SUPPLEMENT TO "Data sharpening method in regression confidence band"

Xuyang He[a,*], Yuexiang Jiang[b]

[a]Department of Mathematics, Zhejiang University, Hangzhou, 310027, China
[b]College of Economics, Zhejiang University, Hangzhou, 310027, China

Abstract

This supplemental material consists largely of the theoretical proofs.

1. Technical details

For the sake of convenience, we define a rule for all the values in this article, just like B, δ_N, δ. No superscript means the value belongs to the new confidence band, superscript $'$ means the value belongs to the old confidence band, and superscript $*$ means the value belongs to the new resampling confidence band. For example, B is a new confidence band, δ'_N is a nominal confidence from old confidence band B', and $\hat{m}^*$ is the function estimator of resampling confidence band B^*.

1.1. Proof of Theorem 4.1

In this section we prove the main theoretical result, Theorem 4.1. The proof is divided into three parts. The first part proves the relationship between nominal confidence δ_N and real confidence δ_R. The second part proves that when the nominal confidence is the same, the widths of the confidence bands B and B' are also of the same order. The third part proves that when the real confidence is the same, the width of the confidence band B is narrower than B' for some δ_N.

1.1.1. part i

Without loss of generality, we assume the nominal confidence of B and B' are the same, $\delta_N = \delta'_N = \delta$. Following the discussion in Sec. 2.2, the confidence at point x is

$$\begin{aligned} P\{(X, m(X)) \in B(\alpha) | X = x\} &= \mathcal{C}_n(x, \alpha) \\ &= \Phi\left(Z_{(1+\delta_N)/2} + \frac{b(x)}{s(x)\hat{\sigma}}\right) - \Phi\left(-Z_{(1+\delta_N)/2} + \frac{b(x)}{s(x)\hat{\sigma}}\right) \end{aligned}$$

*Corresponding author
Email address: hxy220284@zju.edu.cn (Xuyang He)

Let $\pi_\alpha(x) = \Phi(x + Z_\alpha) - \Phi(x - Z_\alpha)$. We have proved that

$$\frac{d\pi_\alpha(x)}{dx} = 0 \iff x = 0, \tag{1}$$

$\pi_\alpha(0)$ is the global maximum, and $\pi_\alpha(x)$ is strictly monotonically decreasing in $|x|$. Since δ_N is the same, the probability that the confidence band holds at point x depends mainly on the item $b(x)/s(x)\hat{\sigma}$.

According to Thm. 3.1 of He et al. (2018), the data-sharpening estimator $\hat{m}$ has the property

$$b(x) = O(h^r), s(x) = O\left(\frac{1}{\sqrt{nhf_X(x)}}\right)$$

if m is $r-2$ times differentiable. Simultaneously the properties of the local-linear estimator

$$b'(x) = O(h^2), s'(x) = O\left(\frac{1}{\sqrt{nhf_X(x)}}\right)$$

are proved in Fan (1992). The estimation of σ is an independent problem, so we may assume that B and B' use the same $\hat{\sigma}$ estimation. Thus

$$\lim_{n\to\infty, h(n)\to 0} \frac{b(x)/s(x)\hat{\sigma}}{b'(x)/s'(x)\hat{\sigma}} = O(h^{r-2}) \tag{2}$$

In particular, when $r = 2$, the data-sharpening estimator degenerates to the local-linear case, and B and B' are the same. In other cases, $O(h^{r-2}) \to 0$, which means for $\forall x \in I^*$, $\exists N \in \mathbb{N}$, $\forall n > N$

$$\left|\frac{b(x)}{s(x)\hat{\sigma}}\right| < \left|\frac{b'(x)}{s'(x)\hat{\sigma}}\right| \tag{3}$$

Define

$$d(x) = \left|\frac{b'(x)}{s'(x)\hat{\sigma}}\right| - \left|\frac{b(x)}{s(x)\hat{\sigma}}\right|;. \tag{4}$$

Note that $d(x)$ is continuous with respect to x, so that $\forall\epsilon > 0$, $\exists\eta > 0$, if $|x_1 - x_2| < \eta$, $|d(x_1) - d(x_2)| < \epsilon$.

