Hybrid Estimator

Thesis Subtitle

Author Name

A thesis presented for the degree of
Doctor of Philosophy

[width=0.4]university

Department Name
University Name
Country

Date

1 Labelled Data

Definition of NW-estimator is generally given by the following equation 1 and it can be divided into two parts, namely 2 and 3.

$$NW = \frac{\hat{\alpha(x)}}{\hat{p(x)}} \tag{1}$$

$$\hat{\alpha(x)} = \frac{1}{nh_n}\Sigma y_i \cdot K(\frac{x - x_i}{h_n}) \tag{2}$$

$$\hat{p(x)} = \frac{1}{nh_n}\Sigma K(\frac{x - x_i}{h_n}) \tag{3}$$

1.1 Asymptotic normality of $\hat{\alpha(x)}$ and $\hat{p(x)}$

Theorem 1.1. *Central Limit Theorem for Triangular Arrays: It statest that, if the scalar random variable z_{in} is independently (but not necessarily identically) distributed with variance $Var(z_{in}) \equiv \sigma_{in}^2$ and r-th absolute central moment $E[|z_{in} - E(z_{in})|^r] \equiv \rho_{in} < \infty$ for some r ¿ 2; and if*

$$\frac{(\Sigma_{i=1}^n \rho_{in})^{1/r}}{(\Sigma_{i=1}^n \sigma_{in}^2)^{1/2}} \to 0$$

then standardized $\{z_{in}\}$ will converge to Normal distribution with mean 0 and variance 1.

Based on this theorem, we are able to deduce the condition under which $\hat{\alpha(x)}$ and $\hat{(x)}$ will be asymptotically normal. The proof for the later one, which is just kernel density estimator, is provided by **James L. Powell** in his **Notes On Nonparametric Density Estimation**.

Lemma 1.2. $\hat{\alpha(x)} = \frac{1}{nh_n}\Sigma y_i \cdot K(\frac{x-x_i}{h_n})$*is asymptotically normal if* $nn_h \to \infty$

Proof. Here $z_{in} = \frac{1}{h_n} y_i K(\frac{xi-x}{h_n})$, $yi = m(x_i) + \epsilon_i$. We already know the result that

$$\sigma_{in}^2 = Var(z_{in}) = nVar(\hat{\alpha(x)}) = O(h^{-1})$$

For all r larger than or equal to 3, we have the following

$$\begin{aligned}
\rho_{in} &= E[|z_{in} - E(z_{in})|^r] \\
&\leq 2^r E(|z_{in}|^r) \\
&= 8E[\frac{1}{h_n^3}|m(x_i) + \epsilon_i|K(\frac{xi - x}{h_n})^3] \\
&\leq \frac{8}{h_n^3}\sqrt{E[|m(x_i) + \epsilon_i|^6]E[K(\frac{xi - x}{h_n})^6]}
\end{aligned}$$

$$
\begin{aligned}
\rho_{in} &= E[|z_{in} - E(z_{in})|^r] \\
&\leq E[(|z_{in}| + |E(z_{in})|)^r] \\
&= \Sigma C_k^r E(|z_{in}|^k |E(z_{in})|^{r-k}) \\
&\leq \Sigma C_k^r E(|z_{in}|^r)^{\frac{k}{r}} E(|z_{in}|^r)^{\frac{r-k}{r}} \\
&= 2^r E(|z_{in}|^r)
\end{aligned}
$$

$$
\begin{aligned}
\rho_{in}^{\frac{1}{r}} &\leq 2E(|\frac{1}{nh_n} y_i K(\frac{x-x_i}{h_n})|^r)^{\frac{1}{r}} \\
&= 2\frac{1}{h_n} E(|(m(y)+\epsilon)K(\frac{x-y}{h_n})|^r)^{\frac{1}{r}} \\
&= 2\frac{1}{h_n} E(\Sigma C_q^r |m(y)^q \epsilon^{r-q} K(\frac{x-y}{h_n})^r|)^{\frac{1}{r}}
\end{aligned}
$$

$$
\begin{aligned}
&E(|m(y)^q \epsilon^{r-q} K(\frac{x-y}{h_n})^r|) \\
&\leq E(\epsilon^{r-q}) E(|m(y)^q \epsilon^{r-q} K(\frac{x-y}{h_n})^r|) \\
&= C_\epsilon[\int_{C1} m(y)^q \epsilon^{r-q} K(\frac{x-y}{h_n})^r p(y)dy - \int_{C2} m(y)^q \epsilon^{r-q} K(\frac{x-y}{h_n})^r p(y)dy] \\
&= O(h_n)
\end{aligned}
$$

