[K\left(\frac{Y_i-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&\qquad\sum_{j=1}^n E\Big[[I(X_j\le X_i)-G(X_i)]-[I(X_j\le x_0)-G(x_0)]\,\Big|\,Y_i,X_i\Big]\Bigg]\\
\approx&\frac{1}{nh^3}E\left[K\left(\frac{Y_i-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\{[1-G(X_i)]-[I(X_i\le x_0)-G(x_0)]\}\right]\\
=&\frac{1}{nh^2}\iint K\left(\frac{z-y_0}{h}\right)K'(u)\{[1-G(\psi(u))]-[I(\psi(u)\le x_0)-G(x_0)]\}\frac{f(\psi(u),z)}{g(\psi(u))}\mathrm{d}z\mathrm{d}u\\
=&O\left(\frac{1}{n}\right).
\end{aligned}$$

The squared value of J_2 is given by

$$\begin{aligned}
(J_2)^2=&\frac{1}{n^4h^6}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{m=1}^n K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\{[I(X_j\le X_i)-G(X_i)]-[I(X_j\le x_0)-G(x_0)]\}\\
&\{[I(X_m\le X_k)-G(X_k)]-[I(X_m\le x_0)-G(x_0)]\}\\
=:&\frac{1}{n^4h^6}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{m=1}^n\Xi(i,j,k,m).
\end{aligned}$$

We can find the following:

if all of (i,j,k,m) are different, $E[\Xi(i,j,k,m)] = E[E[\Xi(i,j,k,m)|X_i, X_k]] = 0$
if $i = j$ and all of (i,k,m) are different, $E[\Xi(i,j,k,m)] = E[E[\Xi(i,j,k,m)|X_i, X_k]] = 0$
if $i = k$ and all of (i,j,m) are different, $E[\Xi(i,j,k,m)] = E[E[\Xi(i,j,k,m)|X_i]] = 0$
if $i = m$ and all of (i,j,k) are different, $E[\Xi(i,j,k,m)] = E[E[\Xi(i,j,k,m)|X_i]] = 0.$

From the above results, we can see that the terms in which $j = m$ and all (i,j,k) are different is the main of $E[(J_2)^2]$. For the term in which $j = m$ and all (i,j,k) are different, the expectation is given by

$$\begin{aligned}
&E[\Xi(i,j,k,m)]\\
=&\frac{n(n-1)(n-2)}{n^4h^4}E\Bigg[K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&\quad K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\{[I(X_j\le X_i)-G(X_i)]-[I(X_j\le x_0)-G(x_0)]\}\\
&\quad \{[I(X_j\le X_k)-G(X_k)]-[I(X_j\le x_0)-G(x_0)]\}\Bigg].
\end{aligned}$$

It holds that

$$\begin{aligned}
&E\Bigg[E\Bigg[\{[I(X_j\le X_i)-G(X_i)]-[I(X_j\le x_0)-G(x_0)]\}\\
&\qquad\times\{[I(X_j\le X_k)-G(X_k)]-[I(X_j\le x_0)-G(x_0)]\}\;\Bigg|\;X_i,X_k\Bigg]\Bigg]\\
=&E\Big[G(\min(X_i,X_k))-G(\min(X_i,x_0))-G(\min(X_k,x_0))-G(X_i)G(X_k)+G(X_i)G(x_0)\\
&+G(X_k)G(x_0)-G^2(x_0)+G(x_0)\Big].
\end{aligned}$$

Let us evaluate the following first term (with respect to $G(\min(X_i, X_k))$)

$$\begin{aligned}
&E\Bigg[K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\\
&\quad\times G(\min(X_i,X_k))\Bigg]\\
=&\int\cdots\int_{-\infty}^{v}K\left(\frac{z_i-y_0}{h}\right)K\left(\frac{z_k-y_0}{h}\right)K'\left(\frac{G(w)-G(x_0)}{h}\right)K'\left(\frac{G(v)-G(x_0)}{h}\right)\\
&\quad\times G(w)f(w,z_i)f(v,z_k)\mathrm{d}w\mathrm{d}v\mathrm{d}z_i\mathrm{d}z_k\\
&+\int\cdots\int_{v}^{\infty}K\left(\frac{z_i-y_0}{h}\right)K\left(\frac{z_k-y_0}{h}\right)K'\left(\frac{G(w)-G(x_0)}{h}\right)K'\left(\frac{G(v)-G(x_0)}{h}\right)\\
&\quad\times G(v)f(w,z_i)f(v,z_k)\mathrm{d}w\mathrm{d}v\mathrm{d}z_i\mathrm{d}z_k
\end{aligned}$$

