On the Variance-stabilizing Non-parametric Regression : The case of the Nadaraya-Watson (N-W) estimator.

Kiheiji NISHIDA
Ph.D. student, Doctoral Program in Social Systems & Management, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Ten-noh-dai, Tsukuba, 351-8973, Japan, knishida@sk.tsukuba.ac.jp
Yuichiro KANAZAWA
University of Tsukuba

1. Motivation. Suppose that we wish to estimate Engel curve, which expresses the relation between income (X) and food expenses (Y) of a household, by one of non-parametric models, the N-W estimator. Then, we have to deal with the data that has two characteristics: One is that sample of X has distribution. Another is that the variability of Y depends on X.

2. Introduction of the Nadaraya-Watson (N-W) estimator. Suppose that we have n-pairs of random variables (X_i, Y_i), $i = 1, \cdots, n$. We assume x_i, $i = 1, \cdots, n$, are the realizations of i.i.d. random variable X whose density function is denoted as $f_X(x)$. Let u_i, $i = 1, \cdots, n$, be the realizations of the disturbance random variable U_i. We assume that $U_i|X_i$ are i.i.d. given X_i and that U_i are independent with X_j, $j \neq i$. We further assume that the conditional moments of $U|X$ as

$$E_{U|X}[U|X = x] = 0, \quad E_{U|X}[U^2|X = x] = \sigma^2(x).$$

We assume that the response Y is influenced by the explanatory variable X as

$$Y_i = m(X_i) + U_i = E(Y_i|X_i) + U_i,$$

where $m(X_i) = E(Y_i|X_i) = \int y f_{Y|X}(y|x)dy = \int y \frac{f_{X,Y}(x,y)}{f_X(x)}dy$. By replacing $f_{X,Y}(x, y)$ and $f_X(x)$ with the corresponding kernel bivariate and univariate estimates with the multiplicative kernel $K_{X,Y}(x, y) = K_X(x)K_Y(y)$ on the numerator and the kernel $K_X(x)$ on the denominator with the bandwidth on X as h_x, we arrive at the N-W estimator at $X = x$ as,

$$\hat{m}_{h_x}(x) = \frac{\sum_{i=1}^{n} K_X\left(\frac{x-X_i}{h_x}\right) Y_i}{\sum_{i=1}^{n} K_X\left(\frac{x-X_i}{h_x}\right)}.$$

3. The N-W with Mean Squared Error(MSE)-optimized bandwidth. The bandwidth that minimizes MSE $E_{\mathbf{X},\mathbf{Y}}\left[m(x) - \hat{m}_{h_x}(x)\right]^2$ is obtained as,

$$h^{MSE}(x) = \left[\frac{\left[\int K_X^2(t)dt\right]\sigma^2(x)}{\left[\int t^2 K_X(t)dt\right]^2 \frac{\left(2m^{(1)}(x)f_X^{(1)}(x)+m^{(2)}(x)f_X(x)\right)^2}{f_X(x)}}\right]^{\frac{1}{5}} n^{-\frac{1}{5}}, \tag{1}$$

and the asymptotic MSE and the asymptotic variance respectively as,

$$V_{\mathbf{X},\mathbf{Y}}(\hat{m}_{h^{MSE}(x)}(x)) \approx \left[\frac{\left[\int K_X^2(t)dt\right]^{\frac{4}{5}}\left[\sigma^2(x)\right]^{\frac{4}{5}}}{\left[\int t^2 K_X(t)dt\right]^{-\frac{2}{5}} \frac{\left(2m^{(1)}(x)f_X^{(1)}(x)+m^{(2)}(x)f_X(x)\right)^{-\frac{2}{5}}}{f_X^{-\frac{6}{5}}(x)}}\right] n^{-\frac{4}{5}}. \tag{2}$$

The N-W with MSE bandwidth has two problems. First, the variance is not uniform across the domain of the regression because the term $\sigma^{\frac{8}{5}}(x)(2m^{(1)}(x)f_X^{(1)}(x)+m^{(2)}(x)f_X(x))^{-\frac{2}{5}}/f_X^{-\frac{6}{5}}(x)$ that appears in (2) depends on x. Second, the N-W might jump at the points where $2m^{(1)}(x)f_X^{(1)}(x)+m^{(2)}(x)f_X(x)=0$ is satisfied because MSE bandwidth $h^{MSE}(x)$ goes to infinity and the N-W goes to sample average $\bar{Y}=(1/n)\sum_{i=1}^{n}Y_i$ at the x.