Therefore, $\rho_{in} \leq O(h_n^{-r+1})$ and if $nh_n \to \infty, then \frac{(\Sigma_{i=1}^n \rho_{in})^{1/r}}{(\Sigma_{i=1}^n \sigma_{i\,n}^2)^2} \to 0, since$

$$
\begin{aligned}
&\frac{(\Sigma_{i=1}^n \rho_{in})^{1/r}}{(\Sigma_{i=1}^n \sigma_{i\,n}^2)^2} \\
&\leq \frac{C_1 n^{1/r} h_n^{-1+1/r}}{C_2 n^{1/2} h_n^{-1/2}} \\
&= Cn^{-1/2+1/r} h_n^{-1/2+1/r} \\
&= C(nh_n)^{-1/2+1/r} \to 0
\end{aligned}
$$

□

1.2 Asymptotic Distribution of NW-Estimator

After proving normality of and , we are able to discuss the distribution of .

$$
A = (nh_n)^{1/2}[\alpha(\hat{x}) - E[\alpha(\hat{x})]] \tag{4}
$$

$$
B = (nh_n)^{1/2}[p(\hat{x}) - E[p(\hat{x})]] \tag{5}
$$

Thus we can write NW in the following way,

$$\begin{aligned}
m\hat{(x)} - m(x) &= \frac{\alpha\hat{(x)}}{p\hat{(x)}} - m(x) \\
&= \frac{(nh_n)^{-1/2})A + E[\alpha\hat{(x)}]}{(nh_n^{-1/2})B + E[p\hat{(x)}]} - m(x) \\
&= \frac{(nh_n^{-1/2})(A - Bm(x)) + E[\alpha\hat{(x)} - m(x)p\hat{(x)}]}{(nh_n^{-1/2})B + E[p\hat{(x)}]}
\end{aligned}$$

when $nh_n^2 \to \infty, A \xrightarrow{d} N(0, \sigma_a^2), B \xrightarrow{d} N(0, \sigma_b^2)$, Therefore $(nh_n)^{-1/2}Aand(nh_n)^{-1/2}B \xrightarrow{p} 0$.

$$\nabla G(E(\alpha\hat{(x)}), E(p\hat{(x)}) = \begin{bmatrix} \frac{1}{E(p\hat{(x)})} \\ -\frac{E(\alpha\hat{(x)})}{E(p\hat{(x)})^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{p(x)} + O(h_n^2) \\ -\frac{m(x)}{p(x)} + O(h_n^2) \end{bmatrix}$$

$$\begin{aligned}
m\hat{(x)} - m(x) &= G((nh_n)^{-1/2}A + E(\alpha\hat{(x)}), (nh_n)^{-1/2}B + E(p\hat{(x)})) - m(x) \\
&= G(E(\alpha\hat{(x)}), E(p\hat{(x)}) - m(x) + \begin{bmatrix} (nh_n)^{-1/2}A \\ (nh_n)^{-1/2}B \end{bmatrix}^T \cdot \nabla G(E(\alpha\hat{(x)}), E(p\hat{(x)}) + O_p((nh_n)^{-1}) \\
&= \frac{E(\alpha\hat{(x)} - m(x)p\hat{(x)})}{E(p\hat{(x)})} + \begin{bmatrix} (nh_n)^{-1/2}A \\ (nh_n)^{-1/2}B \end{bmatrix}^T \cdot \begin{bmatrix} \frac{1}{p(x)} + O(h_n^2) \\ -\frac{m(x)}{p(x)} + O(h_n^2) \end{bmatrix} + O_p((nh_n)^{-1}) \\
&= E + (nh_n)^{-1/2}FA + (nh_n)^{-1/2}GB + (nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} + O_p((nh_n)^{-1})
\end{aligned}$$

Where $E = [h_n^2 m'(x)p'(x)\mu_2(k) + \frac{h_n^2}{2}m''(x)p(x)\mu_2(k)]\frac{1}{p(x)} + O(h_n^4)$, $F = \frac{h_n^2}{2}p''(x)\mu_2(k) + O(h_n^4)$ and $G = [\frac{3h_n^2}{2}m(x)p''(x)\mu_2(k) + h^2 m'(x)p'(x)\mu_2(k) + \frac{h_n^2}{2}m''(x)p(x)\mu_2(k)]\frac{1}{p(x)^2} + O(h_n^4)$. Here $\mu_2(k) = \int z^2 K(z)dz$

Combining all information together, $m\hat{(x)}$ will have an asymptotically normal distribution if $h_n^2 \cdot (nh_n)^{1/2}$ is bounded.