$$
\begin{aligned}
=&-h^2\int\cdots\int K\left(\frac{z_i-y_0}{h}\right)K\left(\frac{z_k-y_0}{h}\right)W\left(\frac{G(v)-G(x_0)}{h}\right)K'\left(\frac{G(v)-G(x_0)}{h}\right)\\
&\times\left(\frac{f(x_0,z_i)}{g(x_0)}+\frac{G(x_0)}{g^3(x_0)}\left(g(x_0)f^{(1,0)}(x_0,z_i)-f(x_0,z_i)g'(x_0)\right)\right)f(v,z_k)\mathrm{d}v\mathrm{d}z_i\mathrm{d}z_k\\
&-h^2\int\cdots\int K\left(\frac{z_i-y_0}{h}\right)z_k\left(1-W\left(\frac{G(v)-G(x_0)}{h}\right)\right)K'\left(\frac{G(v)-G(x_0)}{h}\right)\\
&\times\frac{G(v)}{g^3(x_0)}\left(g(x_0)f^{(1,0)}(x_0,z_i)-f(x_0,z_i)g'(x_0)\right)f(v,z_k)\mathrm{d}v\mathrm{d}z_i\mathrm{d}z_k+O(h^3),
\end{aligned}
$$

where W is the c.d.f. of K. It follows that

$$
\begin{aligned}
&E\Bigg[K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\\
&\qquad\times G(\min(X_i,X_k))\Bigg]\\
=&h^3\iint K\left(\frac{z_i-y_0}{h}\right)K\left(\frac{z_k-y_0}{h}\right)\Bigg(A_{2,0}\frac{f(x_0,z_k)}{g(x_0)}\\
&\qquad+\frac{h}{2g^3(x_0)}\left(g(x_0)f^{(1,0)}(x_0,z_k)-f(x_0,z_k)g'(x_0)\right)\Bigg)\\
&\quad\times\left(\frac{f(x_0,z_i)}{g(x_0)}+\frac{G(x_0)}{g^3(x_0)}\left(g(x_0)f^{(1,0)}(x_0,z_i)-f(x_0,z_i)g'(x_0)\right)\right)\mathrm{d}z_i\mathrm{d}z_k\\
&+h^3\iint z_iz_k\Bigg(-A_{2,0}G(x_0)\frac{f(x_0,z_k)}{g(x_0)}\\
&\qquad+\frac{h}{2g^3(x_0)}G(x_0)\left(g(x_0)f^{(1,0)}(x_0,z_k)-f(x_0,z_k)g'(x_0)\right)+h\frac{f(x_0,z_k)}{2g(x_0)}\Bigg)\\
&\quad\times\frac{1}{g^3(x_0)}\left(g(x_0)f^{(1,0)}(x_0,z_i)-f(x_0,z_i)g'(x_0)\right)\mathrm{d}z_i\mathrm{d}z_k+O(h^4)\\
=&h^3A_{2,0}\iint K\left(\frac{z_i-y_0}{h}\right)K\left(\frac{z_k-y_0}{h}\right)\frac{f(x_0,z_k)}{g(x_0)}\frac{f(x_0,z_i)}{g(x_0)}\mathrm{d}z_i\mathrm{d}z_k+O(h^4).
\end{aligned}
$$

By a similar calculation, we can see

$$
\begin{aligned}
&E\Bigg[K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\\
&\qquad G(\min(X_i,x_0))\Bigg]=O(h^4)
\end{aligned}
$$

and

$$
\begin{aligned}
&E\Bigg[K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\\
&\qquad G(X_i)G(X_k)\Bigg]=O(h^4).
\end{aligned}
$$

Therefore, it holds that for the term in which $j=m$ and all (i,j,k) are different,

$$
E[\Xi(i,j,k,m)]=\frac{A_{2,0}}{nh}E[Y|x]^2
$$

and that the other combinations of (i,j,k,m) are of the order n^{-1}. As a result, we obtain $E[(J_2)^2] = O((nh)^{-1})$ and

$$V[J_2] = \frac{A_{2,0}}{nh} E[Y|x]^2.$$

The moments of J_3 are obtained in a similar manner, which are given by

$$E[J_3] = O(n^{-1}) \quad \text{and} \quad V[J_3] = O(n^{-2}h^{-1}).$$

Finally, we obtain upper bounds of $|E[J_4^*]|$ and $E[(J_4^*)^2]$. From the assumption of Theorem 1, we can see that