For any $x \in I^*$, choose N_x to satisfy Eq. 3. Choose $\epsilon_x = d(x)$, $\exists\eta_x > 0$, $\forall x'$ in open interval $(x-\eta_x, x+\eta_x)$ satisfies $d(x') > 0$ when $n > N_x$. Consider all open intervals $\{(x-\eta_x, x+\eta_x)\}_{x\in I^*}$, $I^* \subseteq \cup(x-\eta_x, x+\eta_x)$. These intervals constitute an open cover of I^*. By the HeineBorel theorem, we can choose a finite number of intervals $\{(x_j - \eta_{x_j}, x_j + \eta_{x_j})\}_{j=1}^M$, $I^* \subseteq \cup_{j=1}^M (x_j - \eta_{x_j}, x_j + \eta_{x_j})$, which is a finite open cover of I^*. Let $N_0 = \max_{j=1,2\ldots M} N_{x_j}$. When $n > N_0$, $d(x) > 0$ is true for any interval $(x - \eta_x, x + \eta_x)$, thus is true on I^*.

Note the real confidence $\delta_R = \int \mathcal{C}_n(x,\alpha) f_X(x)dx$. Because $d(x) > 0$ always holds, $\mathcal{C}_n(x,\alpha) > \mathcal{C}'_n(x,\alpha)$, $\delta_R > \delta'_R$ always holds, too.

Next we consider the case $\delta_R = \delta'_R$. For the old confidence band B', a simple fact is

$$\frac{d \int \mathcal{C}_n(x,\alpha) f_X(x) dx}{d\delta_N} > 0 \,. \tag{5}$$

We proved this in Thm. 2.2. Let $\delta_N = 1 - \alpha$ be the nominal confidence of B, and construct an old band B'' with nominal confidence $\delta''_N = \delta_N$. From the discussion above we know $\delta_N = \delta''_N$, so $\delta_R > \delta''_R$. Next we compare two similar confidence bands B' and B''. We know $\delta'_R = \delta_R > \delta''_R$, and by Eq. 5, δ_R monotonically increases with respect to δ_N. Because $\delta'_R > \delta''_R$, $\delta'_N > \delta''_N = \delta_N$.

1.1.2. part ii

In this part we discuss the width of the confidence band when the nominal confidence is the same. Assume $\delta_N = \delta'_N = \delta$. From the discussion in Section 2, the width of B is

$$\mathcal{W}(B)(x) = 2 * s(x)\hat{\sigma} * Z_{1/2+\delta/2}$$

where $\hat{\sigma}$ is the estimator of σ and Z_β the β-quantile of the standard normal distribution. Note

$$\frac{\mathcal{W}(B)(x)}{\mathcal{W}'(B)(x)} = \frac{s(x)}{s'(x)} \,.$$

We know from He et al. (2018) that

$$s(x) \to \frac{\int \left\{K(u) - \int K'(v) N_r(x - hv, u - v) f_X(x - hv) dv\right\}^2 du}{\sqrt{nhf_X(x)}},$$

$N_r(x, y)$ is a function determined by X_i, Y_i and K, and Fan (1992) proves that $s'(x) = \frac{\int K^2(u)du}{\sqrt{nhf_X(x)}}$. When $r = 2$, $s(x) = s'(x)$ and $C = 1$, otherwise

$$\begin{aligned}\lim_{n\to\infty} \frac{s(x)}{s'(x)} &= \frac{\int \left\{K(u) - \int K'(v) N_r(x - hv, u - v) f_X(x - hv) dv\right\}^2 du/\sqrt{nhf_X(x)}}{\int K^2(u)du/\sqrt{nhf_X(x)}} \\ &= C\end{aligned}$$

1.1.3. part iii

Assume B, B' and B'' are the confidence bands from 1.1.1, $\delta_N = \delta''_N = \delta$, $\delta'_R = \delta_R$. The relationship between B and B' is what we want to study. From 1.1.2 we know

$$\lim_{n\to\infty} \frac{s(x)}{s''(x)} = C$$

Assume $\Delta = \delta'_R - \delta''_R > 0$. We use B'' as a bridge in the proof.

$$\begin{aligned}
\lim_{n\to\infty} \frac{\mathcal{W}(B)(x)}{\mathcal{W}'(B)(x)} &= \lim_{n\to\infty} \frac{\mathcal{W}(B)(x)}{\mathcal{W}''(B)(x)} * \lim_{n\to\infty} \frac{\mathcal{W}(B'')(x)}{\mathcal{W}'(B)(x)} \\
&= C * \lim_{n\to\infty} \frac{\mathcal{W}(B'')(x)}{\mathcal{W}'(B)(x)} \\
&= C * \frac{2 * s(x)\hat{\sigma} * Z_{1/2+\delta''/2}}{2 * s(x)\hat{\sigma} * Z_{1/2+\delta'/2}} \\
&= C * \frac{Z_{1/2+\delta''/2}}{Z_{1/2+\delta'}/2} \\
&= C * \frac{Z_{1/2+\delta''/2}}{Z_{1/2+(\delta''+\Delta)/2}}
\end{aligned}$$