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

1.3 Joint Distribution of $\hat{\alpha}(x)$ and $\hat{p}(x)$

Intuitively, we know that they have a bivariate normal distribution. To prove this, we only need to show that every linear combination of these two is normal.

$$X = x_1\hat{\alpha}(x) + x_2\hat{p}(x) = \frac{1}{nh_n}\Sigma(x_1 y_i + x_2)K(\frac{x - x_i}{h_n}) = \frac{1}{nh_n}\Sigma(t(xi) + \eta_i)K(\frac{x - x_i}{h_n})$$

where $t(x) = x_1 m(x) + x_2$ and $\eta_i = x_1 \epsilon_i$. Utilizing our previous formulas, it is easy to obtain the following.

$$\begin{aligned}E(X) &= p(x)t(x) + \frac{h_n^2}{2}t(x)p''(x)\mu_2(k) + h^2 t'(x)p'(x)\mu_2(k)\\&+ \frac{h_n^2}{2}t''(x)p(x)\mu_2(k) + O(h_n^4)\end{aligned}$$

$$Var(X) = n(nh_n)^{-2}\int t(y)^2 K(\frac{y-x}{h_n})^2 p(y)dy + n(nh_n)^{-2}\sigma_\eta^2 \int K(\frac{y-x}{h_n})^2 p(y)dy - n^{-1}E^2(X)$$

Then the result follows according to **Lemma 2.2**.

Moreover, we can easily calculate $Cov(\hat{\alpha}(x), \hat{p}(x))$ as below,

$$\begin{aligned}Cov(\hat{\alpha}(x), \hat{p}(x)) &= \frac{1}{2}[Var(X) - Var(\hat{\alpha}(x)) - Var(\hat{p}(x))]\\&= \frac{1}{2}[n(nh_n)^{-2}\int[(m(y)+1)^2 - m(y)^2 - 1]K(\frac{y-x}{h_n})^2 p(y)dy\\&+ n(nh_n)^{-2}(\sigma_\epsilon^2 - \sigma_\epsilon^2 - 0)\int K(\frac{y-x}{h_n})^2 p(y)dy\\&- n^{-1}[E^2(\hat{\alpha}(x) + \hat{p}(x)) - E^2(\hat{\alpha}(x)) - E^2(\hat{p}(x))]]\\&= n(nh_n)^{-2}\int m(y)K(\frac{y-x}{h_n})^2 p(y)dy - n^{-1}E(\hat{\alpha}(x))E(\hat{p}(x))\\&= (nh_n)^{-1}[m(x)p(x)r(k) + \frac{1}{2}m(x)p''(x)\sigma_k^2 h_n^2 + m'(x)p'(x)\sigma_k^2 h_n^2\\&+ \frac{1}{2}m''(x)p(x)\sigma_k^2 h_n^2 + O(h_n^4)] - n^{-1}E(\hat{\alpha}(x))E(\hat{p}(x))\end{aligned}$$

Here $r(k) = \int K^2(z)dz$ and $\sigma_k^2 = \int K^2(z)z^2 dz$
and

$$\begin{aligned}Cov(A, B) &= (nh_n)Cov(\hat{\alpha}(x), \hat{p}(x))\\&= [m(x)p(x)r(k) + \frac{1}{2}m(x)p''(x)\sigma_k^2 h_n^2 + m'(x)p'(x)\sigma_k^2 h_n^2\\&+ \frac{1}{2}m''(x)p(x)\sigma_k^2 h_n^2 + O(h_n^4)] - h_n E(\hat{\alpha}(x))E(\hat{p}(x))\end{aligned}$$

