SEMI-SUPERVISED LEARNING BASED ON NW ESTIMATOR - A TRADE-OFF

Junshi Wang

November 11, 2022

The University of Hong Kong
Department of Statistics and Actuarial Science

Contents

1 The case of MSE

Recall the definition of hybrid estimator,

$$\hat{y}_c(x) = \lambda \hat{m}(x) + (1-\lambda)\hat{r}(x) \tag{1}$$

$$\begin{aligned}\hat{m}(x) = m(x) + Eh_n^2 + O(h_n^4) + (nh_n)^{-1/2}h_n^2[FA+GB] \\ + (nh_n)^{-1/2}\frac{A-m(x)B}{p(x)} + O_p(\frac{1}{nh_n})\end{aligned} \tag{2}$$

$$\begin{aligned}\hat{r}(x) = \hat{m}(x) + Hg_m^2 + O_p(\frac{g_m^2}{\sqrt{nh_n^5}} + g_m^4) + (mg_m)^{-1/2}g_m^2[IC+JD] \\ + (mg_m)^{-1/2}g_m\frac{C-\hat{m}(x)D}{q(x)g_m} + O_p(\frac{1}{m})\end{aligned} \tag{3}$$

where $E = \frac{m'(x)p'(x)\mu_2(k)}{p(x)} + \frac{m''(x)\mu_2(k)}{2}, H = \frac{m'(x)q'(x)\mu_2(k)}{q(x)} + \frac{m''(x)\mu_2(k)}{2}$. F, G, I and J are all $O(1)$ and A, B, C, D are all $O_p(1)$.

To minimize $Bias(\hat{y}_c(x))$, we choose λ to be $1 + \frac{h_n^2 E}{g_m^2 H} \approx 1 + \frac{h_n^2}{g_m^2}$. Then we are able to derive the bias and variance of hybrid estimator.

$$\begin{aligned}\hat{y}_c(x) = m(x) + O(h_n^4) + (nh_n)^{-1/2}h_n^2[FA+GB] + (nh_n)^{-1/2}\frac{A-m(x)B}{p(x)} + O_p(\frac{1}{nh_n}) \\ - \frac{h_n^2}{g_m^2}[O_p(\frac{g_m^2}{\sqrt{nh_n^5}} + g_m^4) + (mg_m)^{-1/2}g_m^2[IC+JD] + (mg_m)^{-1/2}g_m\frac{C-\hat{m}(x)D}{q(x)g_m} + O_p(\frac{1}{m})]\end{aligned}$$

$$Bias(\hat{y}_c(x)) = O(h_n^4 + \frac{1}{\sqrt{nh_n}} + h_n^2 g_m^2 + \frac{h_n^2}{mg_m^2})$$

$$\begin{aligned}Var(\hat{y}_c(x)) &= Var\{(nh_n)^{-1/2}\frac{A-m(x)B}{p(x)} - [O_p(\frac{1}{\sqrt{nh_n}} + h_n^2 g_m^2) + \frac{h_n^2(mg_m)^{-1/2}}{g_m}\frac{C-\hat{m}(x)D}{q(x)g_m}]\} \\ &= O(Var\{(nh_n)^{-1/2}\frac{A-m(x)B}{p(x)}\} + Var\{O_p(\frac{1}{\sqrt{nh_n}} + h_n^2 g_m^2)\} + Var\{\frac{h_n^2}{(mg_m^3)^{-1/2}}\frac{C-\hat{m}(x)D}{q(x)g_m}]\}) \\ &= O(\frac{1}{nh_n} + h_n^4 g_m^4 + \frac{h_n^4}{mg_m^3})\end{aligned}$$

Then the mean square error(MSE) can be expressed as,

$$MSE = O(h_n^8 + \frac{1}{nh_n} + h_n^4 g_m^4 + \frac{h_n^4}{mg_m^3})$$

with restrictions $nh_n \to \infty, mg_m \to \infty, nh_n^5 \to \infty$ and $h_n, g_m \to 0$. The optimal (here what optimal means is optimal worst case or minimax, since we are using big O notation) value is $O(n^{-8/9} + n^{-4/5}m^{-4/35})$, where $h = n^{-1/9}$ or $n^{-1/5}m^{4/35}$ and $g = m^{-1/7}$.