Let $C * \frac{Z_{1/2+\delta/2}}{Z_{1/2+(\delta+\Delta)/2}} < 1$. We try to find the appropriate $\delta''_N = \delta$. By the definition of quantile we have

$$\begin{aligned}
\int_0^{Z_{1/2+\delta/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du =& \delta/2 \\
\int_{Z_{1/2+\delta/2}}^{Z_{1/2+(\Delta+\delta)/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du =& \Delta/2
\end{aligned}$$

Let $g(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$, and let $g(u)$ be continuous. By the monotonically decreasing nature of g on $(0, +\infty)$, $\exists u_1 \in (0, Z_{1/2+\delta/2})$,

$$\begin{aligned}
\delta/2 =& Z_{1/2+\delta/2} * g(u_1) \\
>& Z_{1/2+\delta/2} * g(Z_{1/2+\delta/2})
\end{aligned}$$

and $\exists u_2 \in (Z_{1/2+\delta/2}, Z_{1/2+(\Delta+\delta)/2})$,

$$\begin{aligned}
\Delta/2 =& (Z_{1/2+(\Delta+\delta)/2} - Z_{1/2+\delta/2}) * g(u_2) \\
<& (Z_{1/2+(\Delta+\delta)/2} - Z_{1/2+\delta/2}) * g(Z_{1/2+\delta/2})
\end{aligned}$$

Making a simple transformation,

$$\begin{aligned}
Z_{1/2+\delta/2} <& \frac{\delta}{2 * g(Z_{1/2+\delta/2})} \\
Z_{1/2+(\Delta+\delta)/2} - Z_{1/2+\delta/2} >& \frac{\Delta}{2 * g(Z_{1/2+\delta/2})}
\end{aligned}$$

Obviously, both sides of the inequality sign are positive. So

$$\begin{aligned}
\frac{Z_{1/2+(\Delta+\delta)/2}}{Z_{1/2+\delta/2}} &> \frac{\Delta + \delta}{\delta} \\
C * \frac{Z_{1/2+\delta/2}}{Z_{1/2+(\delta+\Delta)/2}} &< C * \frac{\delta}{\Delta + \delta} .
\end{aligned}$$

If

$$C * \frac{\delta}{\Delta + \delta} < 1$$

we have

$$\delta < \frac{\Delta}{C - 1}\ . \tag{6}$$

Next we discuss Δ. We know $\Delta = \delta'_R - \delta''_R$, and the corresponding confidence bands B' and B'' are same band with different δ_N. So the relation functions $\mathcal{Q}_\xi$ are the same,

$$\begin{aligned}\Delta =& \mathcal{Q}_\xi(\delta'_N) - \mathcal{Q}_\xi(\delta''_N)\\ =& \mathcal{Q}_\xi^{(1)}(\delta_0) * (\delta'_N - \delta''_N)\end{aligned}$$

where $\delta_0 \in (\delta''_N, \delta'_N)$. Substituting Δ into Eq. 6, we have

$$\delta < \frac{\delta'_N}{(C-1)/D + 1} \tag{7}$$

for some $D > 0$.

If $C \leq 1$, (7) is always true, which means $\mathcal{W}(B)(x) < \mathcal{W}'(B)$ always holds. In other cases, this is true only if δ_N is within a certain range.

Here we can only find a narrow range of δ, but this is not the true performance of the new confidence band. In fact, in the simulation, the range of δ where the new confidence band performs better than the old one is very wide. The shortcoming here is mainly because our estimates of the quantiles of the normal distribution Z_β are not accurate enough, which we will improve in subsequent studies. The result here is just a sufficient condition.

1.2. Proof of Theorem 4.2

In this section we prove the strong approximation of $\hat{m}^*$. For the sake of clarity we give the proof only in the case $r = 2$.

We separate the derivation into two parts. First, because $E\{\hat{m}^*(x)|Z\} = m(x) + b^*(x)$ and $\hat{m}(x) = m(x) + b(x) + \Psi_x$, the formula $E\{\hat{m}^*(x)|Z\} - \hat{m}(x)$ equals

$$b^*(x) - b(x) - \Psi_x \tag{8}$$

and $b^*(x) - b(x)$ is not a random variable, $E\{\Psi_x\} = 0$. So

$$E\left\{E\{\hat{m}^*(x)|Z\} - \hat{m}(x)\right\} = b^*(x) - b(x) \tag{9}$$

We also know

$$\begin{aligned}\hat{m}(x) =& \frac{1}{nh}\sum_{i=1}^{N} \frac{S_{n,2}(x) - S_{n,1}(x)\frac{x - X_i}{h}}{S_{n,2}(x)S_{n,0}(x) - S_{n,1}(x)^2} K\left(\frac{x - X_i}{h}\right) Y_i\\ =& \frac{1}{n}\sum_{i=1}^{N} A_i(x) * Y_i\end{aligned} \tag{10}$$