$$
\begin{aligned}
Var(A - m(x)B) &= Var(A) + Var(m(x)B) - 2Cov(A, m(x)B) \\
&= \{m^2(x)p(x)r(k) + [2m(x)m^{'}(x)p'(x) + m''(x)m(x)p(x) + m'(x)^2 p(x) \\
&+ m^2(x)\frac{p''(x)}{2}]\sigma_k^2 h_n^2 + O(h_n^4)\} + \sigma_\epsilon^2\{p(x)r(k) + m^2(x)\frac{p''(x)}{2}\sigma_k^2 h_n^2 + O(h_n^4)\} - h_n E^2(\alpha(\hat{x})) \\
&+ m(x)^2\{p(x)r(k) + \frac{p''(x)}{2}\sigma_k^2 h_n^2 + O(h_n^4) - h_n E^2((\hat{x}))\} \\
&- 2m(x)\{[m(x)p(x)r(k) + \frac{1}{2}m(x)p''(x)\sigma_k^2 h_n^2 + m'(x)p'(x)\sigma_k^2 h_n^2 \\
&+ \frac{1}{2}m''(x)p(x)\sigma_k^2 h_n^2 + O(h_n^4)] - h_n E(\hat{\alpha}(x))E(\hat{p}(x))\} \\
&= \sigma_\epsilon^2 p(x)r(k) + O(h_n^2)
\end{aligned}
$$

2 Unlabelled Data

Here is the equation for estimator using unlabeled data

$$NW_{Unlabled} = \hat{r}(x) = \frac{\hat{\beta}(x)}{\hat{q}(x)} \tag{6}$$

$$\hat{\beta}(x) = \frac{1}{mg_m}\Sigma w_i K(\frac{x - x_i}{g_m}) \tag{7}$$

$$\hat{q}(x) = \frac{1}{mg_m}\Sigma K(\frac{x - x_i}{g_m}) \tag{8}$$

$$w_i = \hat{m}(x_i) \tag{9}$$

2.1 Asymptotic normality of $\hat{\beta}(x)$ and $\hat{q}(x)$ conditioned on X and Y

First of all, $\hat{q}(x)$ itself is actually not dependent on labeled dataset. Therefore, similar the previous derivation for $\hat{p}(x)$, it is asymptotically normal and independent of $\hat{\alpha}(x)$ and $\hat{p}(x)$.

Regarding (x), we shall first examine its conditional distribution on $\hat{\alpha}(x)$ and $\hat{p}(x)$, then figure out their joint distribution.

$$E(\hat{\beta}(x)|X,Y) = q(x)\hat{m}(x) + \frac{g_m^2}{2}\hat{m}(x)q''(x)\mu_2(k) + g_m^2\hat{m}'(x)q'(x)\mu_2(k) + \frac{g_m^2}{2}\hat{m}''(x)q(x)\mu_2(k) + O(g_m^4)$$

$$Var(\hat{\beta}(x)|X,Y) = \frac{1}{mg_m}[\hat{m}^2(x)q(x)\mu_2(k) + [2\hat{m}(x)\hat{m}'(x)q'(x) + \hat{m}''(x)\hat{m}(x)q(x) + \hat{m}'(x)^2q(x) + \hat{m}^2(x)\frac{q''(x)}{2}]\sigma_k^2 g_m^2] + O(g_m^4)] - \frac{1}{m}E^2(\hat{\beta(x)}|X,Y)$$
$$= O((mg_m)^{-1})$$

Since we have the result below from Miss.Tang's **8.4(1)**,

$$\hat{m}(x) = m(x) + O_p(h_n^2 + \frac{1}{\sqrt{nh_n}})$$

$$\hat{m}'(x) = m'(x) + O_p(h_n^2 + \frac{1}{\sqrt{nh_n^3}})$$

$$\hat{m}''(x) = m''(x) + O_p(h_n^2 + \frac{1}{\sqrt{nh_n^5}})$$

,

conditional expectation and variance can be then written as,

$$E(\hat{\beta}(x)|X,Y) = q(x)\hat{m}(x) + \frac{g_m^2}{2}m(x)q''(x)\mu_2(k) + g_m^2 m'(x)q'(x)\mu_2(k)$$

$$+ \frac{g_m^2}{2}m''(x)q(x)\mu_2(k) + O_p(g_m^2 h_n^2 + \frac{g_m^2}{\sqrt{nh_n^5}} + g_m^4)$$

Quoting the result for the asymptotically normal $\hat{\alpha}(x)$, we shall easily see that a transformation of $\hat{\beta}(x)|X,Y$ is also asymptotically normal.