2 The case of Coverage Error

Let r_n denote the normalizing constant. We have $\hat{y_c}(x)$ as an asymptotic normal random variable which can be used to obtain the confidence interval for a point estimation.

$$Z_n = r_n(\hat{y_c}(x) - m(x)) \overset{approx}{\sim} \mathbf{N}(b, \sigma^2)$$

In general, $r_n^2 = MSE = O(h_n^8 + \frac{1}{nh_n} + h_n^4 g_m^4 + \frac{h_n^4}{mg_m^3})$. But the concept of confidence interval only makes sense when $\sigma^2 > 0$. As we can see, bias term is very complicated so hard to estimate. Therefore, we propose a naive estimator $\hat{b} = 0$.

2.1 $b = o(1)$

As we can observe from the expression of MSE, $b = o(1)$ if $\frac{h^4}{mg^3} \succ h_n^8 + \frac{1}{nh_n} + h_n^4 g_m^4$. In this case, $r_n^2 = \frac{mg_m^3}{h^4}$.

$$b = O_p(\delta_b) = Bias(\hat{y_c}(x))r_n = O(h_n^2 m^{1/2} g_m^{3/2} + n^{-1/2} h_n^{-5/2} m^{1/2} g_m^{3/2} + m^{1/2} g_m^{7/2} + m^{-1/2} g_m^{1/2})$$

Regarding the case of variance, we can do make a better estimation.

$$\begin{aligned}
Var(\hat{y_c}(x)) &= Var\{(nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} - [O_p(\frac{1}{\sqrt{nh_n}} + h_n^2 g_m^2) + \frac{h_n^2(mg_m)^{-1/2}}{g_m}\frac{C - \hat{m}(x)D}{q(x)g_m}]\} \\
&= Var\{(nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} - \frac{h_n^2(mg_m)^{-1/2}}{g_m}\frac{C - \hat{m}(x)D}{q(x)g_m}\} + Var(O_p(\frac{1}{\sqrt{nh_n}} + h_n^2 g_m^2)) \\
&- 2Cov\{(nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} - \frac{h_n^2(mg_m)^{-1/2}}{g_m}\frac{C - \hat{m}(x)D}{q(x)g_m}], O_p(\frac{1}{\sqrt{nh_n}} + h_n^2 g_m^2)\} \\
&= E(Var(\frac{h_n^2(mg_m)^{-1/2}}{g_m}\frac{C - \hat{m}(x)D}{q(x)g_m}|X, Y)) + Var((nh_n)^{-1/2}\frac{A - m(x)B}{p(x)}) \\
&+ O(\frac{1}{nh_n} + h_n^4 g_m^4 + \frac{h_n^2(mg_m)^{-1/2}}{g_m}\frac{1}{\sqrt{nh_n}} + \frac{h_n^2(mg_m)^{-1/2}}{g_m}h_n^2 g_m^2) \\
&= E[(\hat{m}'(x)^2)\frac{\sigma_k^2}{q(x)}\frac{h_n^4}{mg_m^3} + O(\frac{h^4}{mg_m})] + [\frac{1}{nh_n}\sigma_\varepsilon^2\frac{r(k)}{p(x)} + O(\frac{h_n}{n})]
\end{aligned}$$

$$\begin{aligned}
\sigma^2 &= E[\hat{m}'(x)^2\frac{\sigma_k^2}{q(x)} + O(g_m^2)] + [\frac{mg_m^3}{nh_n^5}\sigma_\varepsilon^2\frac{r(k)}{p(x)} + O(\frac{mg_m^3}{nh_n^3})] \\
\hat{\sigma^2} &= \hat{m}'(x)^2\frac{\sigma_k^2}{\hat{q}(x)} + \frac{mg_m^3}{nh_n^5}\hat{\sigma_\varepsilon^2}\frac{r(k)}{\hat{p}(x)}
\end{aligned}$$

Let $\hat{p}(x) = p(x) + O_p(\delta_p), \hat{q}(x) = q(x) + O_p(\delta_q), \hat{\sigma_\varepsilon^2} = \sigma_\varepsilon^2 + O_p(\delta_\varepsilon)$.

$$
\begin{aligned}
\hat{\sigma^2} &= \sigma^2 + O(\delta_\sigma) \\
&= \sigma^2 + O_p(h^2 + \frac{1}{\sqrt{nh_n^3}}) + O(h_n^2 + \frac{1}{\sqrt{nh_n^3}} + \frac{mg_m^3}{nh_n^3}) + O_p(\delta_q + \frac{mg_m^3}{nh_n^5}\delta_\varepsilon + \frac{mg_m^3}{nh_n^5}\delta_p)
\end{aligned}
$$

We can derive the coverage error subsequently, but the true coverage probability can only be estimated with the following theorem.