In the process of bootstrapping, we replaced Y_i by $\hat{m}(X_i) + \tilde{\epsilon}_i$ and $E\{\tilde{\epsilon}_i\} = 0$. So

$$
\begin{aligned}
E\{\hat{m}^*(x)\} =& \frac{1}{n} \sum_{i=1}^{N} A_i(x) * E\{\hat{m}(X_i) + \tilde{\epsilon}_i\} \\
=& \frac{1}{n} \sum_{i=1}^{N} A_i(x) * \hat{m}(X_i) \\
=& \frac{1}{n^2} \sum_{i=1}^{N} A_i(x) \sum_{j=1}^{N} A_j(x) * (m(X_j) + \epsilon_j) \\
=& \frac{1}{n^2} \sum_{i=1}^{N} A_i(x) \sum_{j=1}^{N} A_j(x) * m(X_j) + \frac{1}{n^2} \sum_{i=1}^{N} A_i(x) \sum_{j=1}^{N} A_j(x) * \epsilon_j \\
=& g(x) + e_1
\end{aligned}
\tag{11}
$$

Consider the first part g,

$$
\begin{aligned}
g(x) =& \frac{1}{n^2} \sum_{i=1}^{N} A_i(x) \sum_{j=1}^{N} A_j(x) \\
& * \left[m(X_i) + (X_i - X_j) m'(X_i) + \frac{1}{2}(X_i - X_j)^2 m''(X_i) + R(x, X_i) \right] \\
=& \frac{1}{n} \sum_{i=1}^{N} A_i(x) \left[m(X_i) + \frac{1}{2} h^2 m''(X_i) + R(x, X_i) \right] \\
=& \frac{1}{n} \sum_{i=1}^{N} A_i(x) \left[Y_i - \epsilon_i + \frac{1}{2} h^2 m''(X_i) + R(x, X_i) \right] \\
=& \hat{m}(x) - e_2 + \frac{h^2 \int u^2 K(u) du}{2} + R(x)
\end{aligned}
\tag{12}
$$

where

$$R(x) = \frac{1}{n} \sum_{i=1}^{N} A_i(x) * R(x, X_i)$$

is the sum of the Peano residual,

$$e_2 = \frac{1}{n} \sum_{i=1}^{N} A_i(x) * \epsilon_i$$

.

The bound of $R(x)$ is given in Supplement 1.3 of He et al. (2018), for some $C > 0$,

$$P\{\sup_{x \in I^*} |R(x)| > C h^{r+1}\} = O(h^{-\lambda})$$

For any $\lambda > 0$, because $E\{e_1\} = E\{e_2\} = 0$, note that the order of $R(x)$ is higher than the constant term so we omit it. So

$$E\left\{E\{\hat{m}^*(x)|Z\} - \hat{m}(x)\right\} = \frac{h^2 \int u^2 K(u)du}{2} = b^*(x) - b(x) \tag{13}$$

1.2.1. Proof of Lemma 2

The proof of Lemma 2 is trivial. We know from 1.2 that

$$E\{\hat{m}^*(x)|Z\} - \hat{m}(x) = b^*(x) - b(x) - \Psi_x = \frac{h^2 \int u^2 K(u)du}{2} + e_1 - e_2 \tag{14}$$

and $-\Psi_x = e_1 - e_2$. Noting Lemma 1, we can use the same transform as in Sec 1.2 and compare the two formulas. The constant term $b^* - b$ is different while the random term $e_1 - e_2$ are same. In Lemma 1 it is proved that $e_1 - e_2 \to D(nh)^{-1/2} f_X(x)^{-1/2} W(x/h)$ when $n \to \infty$, and the e_1, e_2 are same as in Lemma 2, so Lemma 2 is true.

Combining the above-mentioned random variable part Ψ_x and non-random variable part $b^* - b$, the proof of Theorem 4.2 is completed.

References

Fan, J., 1992. Design-adaptive nonparametric regression. Journal of the American Statistical Association 87, 998–1004.

He, X., Jiang, Y., Wang, J., 2018. High-order data sharpening with dependent errors for regression bias reduction. Communications in Statistics - Theory and Methods 0 (0), 1–8.
URL https://doi.org/10.1080/03610926.2018.1520885