2.2 Asymptotic normality of $\hat{\beta}(x)$ conditioned on $\hat{m}(x)$

First, introduce a new variable with conditionally asymptotically normal distribution,

$$C = (mg_m)^{1/2}[\hat{\beta}(x) - E[\hat{\beta}(x)|X,Y]]$$

Based on the conditional distribution of $\hat{\beta}(x)|X,Y$, we are able to derive $F(\hat{\beta}(x)|X,Y)$,

$$F(C|X,Y) = P(C \le c|X,Y) = \Phi(c/\sigma_c)$$

$$\sigma_c^2 = (mg_m)^{1/2}Var(\hat{\beta}(x)|X,Y)$$

$$= \hat{m}^2(x)q(x)\mu_2(k) + [2\hat{m}(x)\hat{m}'(x)q'(x) + \hat{m}''(x)\hat{m}(x)q(x)$$

$$+ \hat{m}'(x)^2 q(x) + \hat{m}^2(x)\frac{q''(x)}{2}]\sigma_k^2 g_m^2 + O(g_m^4) - g_m E^2(\hat{\beta(x)}|X,Y)$$

$$= m^2(x)q(x)\mu_2(k) + O_p(h_n^2 + \frac{1}{\sqrt{nh_n}}) + [2m(x)m'(x)q'(x) + m''(x)m(x)q(x)$$

$$+ m'(x)^2 q(x) + m^2(x)\frac{q''(x)}{2}]\sigma_k^2 g_m^2 + O(g_m^4) - g_m(q(x)m(x) + O_p(h_n^2 + \frac{1}{\sqrt{nh_n}}))$$

$$= m^2(x)q(x)\mu_2(k) + O_p(h_n^2 + \frac{1}{\sqrt{nh_n}} + g_m + g_m^2)$$

So the asymptotic distribution of C is actually independent of X,Y and $\hat{m}(x)$.

2.3 Asymptotic normality of $\hat{\beta}(x)$

$$\hat{\beta}(x) = (mg_m)^{-1/2}C + E[\hat{\beta}(x)|X,Y]$$

Moreover, $E(\hat{\beta}(x)|X,Y)$ can be expressed in $\hat{m}(x)$, which leads to their correlation and joint distribution,

$$E(\hat{\beta}(x)|X,Y) = q(x)\hat{m}(x) + \frac{g_m^2}{2}m(x)q''(x)\mu_2(k) + g_m^2 m'(x)q'(x)\mu_2(k)$$

$$+ \frac{g_m^2}{2}m''(x)q(x)\mu_2(k) + O_p(g_m^2 h_n^2 + \frac{g_m^2}{\sqrt{nh_n^5}} + g_m^4)$$

Even though $\hat{m}(x)$ in the formula above, we can still see that the whole term will depend on only constant terms when n is large enough.

2.4 Asymptotic normality of $\hat{r}(x)$ conditioned on $\hat{m}(x)$

Given $\hat{m}(x)$, we can derive the conditional normality of $\hat{r}(x)$ using the same method as that of $\hat{m}(x)$,

**Note that this calculation contains some mistakes.

$$\begin{aligned}
\hat{r}(x) &= \frac{(mg_m)^{-1/2}C + E[\hat{\beta}(x)|X,Y]}{(mg_m)^{-1/2}D + E[\hat{q}(x)]} \\
&= \frac{(mg_m)^{-1/2}C + q(x)\hat{m}(x)}{(mg_m)^{-1/2}D + q(x)} \\
&= G(q(x)\hat{m}(x), q(x)) - m(x) + \begin{bmatrix}(mg_m)^{-1/2}C + F \\ (mg_m)^{-1/2}D + G\end{bmatrix}^T \cdot \nabla G(q(x)\hat{m}(x), q(x)) + O_p((mg_m)^{-1}) \\
&= \hat{m}(x) + \begin{bmatrix}(mg_m)^{-1/2}C + F \\ (mg_m)^{-1/2}D + G\end{bmatrix}^T \cdot \begin{bmatrix}\frac{1}{q(x)} \\ -\frac{\hat{m}(x)}{q(x)}\end{bmatrix} + O_p((mg_m)^{-1}) \\
&= \hat{m}(x) + (mg_m)^{-1/2}[\frac{C}{q(x)} - \frac{1}{q(x)}(D + \frac{(mg_m^5)^{1/2}}{2}q''(x)\mu_2(k))\hat{m}(x) + \frac{(mg_m^5)^{1/2}}{2}\frac{1}{q(x)}q''(x)\mu_2(k)m(x)] \\
&= \hat{m}(x) + (mg_m)^{-1/2}[\frac{C}{q(x)} - \frac{D}{q(x)}\hat{m}(x) - \frac{(mg_m^5)^{1/2}}{2}\frac{q''(x)}{q(x)}\mu_2(k)(\hat{m}(x) - m(x))] \\
&= \hat{m}(x) + (mg_m)^{-1/2}[\frac{C}{q(x)} - \frac{D}{q(x)}m(x) - (nh_n)^{-1/2}\frac{D}{q(x)}E - \frac{(mg_m^5)^{1/2}(nh_n)^{-1/2}}{2}\frac{q''(x)}{q(x)}\mu_2(k)E] \\
&= \hat{m}(x) + (mg_m)^{-1/2}\hat{s}(x)
\end{aligned}$$