There are two possible ways to formulate the coverage error. First, estimating the bias and variance of standardized hybrid estimator directly. But one problem we may encounter is that although the hybrid estimator follows an asymptotic normal distribution,its approximation error(approximating the asymptotic normal using normal) is hard to calculate using the Berry-Esséen theorem.

$$
\begin{aligned}
&\mathbf{P}\{\frac{r_n(\hat{y}_c - m(x)) - \hat{b}}{\hat{\sigma^2}} \leq z_\alpha\} \\
=&\mathbf{P}\{\frac{r_n(\hat{y}_c - m(x)) - (b + O_p(\delta_b))}{\sigma^2 + O_p(\delta_\sigma)} \leq z_\alpha\} \\
=&\mathbf{P}\{\frac{r_n(\hat{y}_c - m(x)) - b}{\sigma^2} \leq z_\alpha + O_p(\delta_b + \delta_\sigma)\} \\
=&\mathbf{P}\{Z_n \leq z_\alpha\} + O(\delta_b + \delta_\sigma) \\
=&\Phi(z_\alpha) + O(\delta_n + \delta_b + \delta_\sigma)
\end{aligned}
$$

The second inequality is due to delta method. And δ_n is the approximation error.

Second method is more or less similar, but try to make use of the Berry-Esseen theorem directly by conditioning on (X,Y). Considering writing the self-supervised estimator in terms of a triangular array $\{\xi_{j,m}\}_{j=1}^n$:

$$
\begin{aligned}
\frac{C - \hat{m}(x)D}{q(x)g_m} &= (mg_m^3 q^2(x))^{-1/2} \sum_{j=1}^m [\hat{m}(u_j)K(\frac{x-u_j}{g_m}) - \hat{m}(x)K(\frac{x-u_j}{g_m}) \\
&\quad + \mathbf{E}(\hat{m}(x)K(\frac{x-u_j}{g_m})) - \mathbf{E}(\hat{m}(u_j)K(\frac{x-u_j}{g_m})|X,Y)] \\
&= \frac{1}{m}\sum_{j=1}^m \xi_{j,m}
\end{aligned}
$$

It is possible to show that

$$
\begin{aligned}
\mathbf{E}[|\xi_{j,m}|^2] &= \Theta(\hat{m}'(x)^2 m g_m^2 \frac{\sigma_k^2}{q(x)}) \\
\mathbf{E}[|\xi_{j,m}|^3] &= O(\hat{m}'(x)^3 m^{3/2} g_m^{5/2})
\end{aligned}
$$

Then apply the Berry Esseen's Theorem,

$$
\begin{aligned}
&\mathbf{P}\{\frac{r_n(\hat{y}_c - m(x)) - \hat{b}}{\hat{\sigma}^2} \leq z_\alpha\} \\
=&\mathbf{P}\{r_n(\hat{y}_c - m(x) \leq z_\alpha \hat{\sigma}^2 + \hat{b}\} \\
=&\mathbf{P}\{r_n[(nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} - \frac{h_n^2}{(mg_m^3)^{-1/2}}\frac{C - \hat{m}(x)D}{q(x)g_m}] + O_p(\delta) \leq z_\alpha \sigma^2(1+\delta_\sigma) + O_p(\delta_b)\} \\
=&\mathbf{EP}\{\frac{r_n}{\sigma^2}[(nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} - \frac{h_n^2}{(mg_m^3)^{-1/2}}\frac{C - \hat{m}(x)D}{q(x)g_m}] \leq z_\alpha | X, Y\} + O(\delta + \delta_b + \delta_\sigma) \\
=&\mathbf{EP}\{\frac{1}{\sigma_{CD}^2}[-\frac{C - \hat{m}(x)D}{q(x)g_m}] \leq [z_\alpha + O_p([\frac{mg_m^3}{nh_n^5}]^{1/2})]\frac{\sigma^2}{\sigma_{CD}^2} | X, Y\} + O(\delta + \delta_b + \delta_\sigma) \\
=&\mathbf{E}\{\Phi(z_\alpha + O_p[\frac{(mg_m^3)^{1/2}}{(nh_n^5)^{1/2}}] + O(\delta_{CD})) + O_p(m^{-1/2}g_m^{-1/2})\} + O(\delta + \delta_b + \delta_\sigma) \\
=&\Phi(z_\alpha) + O(\delta + \delta_b + \delta_\sigma + \delta_{CD} + \frac{(mg_m^3)^{1/2}}{(nh_n^5)^{1/2}} + m^{-1/2}g_m^{-1/2})
\end{aligned}
$$

