Variance estimation after Kernel Ridge Regression Imputation

Hengfang Wang & Jae Kwang Kim

Center for Survey Statistics and Methodology
Iowa State University

07/17/2020
**Introduction**

**Data Structure**
- Imputation is a popular technique for handling item nonresponse.
- Suppose we have the dataset \( \{(\delta_i, x_i, \delta_i y_i), i = 1, \ldots, n\} \).
- \( x_i \in \mathbb{R}^d \): is fully observed covariate.
- \( y_i \in \mathbb{R} \): response subject to missingness.
- \( \delta_i \): response indicator for \( y_i \), \( \delta_i = 1 \) if \( y_i \) is observed, for \( i = 1, \ldots, n \).

**Target Estimator**
Under MAR (missing at random), we can develop a nonparametric estimator \( \hat{m}(\mathbf{x}) \) of \( m(\mathbf{x}) = \mathbb{E}(Y | \mathbf{x}) \) and construct the following imputation estimator:

\[
\hat{\theta}_I = \frac{1}{n} \sum_{i=1}^{n} \{ \delta_i y_i + (1 - \delta_i) \hat{m}(\mathbf{x}_i) \} .
\]

**KRR Imputation**
- We use kernel ridge regression (KRR) to get the corresponding \( \hat{m}(\cdot) \) where

\[
\hat{m} = \arg\min_{m \in \mathcal{H}} \left[ \sum_{i=1}^{n} \delta_i \{ y_i - m(\mathbf{x}_i) \}^2 + \lambda \|m\|^2_{\mathcal{H}} \right],
\]

where \( \lambda > 0 \) is tuning parameter and \( \mathcal{H} \) is a Hilbert space.
Main Results

**Theorem 1**

*Under regularity conditions, for a Sobolev kernel of order $\ell$, $\lambda = n^{1-\ell}$, we have*

$$\sqrt{n}(\hat{\theta}_I - \tilde{\theta}_I) = o_P(1),$$  \hspace{2cm} (3)

*where*

$$\tilde{\theta}_I = \frac{1}{n} \sum_{i=1}^{n} \left[ m(x_i) + \delta_i \frac{1}{\pi(x_i)} \{y_i - m(x_i)\} \right].$$  \hspace{2cm} (4)

*Furthermore, as $n \to \infty$,*

$$\sqrt{n}(\hat{\theta}_I - \theta) \overset{d}{\longrightarrow} N(0, \sigma^2),$$

*where*

$$\sigma^2 = V\{E(Y \mid \mathbf{x})\} + E\{V(Y \mid \mathbf{x})/\pi(\mathbf{x})\}$$

*and $\pi(\mathbf{x}) = E(\delta \mid \mathbf{x})$.***
Main Results

Methodology

By Theorem 1, we use the following estimator to estimate the variance of $\hat{\theta}_I$ in (1):

$$\hat{\sigma}^2(\hat{\theta}_I) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\eta}_i - \bar{\eta})^2$$  \hspace{1cm} (5)

where

$$\hat{\eta}_i = \hat{m}(x_i) + \delta_i \hat{\omega}_i \{y_i - \hat{m}(x_i)\},$$

and $\hat{\omega} = (\hat{\omega}_1, \ldots, \hat{\omega}_n)^T \in \mathbb{R}^n$ is estimated by

$$\hat{\omega} = \arg \min_{\omega \geq 1} \left[ \max_{u \in \mathcal{H}_n} \left\{ S_n(\omega, u) - \lambda \| u \|_{\mathcal{H}}^2 \right\} + \tau V_n(\omega) \right].$$  \hspace{1cm} (6)

where $S_n(\omega, u) = \left\{ \frac{1}{n} \sum_{i=1}^{n} (\delta_i \omega_i - 1) u(x_i) \right\}^2$, $V_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} \delta_i \omega_i^2$, $\mathcal{H}_n = \{ u \in \mathcal{H} : \| u \|_n = 1 \}$ and $\| u \|_n^2 = \frac{1}{n} \sum_{i=1}^{n} u(x_i)^2$. The weight $\hat{\omega}_i$ estimates $\{ \pi(x_i) \}^{-1}$ nonparametrically.