An more accurate calculation is:

$$
\begin{aligned}
\hat{r}(x) - \hat{m}(x) &= G((mg_m)^{-1/2}C + E(\widehat{\beta(x)}|X,Y), (mg_m)^{-1/2}D + E(\widehat{q(x)})) - \hat{m}(x) \\
&= G(E(\widehat{\beta(x)}|X,Y), E(\widehat{q(x)}) - \hat{m}(x) + \begin{bmatrix}(mg_m)^{-1/2}C \\ (mg_m)^{-1/2}D\end{bmatrix}^T \cdot \nabla G(E(\widehat{\beta(x)}|X,Y), E(\widehat{q(x)}) + O_p((mg_m)^{-} \\
&= \frac{E(\widehat{\beta(x)} - \hat{m}(x)\widehat{q(x)}|X,Y)}{E(\widehat{q(x)})} + \begin{bmatrix}(mg_m)^{-1/2}C \\ (mg_m)^{-1/2}D\end{bmatrix}^T \cdot \begin{bmatrix}\frac{1}{q(x)} + O(g_m^2) \\ -\frac{\hat{m}(x)}{q(x)} + O(g_m^2)\end{bmatrix} + O_p((mg_m)^{-1}) \\
&= H + (mg_m)^{-1/2}[IC + JD + \frac{C - \hat{m}(x)D}{q(x)}] + O_p((mg_m)^{-1}) \\
&= (mg_m)^{-1/2}\hat{s}(x)
\end{aligned}
$$

The second order expansion is $\begin{bmatrix}(mg_m)^{-1/2}C \\ (mg_m)^{-1/2}D\end{bmatrix}^T \cdot \nabla^2 G(E(\widehat{\beta(x)}|X,Y), E(\widehat{q(x)}) \begin{bmatrix}(mg_m)^{-1/2}C \\ (mg_m)^{-1/2}D\end{bmatrix}$ which is approximately $(mg_m)^{-1}[2D\frac{C-\hat{m}(x)D}{q(x)^2} + O(g_m^2)]$. So it is of order $O_p((mg_m)^{-1}g_m)$

Where $H = [g_m^2 m'(x)q'(x)\mu_2(k) + \frac{g_m^2}{2}m''(x)q(x)\mu_2(k)]\frac{1}{q(x)} + O_p(\frac{g_m^2}{\sqrt{nh_n^5}} + g_m^4)$, $I = \frac{g_m^2}{2}q''(x)\mu_2(k) + O(g_m^4)$ and $J = [\frac{3g_m^2}{2}m(x)q''(x)\mu_2(k) + g_m^2 m'(x)q'(x)\mu_2(k) + \frac{g_m^2}{2}m''(x)q(x)\mu_2(k)]\frac{1}{q(x)^2} + O_p(h_n^2 g_m^2 + \frac{g_m^2}{\sqrt{nh_n^5}} + g_m^4)$

When $(mg_m^5)^{1/2}(nh_n)^{-1/2}$ goes to infinity, it becomes $-\frac{(mg_m^5)^{1/2}(nh_n)^{-1/2}}{2}\frac{q''(x)}{q(x)}\mu_2(k)E$.

To further understand the distribution of C and D, we shall write out their joint distribution.