$$\begin{aligned} E[J_4^*] = & \Bigg| E\Bigg[\frac{1}{h^5} K\left(\frac{Y_i - y_0}{h}\right) K^{(3)}\left(\frac{G_n{}^*(X_i) - G_n{}^*(x_0)}{h}\right) \\ & \qquad \{[G_n(X_i) - G(X_i)] - [G_n(x_0) - G(x_0)]\}^3 \Bigg]\Bigg| \\ \leq & \frac{\max_u K(u) \times \max_v |K^{(3)}(v)|}{h^5} E\left[\{[G_n(X_i) - G(X_i)] - [G_n(x_0) - G(x_0)]\}^3\right] \\ = & O\left(\frac{1}{n^2h^5}\right). \end{aligned}$$

Similarly, it follows that

$$\begin{aligned} & E[(J_4^*)^2] \\ \leq & \left(\frac{\max_u K(u) \times \max_v |K^{(3)}(v)|}{h^5}\right)^2 E\Bigg[\{[G_n(X_i) - G(X_i)] - [G_n(x_0) - G(x_0)]\}^3 \\ & \quad \times \{[G_n(X_j) - G(X_j)] - [G_n(x_0) - G(x_0)]\}^3 \Bigg] \\ = & O\left(\frac{1}{n^4h^{10}}\right). \end{aligned}$$

To sum up, we conclude that $J_2 + J_3 + J_4^*$ is asymptotically negligible for fixed x_0. The main bias of $\widehat{f}^{\star}_{Y|x_0}(y_0)$ comes from J_1. Since

$$\begin{aligned} E[J_1J_2] = & \frac{1}{n^3h^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n K\left(\frac{Y_i - y}{h}\right) K\left(\frac{Y_j - y_0}{h}\right) K'\left(\frac{G(X_i) - G(x_0)}{h}\right) \\ & \quad K'\left(\frac{G(X_j) - G(x_0)}{h}\right) \times \{[I(X_k \leq X_j) - G(X_j)] - [I(X_k \leq x_0) - G(x_0)]\} \\ = & E[J_1]E[J_2] + O\left(\frac{1}{n}\right), \end{aligned}$$

using the Cauchy-Schwarz inequality, the asymptotic variance is given by $V[J_1]$. Thus, we have

$$V[\widehat{f}^{\star}_{Y|x_0}(y_0)] = \frac{A_{2,0}^2}{nh^2} f_{Y|x_0}(y_0) + O((nh)^{-1}),$$

and Theorem 1 has been proved. □

Proof of Theorem 2

The same as the proof of Theorem 1, the $AMSE$ of $\widehat{f}^{\diamond}_{Y|x_0}(y_0)$ is obtained. In a similar manner, we decompose $\widehat{f}^{\diamond}_{Y|x_0}(y_0)$ as follows:

$$\begin{aligned}
\widehat{f}^{\diamond}_{Y|x_0}(y_0) =& \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K\left(\frac{\widehat{G}(X_i)-\widehat{G}(x_0)}{h}\right)\\
=& \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&+\frac{1}{nh^3}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&\quad\{[\widehat{G}(X_i)-G(X_i)]-[\widehat{G}(x_0)-G(x_0)]\}\\
&+\frac{1}{nh^4}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K''\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&\quad\{[\widehat{G}(X_i)-G(X_i)]-[\widehat{G}(x_0)-G(x_0)]\}^2\\
&+\frac{1}{nh^5}\sum_{i=1}^{n} K\left(\frac{Y_i-y_0}{h}\right) K^{(3)}\left(\frac{\widehat{G}^{*}(X_i)-G(x_0)}{h}\right)\\
&\quad\{[\widehat{G}(X_i)-G(X_i)]-[\widehat{G}(x_0)-G(x_0)]\}^3\\
=& J_1+J_2^{\diamond}+J_3^{\diamond}+{J_4}^{\diamond *}
\end{aligned}$$

where $\widehat{G}^{*}(X_i)$ is an $r.v.$ between $\widehat{G}(X_i)$ and $G(X_i)$ with probability 1, and ${G_n}^{*}(x_0)$ is between $G_n(x_0)$ and $G(x_0)$ with probability 1. The moments of J_1 are given in the proof of Theorem 1. We need to calculate the expectation of $J_2^{\diamond}$ again in this case. By using conditional expectation, we have the following:

$$\begin{aligned}
&E[J_2^{\diamond}]\\
=&\frac{1}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n} E\Bigg[K\left(\frac{Y_i-y_0}{h}\right) K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&\quad\left\{\left[W\left(\frac{X_i-X_j}{h}\right)-G(X_i)\right]-\left[W\left(\frac{x_0-X_j}{h}\right)-G(x_0)\right]\right\}\Bigg]\\
=&\frac{1}{nh^3}E\Bigg[K\left(\frac{Y_i-y_0}{h}\right) K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\\
&\quad\times\sum_{j=1}^{n} E\left[\left\{W\left(\frac{X_i-X_j}{h}\right)-G(X_i)\right\}-\left\{W\left(\frac{x_0-X_j}{h}\right)-G(x_0)\right\}\,\Big|\, Y_i,X_i\right]\Bigg]\\
=&\frac{1}{nh^3}E\left[K\left(\frac{Y_i-y_0}{h}\right) K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\left[\{nh^2A_{1,2}\left[g'(X_i)-g'(x_0)\right]\}+O(nh^4+1)\right]\right]\\
=&\iint A_{1,2}K\left(\frac{z-y_0}{h}\right)K'(u)\left[g'(\psi(u))-g'(x_0)\right]\frac{f(\psi(u),z)}{g(\psi(u))}\mathrm{d}z\mathrm{d}u+O\left(h^4+\frac{1}{nh}\right)
\end{aligned}$$

$$=h^2 A_{1,2}\frac{f(x_0,y_0)g''(x_0)}{g'(x_0)}+O\left(h^3+\frac{1}{nh}\right).$$

Now, we can calculate $E[(J_2^\diamond)^2]$ and $V[J_2^\diamond]$. The squared value of $J_2^\diamond$ is given by

$$\begin{aligned}(J_2^\diamond)^2 =& \frac{1}{n^4h^6}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{m=1}^n K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\\ &\times K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\left\{\left[W\left(\frac{X_i-X_j}{h}\right)-G(X_i)\right]-\left[W\left(\frac{x_0-X_j}{h}\right)-G(x_0)\right]\right\}\\ &\times\left\{\left[W\left(\frac{X_k-X_m}{h}\right)-G(X_k)\right]-\left[W\left(\frac{x_0-X_m}{h}\right)-G(x_0)\right]\right\}\\ =:& \frac{1}{n^4h^6}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{m=1}^n \Xi^\diamond(i,j,k,m).\end{aligned}$$

For the terms in which all (i,j,k,m) are different, we can show that

$$\begin{aligned}&E[\Xi^\diamond(i,j,k,m)]\\ =&\frac{1}{h^6}E\left[E\left[K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\right.\right.\\ &\times\left\{\left[W\left(\frac{X_i-X_j}{h}\right)-G(X_i)\right]-\left[W\left(\frac{x_0-X_j}{h}\right)-G(x_0)\right]\right\}\\ &\left.\left.\times\left\{\left[W\left(\frac{X_k-X_m}{h}\right)-G(X_k)\right]-\left[W\left(\frac{x_0-X_m}{h}\right)-G(x_0)\right]\right\}\,\middle|\,(X_i,Y_i),(X_k,Y_k)\right]\right]\\ =&\left(\frac{1}{h^3}E\left[K\left(\frac{Y_i-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\left[\left\{h^2A_{1,2}\left[g'(X_i)-g'(x_0)\right]\right\}+O(h^4)\right]\right]\right)^2\\ =&\left(h^2A_{1,2}\frac{f(x_0,y_0)g''(x_0)}{g'(x_0)}+O\left(h^3\right)\right)^2.\end{aligned}$$

For the terms in which $i=j$ and all (i,k,m) are different,

$$\begin{aligned}&E[\Xi^\diamond(i,j,k,m)]\\ =&\frac{1}{nh^6}E\left[E\left[K\left(\frac{Y_i-y_0}{h}\right)K\left(\frac{Y_k-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)K'\left(\frac{G(X_k)-G(x_0)}{h}\right)\right.\right.\\ &\quad\times\left\{\left[W\left(\frac{X_i-X_i}{h}\right)-G(X_i)\right]-\left[W\left(\frac{x_0-X_i}{h}\right)-G(x_0)\right]\right\}\\ &\quad\left.\left.\times\left\{\left[W\left(\frac{X_k-X_m}{h}\right)-G(X_k)\right]-\left[W\left(\frac{x_0-X_m}{h}\right)-G(x_0)\right]\right\}\,\middle|\,(X_i,Y_i),(X_k,Y_k)\right]\right]\\ =&\frac{1}{nh^3}E\left[K\left(\frac{Y_i-y_0}{h}\right)K'\left(\frac{G(X_i)-G(x_0)}{h}\right)\left\{\left[\frac{1}{2}-G(X_i)\right]\right.\right.\\ &\qquad\left.\left.-\left[W\left(\frac{x_0-X_i}{h}\right)-G(x_0)\right]\right\}\right]\times\left(h^2A_{1,2}\frac{f(x_0,y_0)g''(x_0)}{g'(x_0)}+O\left(h^3+\frac{1}{nh}\right)\right)\\ =&O\left(\frac{h}{n}+\frac{1}{n^2h^2}\right),\end{aligned}$$

