STAT 521: Homework 5

Problem 1:

Consider a finite population of size N with measurement $(\mathbf{x}, y)$.

We are interested in estimating $Y = \sum_{i=1}^{N} y_i$ using a linear estimator of the form $\hat{Y}_\omega = \sum_{i \in A} \omega_i y_i$. We wish to impose the following calibration constraints to the final weights:

$$\sum_{i \in A} \omega_i \mathbf{h}(\mathbf{x}_i) = \sum_{i=1}^{N} \mathbf{h}(\mathbf{x}_i) \tag{1}$$

where $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \dots, h_L(\mathbf{x})]$ is a known function of $\mathbf{x}$.

To uniquely determine ω_i, we consider minimizing

$$Q(\boldsymbol{\omega}) = \sum_{i \in A} \left(\omega_i - \pi_i^{-1}\right)^2 q_i$$

subject to (1) where $q_i = q(\mathbf{x}_i)$ is a known function of $\mathbf{x}_i$ and π_i is the first-order inclusion probability of unit i. Let $\hat{Y}_{\text{cal}} = \sum_{i \in A} \hat{\omega}_i y_i$ be the calibration estimator of Y using $\hat{\omega}_i$ obtained from the above calibration problem.

Answer the following questions.

1. Find the closed-form expression for $\hat{\omega}_i$.

2. Under the assumption of $\mathbf{h}_i' \mathbf{a} = q_i / \pi_i$ holds for some $\mathbf{a}$, show that

$$\sum_{i \in A} \frac{1}{\pi_i} \left(y_i - \mathbf{h}_i' \hat{\boldsymbol{\beta}}_h\right) = 0 \tag{2}$$

   and

$$\sum_{i=1}^{N} \left(y_i - \mathbf{h}_i' B_h\right) = 0 \tag{3}$$

   where $\mathbf{h}_i = \mathbf{h}\left(\mathbf{x}_i\right)$ and

$$\hat{\boldsymbol{\beta}}_h = \left\{ \sum_{i \in A} \mathbf{h}(\mathbf{x}_i) \mathbf{h}(\mathbf{x}_i)' / q_i \right\}^{-1} \sum_{i \in A} \mathbf{h}(\mathbf{x}_i) y_i / q_i$$

   and

$$B_h = \left\{ \sum_{i=1}^{N} \pi_i \mathbf{h}(\mathbf{x}_i) \mathbf{h}(\mathbf{x}_i)' / q_i \right\}^{-1} \sum_{i=1}^{N} \pi_i \mathbf{h}(\mathbf{x}_i) y_i / q_i.$$

3. Show this: If (1) holds, then $\hat{Y}_{\text{cal}} = \sum_{i \in A} \hat{\omega}_i y_i$ is equivalent to the projection estimator of the form

$$\hat{Y}_{\text{proj}} = \sum_{i=1}^{N} \mathbf{h}_i' \hat{\boldsymbol{\beta}}_h.$$

4. Show this: If (2) holds, then we have

$$\hat{Y}_{\text{cal}} = Y + \sum_{i \in A} \frac{1}{\pi_i}\eta_i + \left(\sum_{i=1}^{N} \mathbf{h}_i - \sum_{i \in A} \frac{1}{\pi_i}\mathbf{h}_i\right)' \left(\hat{\boldsymbol{\beta}} - B_h\right)$$

where

$$\eta_i = y_i - \mathbf{h}_i' B_h \tag{4}$$

and

$$B_h = \left\{\sum_{i=1}^{N} \pi_i \mathbf{h}(\mathbf{x}_i)\mathbf{h}(\mathbf{x}_i)'/q_i\right\}^{-1} \sum_{i=1}^{N} \pi_i \mathbf{h}(\mathbf{x}_i) y_i / q_i.$$

5. Now, suppose that we have a superpopulation model with $Y_i \mid \mathbf{x}_i \sim (m(\mathbf{x}_i), q(\mathbf{x}_i)\sigma^2)$. Show that $E(\eta_i \mid \mathbf{X}) = 0$ if $\mathbf{h}_i$ includes $m(\mathbf{x}_i)$ in the sense that $\mathbf{h}_i'\boldsymbol{\alpha} = m(\mathbf{x}_i)$ for some $\boldsymbol{\alpha}$. [Hint: Show that $E_\zeta(B_h) = \boldsymbol{\alpha}$.]

6. If the model is

$$y_i = m(x_i) + e_i$$

with $e_i \sim (0, q(\mathbf{x}_i)\sigma^2)$ and $\mathbf{h}_i'\boldsymbol{\alpha} = m(\mathbf{x}_i)$ for some $\boldsymbol{\alpha}$, then

$$AV\left(\sum_{i \in A} \frac{1}{\pi_i}\eta_i\right) \cong \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) q(\mathbf{x}_i)\sigma^2$$

where $AV\left(\sum_{i \in A} \pi_i^{-1}\eta_i\right)$ is the anticipated variance (=model expectation of the design variance) of $\sum_{i \in A} \pi_i^{-1}\eta_i$ where η_i is defined in (4). [Hint: Use

$$\begin{aligned}
AV\left(\sum_{i \in A} \frac{1}{\pi_i}\eta_i\right) &= E_\zeta\left(\sum_{i=1}^{N}\sum_{j=1}^{N} (\pi_{ij} - \pi_i\pi_j) \frac{\eta_i}{\pi_i}\frac{\eta_j}{\pi_j}\right) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} (\pi_{ij} - \pi_i\pi_j) \frac{E_\zeta(\eta_i)}{\pi_i}\frac{E_\zeta(\eta_j)}{\pi_j} \\
&+ \sum_{i=1}^{N}\sum_{j=1}^{N} (\pi_{ij} - \pi_i\pi_j) \frac{1}{\pi_i}\frac{1}{\pi_j} Cov_\zeta\left(\eta_i, \eta_j\right)
\end{aligned}$$

and check that the first term is equal to zero.]

Problem 2:

Consider a finite population of size N with measurement (x, y), where $x > 0$. We consider the following two-phase sampling.

1. Phase 1: select a sample A_1 of size n_1 by SRS and observe x_i for $i \in A_1$.
2. Phase 2: From A_1, select a Poisson sample A_2 of expected sample size $n_2 (< n_1)$ with the first-order inclusion probability $\pi_i = \pi(x_i) \in (0, 1)$.

We use the following two-phase ratio estimator of $\theta = N^{-1} \sum_{i=1}^{N} y_i$:

$$\hat{\theta}_{\text{tpr}} = \frac{1}{n_1} \sum_{i \in A_1} x_i \hat{\gamma}_2$$

where

$$\hat{\gamma}_2 = \frac{\sum_{i \in A_2} \pi_i^{-1} y_i}{\sum_{i \in A_2} \pi_i^{-1} x_i}.$$

1. Show that $\hat{\theta}_{\text{tpr}}$ is asymptotically design unbiased.
2. Derive the formula for linearization variance estimator of $\hat{\theta}_{\text{tpr}}$.
3. Find the formula for optimal inclusion probability $\pi(x)$ for the second-phase sampling (in the sense that it minimizes the asymptotic variance).