As a matter of fact, $F(C,D|X,Y)$ is a joint normal distribution. Since c is asymptotically independent of (X,Y) and D is independent of Y, $F(C,D|X,Y) = F(C,D)$ asymptotically. Therefore, (C,D) follows a normal distribution asymptotically.

when n and m is large enough, we have $Cov(C,D) = Cov(C,D|X,Y)$ which is in the exactly same form as $Cov(A,B)$.

$$
\begin{aligned}
Cov(C,D|X,Y) = [&\hat{m}(x)q(x)r(k) + \frac{1}{2}\hat{m}(x)q''(x)\sigma_k^2 h_n^2 + \hat{m}'(x)q'(x)\sigma_k^2 g_m^2 \\
&+ \frac{1}{2}\hat{m}''(x)q(x)\sigma_k^2 g_m^2 + O(g_m^4)] - g_m E(\hat{\beta}(x))E(\hat{q}(x))
\end{aligned}
$$

Therefore, we can actually show the term $C - \hat{m}(x)D$ is of order g_m. Since its expectation is 0, we only need to consider its asymptotic variance, which is approximately its conditional variance.

$$
\begin{aligned}
Var(C-\hat{m}(x)D|X,Y) &= Var(C|X,Y)+Var(\hat{m}(x)D|X,Y)-2Cov(C,\hat{m}(x)D)\\
&= \{\hat{m}^2(x)q(x)r(k)+[2\hat{m}(x)\hat{m}^{'}(x)q'(x)+\hat{m}''(x)\hat{m}(x)q(x)+\hat{m}'(x)^2q(x)\\
&+\hat{m}^2(x)\frac{q''(x)}{2}]\sigma_k^2g_m^2+O(g_m^4)\}-g_mE^2(\beta\hat{(x)}|X,Y)\\
&+\hat{m}(x)^2\{q(x)r(k)+\frac{q''(x)}{2}\sigma_k^2g_m^2+O(g_m^4)-g_mE^2(\hat{(x)}|X,Y)\}\\
&-2\hat{m}(x)\{[\hat{m}(x)q(x)r(k)+\frac{1}{2}\hat{m}(x)q''(x)\sigma_k^2h_n^2+\hat{m}'(x)q'(x)\sigma_k^2g_m^2\\
&+\frac{1}{2}\hat{m}''(x)q(x)\sigma_k^2g_m^2+O(g_m^4)]-g_mE(\hat{\beta}(x))E(\hat{q}(x))\}\\
&= \hat{m}'(x)^2q(x)\sigma_k^2g_m^2+O(g_m^4)
\end{aligned}
$$

Hence, $\frac{C-\hat{m}(x)D}{g_m}$ is asymptotically normal.

3 Distribution of Hybrid Estimator $\hat{y_c}(x)$

$$\hat{y_c}(x) = \lambda\hat{m}(x) + (1-\lambda)\hat{r}(x) \tag{10}$$

We can express the estimators like this,

$$\hat{\alpha}(x) = (nh_n)^{-1/2}A + p(x)m(x) + O(h_n^2)$$

$$\hat{p}(x) = p(x) + (nh_n)^{-1/2}B + O(h_n^2)$$

$$\hat{\beta}(x) = (mg_m)^{-1/2}C + q(x)\hat{m}(x) + O(g_m^2) + O_p(g_m^2 h_n^2 + \frac{g_m^2}{\sqrt{nh_n^5}} + g_m^4)$$

$$\hat{q}(x) = q(x) + (mg_m)^{-1/2}D + O(g_m^2)$$

We already know that $\hat{\beta}(x)$ is normal conditioned on $\hat{m}(x)$, then the joint distribution of $\hat{\beta}(x)$ and $\hat{q}(x)$ must also be bivariate normal distribution conditioned on $\hat{\alpha}(x)$ and $\hat{p}(x)$. Therefore, we can conclude that the joint distribution of four r.m. is a multivariate normal(So is A, B, C and D).

$$\hat{t}(x) = (nh_n)^{1/2}h_n^2E + h_n^2FA + h_n^2GB + \frac{A - m(x)B}{p(x)} + O_p((nh_n)^{-1/2})$$

$$\hat{s}(x) = (mg_m)^{1/2}g_m^2H + g_m^2IC + g_m^2JD + g_m\frac{C - \hat{m}(x)D}{q(x)g_m} + O_p((mg_m)^{-1/2}g_m)$$

we choose $1-\lambda = -\frac{h_n^2E}{g_m^2H}$ such that bias of $\hat{y_c}(x)$ is minimized, then