Similarly, for the terms in which $i = k$ and all (i, j, m) are different,

$$\begin{aligned}
&E[\Xi^{\diamond}(i,j,k,m)\left(\frac{Y_i - y_0}{h}\right) K\left(\frac{Y_i - y_0}{h}\right) K'\left(\frac{G(X_i) - G(x_0)}{h}\right) K'\left(\frac{G(X_i) - G(x_0)}{h}\right) \\
&\quad \times \left\{\left[W\left(\frac{X_i - X_j}{h}\right) - G(X_i)\right] - \left[W\left(\frac{x_0 - X_j}{h}\right) - G(x_0)\right]\right\} \\
&\quad \times \left\{\left[W\left(\frac{X_i - X_m}{h}\right) - G(X_i)\right] - \left[W\left(\frac{x_0 - X_m}{h}\right) - G(x_0)\right]\right\} \Bigg| (X_i, Y_i) \Bigg]\Bigg] \\
=&\frac{1}{nh^6} E\left[K^2\left(\frac{Y_i - y_0}{h}\right)(K')^2\left(\frac{G(X_i) - G(x_0)}{h}\right)\left[\left\{h^2 A_{1,2}\left[g'(X_i) - g'(x_0)\right]\right\} + O(h^4)\right]^2\right] \\
=&O\left(\frac{h^2}{n}\right).
\end{aligned}$$

For the terms in which $i = m$ and all (i, j, k) are different,

$$\begin{aligned}
&E[\Xi^{\diamond}(i,j,k,m)] \\
=&\frac{1}{h^6} E\left[E\left[K\left(\frac{Y_i - y_0}{h}\right) K\left(\frac{Y_k - y_0}{h}\right) K'\left(\frac{G(X_i) - G(x_0)}{h}\right) K'\left(\frac{G(X_k) - G(x_0)}{h}\right)\right.\right. \\
&\quad \times \left\{\left[W\left(\frac{X_i - X_j}{h}\right) - G(X_i)\right] - \left[W\left(\frac{x_0 - X_j}{h}\right) - G(x_0)\right]\right\} \\
&\quad \times \left\{\left[W\left(\frac{X_k - X_i}{h}\right) - G(X_k)\right] - \left[W\left(\frac{x_0 - X_i}{h}\right) - G(x_0)\right]\right\} \Bigg| (X_i, Y_i) \Bigg]\Bigg] \\
=&O\left(\frac{h}{n}\right).
\end{aligned}$$

Finally, for the terms in which $j = m$ and all (i, j, k) are different, we have

$$E[\Xi^{\diamond}(i,j,k,m)] = O\left(\frac{1}{nh}\right),$$

and

$$V[J_2^{\diamond}] = O\left(\frac{1}{nh}\right).$$

We can see that $J_3^{\diamond} + J_4^{\diamond *}$ is asymptotically negligible. The main bias of $\widehat{f}^{\diamond}_{Y|x_0}(y_0)$ comes from $J_1 + J_2^{\diamond}$. Since $E[J_1 J_2]$ is given by

$$\begin{aligned}
&\frac{1}{n^3 h^3} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} K\left(\frac{Y_i - y}{h}\right) K\left(\frac{Y_j - y_0}{h}\right) K'\left(\frac{G(X_i) - G(x_0)}{h}\right) \\
&\quad K'\left(\frac{G(X_j) - G(x_0)}{h}\right)\left\{\left[W\left(\frac{X_j - X_k}{h}\right) - G(X_j)\right] - \left[W\left(\frac{x_0 - X_k}{h}\right) - G(x_0)\right]\right\} \\
=&E[J_1]E[J_2] + O\left(\frac{1}{n}\right),
\end{aligned}$$

the asymptotic variance is the same as the discrete version $\widehat{f}^{\star}_{Y|x_0}(y_0)$, and Theorem 2 has been proved. □

Received May 27, 2019
Revised January 10, 2020