4. The N-W with Variance-Stabilizing (V-S) bandwidth. To solve the two problems with the MSE-optimized bandwidth, we propose the following V-S bandwidth,

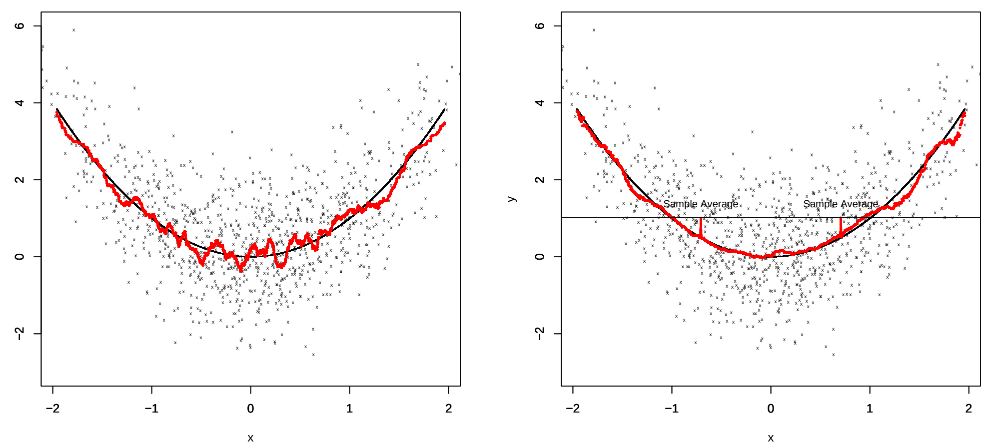
$$h^{V-S}(x)=\left[\frac{\left[\int K_X^2(t)dt\right]}{\left[\int t^2K_X(t)dt\right]^2\left[\int\frac{\sigma^8(x)\left(m^{(2)}(x)f_X(x)+2m^{(1)}(x)f_X^{(1)}(x)\right)^2}{f_X^5(x)}dx\right]}\right]^{\frac{1}{5}}n^{-\frac{1}{5}}\cdot\frac{\sigma^2(x)}{f_X(x)}. \quad (3)$$

and its stabilized variance is written as,

$$V_{\mathbf{X},\mathbf{Y}}[\hat{m}_{h_x^{V-S}(x)}(x)]=\left[\frac{\left[\int K_X^2(t)dt\right]^{\frac{4}{5}}}{\left[\int t^2K_X(t)dt\right]^{-\frac{2}{5}}\left[\int\frac{\sigma^8(x)\left(m^{(2)}(x)f_X(x)+2m^{(1)}(x)f_X^{(1)}(x)\right)^2}{f_X^5(x)}dx\right]^{-\frac{1}{5}}}\right]n^{-\frac{4}{5}}$$
$$+O\left(\frac{1}{n}\right)+o\left(\frac{1}{nh_x}\right).$$

The bandwidth (3) is obtained by extracting the term $\sigma^2(x)/f_X(x)$ from (1) and integrating the variable term except $\sigma^2(x)/f_X(x)$ with its weight being $f_X(x)$. Another interpretation of (3) is the minimizer of Mean Integrated Squared Error (MISE) $\int E_{\mathbf{X},\mathbf{Y}}\left[m(x)-\hat{m}_{h_x}(x)\right]^2 dF_X(x)$ with its bandwidth being constraind by $h_x=(\sigma^2(x)/f_X(x))h_0$.

5. Comparison of V-S with MSE bandwidth. We show a numerical example of the case where $m(x)=x^2I_{[-1.96,1.96]}(x)$, $U_i|X_i\sim N(0,\sigma^2(x))$, $\sigma^2(x)=1$, $X_i\sim N(0,1)$ and $n=1000$. The left panel below is the N-W with V-S bandwidth, the right with MSE-optimized bandwidth. Theoretically, variance stabilization is achieved.



The N-W with V-S bandwidth is less smooth than that with MSE-optimized. In terms of both theoretical and empirical MISE, the V-S bandwidth is inferior to MSE-optimized by about 200%. These are considered to be the penalty of variance stabilization. As for maximum error, the V-S bandwidth is superior to MSE-optimized.