$$
\begin{aligned}
\hat{y_c}(x) - m(x) &= \lambda \hat{m}(x) + (1-\lambda)\hat{r}(x) - m(x) \\
&= \hat{m}(x) - m(x) + (1-\lambda)(\hat{r}(x) - \hat{m}(x)) \\
&= (nh_n)^{-1/2}\hat{t}(x) + (1-\lambda)(mg_m)^{-1/2}\hat{s}(x) \\
&= h_n^2 E + (nh_n)^{-1/2}h_n^2 FA + (nh_n)^{-1/2}h_n^2 GB + (nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} + O_p((nh_n)^{-1}) \\
&+ (1-\lambda)[g_m^2 H + (mg_m)^{-1/2}g_m^2 IC + (mg_m)^{-1/2}g_m^2 JD + (mg_m)^{-1/2}g_m\frac{C - \hat{m}(x)D}{q(x)g_m} \\
&+ O_p((mg_m)^{-1}g_m)] \\
&= (nh_n)^{-1/2}h_n^2 FA + (nh_n)^{-1/2}h_n^2 GB + (nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} + O_p((nh_n)^{-1}) \\
&- h_n^2(mg_m)^{-1/2}\frac{IE}{H}C - h_n^2(mg_m)^{-1/2}\frac{JE}{H}D - \frac{h_n^2(mg_m)^{-1/2}}{g_m^2}g_m\frac{E}{H}\frac{C - \hat{m}(x)D}{q(x)g_m} \\
&- \frac{E}{H}O_p(\frac{h_n^2(mg_m)^{-1})}{g_m^2}]
\end{aligned}
$$

Where $E = [m'(x)p'(x)\mu_2(k) + \frac{1}{2}m''(x)p(x)\mu_2(k)]\frac{1}{p(x)} + O(h_n^2)$, $F = \frac{1}{2}p''(x)\mu_2(k) + O(h_n^2)$, $G = [\frac{3}{2}m(x)p''(x)\mu_2(k) + m'(x)p'(x)\mu_2(k) + \frac{1}{2}m''(x)p(x)\mu_2(k)]\frac{1}{p(x)^2} + O(h_n^2)$ and $H = [m'(x)q'(x)\mu_2(k) + \frac{1}{2}m''(x)q(x)\mu_2(k)]\frac{1}{q(x)} + O_p(\frac{1}{\sqrt{nh_n^5}} + g_m^2)$, $I = \frac{g_m^2}{2}q''(x)\mu_2(k) + O(g_m^4)$ and $J = [\frac{3}{2}m(x)q''(x)\mu_2(k) + m'(x)q'(x)\mu_2(k) + \frac{1}{2}m''(x)q(x)\mu_2(k)]\frac{1}{q(x)^2} + O_p(h_n^2 + \frac{1}{\sqrt{nh_n^5}} + g_m^2)$

Next we shall discuss the asymptotic properties of $\hat{y_c}(x) - m(x)$ under different conditions. We shall fist notice that the leading terms are $(nh_n)^{-1/2}\frac{A-m(x)B}{p(x)}$ and $\frac{h_n^2(mg_m)^{-1/2}}{g_m^2}g_m\frac{E}{H}\frac{C-\hat{m}(x)D}{q(x)g_m}$ which is asymptotically a multi-variable normal distribution.

Therefore we can find the distribution of $\hat{y_c}(x) - m(x)$ through calculating its variance, which is of order $max\{(nh_n)^{-1/2}, (mg_m)^{-1/2}\frac{h_n^2}{g_m^2}g_m\}$. Hence the order of variance is smallest when $mg_m^3 \propto mh_n^5$.When n and m are large enough, we can approximate the variance as

$$
\begin{aligned}
Var(\hat{y_c}(x) - m(x)) &= Var((nh_n)^{-1/2}\frac{A - m(x)B}{p(x)} - \frac{h_n^2(mg_m)^{-1/2}}{g_m^2}g_m\frac{E}{H}\frac{C - \hat{m}(x)D}{q(x)g_m}) \\
&\approx \frac{1}{nh_n}\frac{1}{p^2(x)}Var(A - m(x)B) + \frac{h_n^4(mg_m)^{-1}}{g_m^4}g_m^2\frac{E^2}{H^2 q(x)^2}Var(\frac{C - \hat{m}(x)D}{g_m}) \\
&= \frac{1}{nh_n}\frac{1}{p^2(x)}\sigma_\epsilon^2 p(x)r(k) + \frac{h_n^4}{mg_m^3}\frac{E^2}{H^2 q(x)^2}\hat{m}'(x)^2 q(x)\sigma_k^2
\end{aligned}
$$

Or simply calculate through $E(Y) = E(E(Y|X)), Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$