In the above equation,

$$\delta_{CD} = \sigma^2 - \sigma_{CD}^2 = \frac{mg_m^3}{nh_n^5}\sigma_\varepsilon^2\frac{r(k)}{p(x)} + \frac{mg_m^3}{nh_n^3} \sim \frac{mg_m^3}{nh_n^5} \tag{4}$$

$$\delta = r_n\{h_n^4 + \frac{1}{\sqrt{nh_n}} + h_n^2 g_m^2 + \frac{h_n^2}{(mg_m)^{1/2}} + \frac{h_n^2}{mg_m^2}\} \tag{5}$$

$$\delta_b = h_n^2(mg_m^3)^{1/2} + (nh_n^5)^{-1/2}(mg_m^3)^{1/2} + m^{-1/2}g_m^{1/2} \tag{6}$$

$$\delta_\sigma = h^2 + \frac{1}{\sqrt{nh_n^3}} + \frac{mg_m^3}{nh_n^3} + \delta_q + \frac{mg_m^3}{nh_n^5}\delta_\varepsilon + \frac{mg_m^3}{nh_n^5}\delta_p \tag{7}$$

$$\tag{8}$$

in general, $\delta_p = h_n^2 + (nh_n)^{-1/2}$, $\delta_q = g_m^2 + (mg_m)^{-1/2}$ and δ_ε should be greater than $n^{-1/2}$.

Finally, the coverage error is

$$Error_{coverage} = O(\delta + \delta_\sigma + \frac{(mg_m^3)^{1/2}}{(nh_n^5)^{1/2}} + (mg_m)^{-1/2}) \tag{9}$$

$$= O((mg_m^3)^{1/2}(h_n^2 + g_m^2 + (nh_n^5)^{-1/2}) + (mg_m)^{-1/2} + g_m) \tag{10}$$

2.2 $b = \Theta(1)$

In general, the true bias does not decay to zero, it is fairly difficult to estimate its value and result in $\delta_b \leq 1$. Hence this case is trivial.

$$\mathbf{P}\{\frac{r_n(\hat{y}_c - m(x)) - \hat{b}}{\hat{\sigma}^2} \leq z_\alpha\} - \Phi(z_\alpha)$$
$$\approx \Phi(z_\alpha + C) - \Phi(z_\alpha), \text{where C is some non-trivial constant}$$

2.3 Conclusion

The coverage error $Error_{coverage}$ can be $o(1)$ only if $\frac{h^4}{mg_m^3} \succ h_n^8 + \frac{1}{nh_n} + h_n^4 g_m^4$.

$$Error_{coverage} = O((mg_m^3)^{1/2}(h_n^2 + g_m^2 + (nh_n^5)^{-1/2}) + (mg_m)^{-1/2} + g_m)$$

Recall that the mean square error(MSE) can be expressed as,

$$MSE = O(h_n^8 + \frac{1}{nh_n} + h_n^4 g_m^4 + \frac{h_n^4}{mg_m^3})$$

Therefore, it's not hard to see the coverage error will always be $\Theta(1)$ when MSE is minimized. This implies that it is impossible to achieve the best coverage error and mean square error at the same time. However, some compromises could be made to strike a balance.

If we optimize the coverage error given by Equation 1.5 regardless how large MSE is going to be, when the condition in 1.2.1 is fulfilled, namely $b = o(1)$, we have the following result.

If $m^{-1/6} \succ n^{-1/9}$, the optimal choice should be close to

$$Error_{coverage} = O(m^{-1/3}), MSE = O(m^{-1/6}), h = m^{-1/6}, g = m^{-1/3}$$

If $m^{-1/6} \prec n^{-1/9}$, it should be roughly,

$$Error_{coverage} = O(m^{-1/4}n^{-1/18}), MSE = O(m^{1/2}n^{-7/9}), h = n^{-1/9}, g = m^{-1/2}n^{1/9}$$

mybibs.bib