Background

We consider a survey population $U = \{1, \ldots, N\}$ with N known. On U, a real variable of interest y is defined. We suppose that a sample s is drawn from U w.p $p(s)$.

To estimate $Y = \sum_{i=1}^{N} Y_i$, one may employ a homogeneous linear unbiased estimator (HLE) $t_s = \sum_{i \in s} y_i b_{si} = \sum_{i=1}^{N} Y_i b_{si} I_{si}$

where b_{si}'s are constants free of $\underset{\sim}{Y} = (Y_1, \ldots, Y_N)$ and $I_{si} = 1$ if $i \in s$; and 0 otherwise. Then the MSE is

$$M(t_s) = E_p\left[t_s - Y\right]^2 = \sum_{i}^{N}\sum_{j} d_{ij} Y_i Y_j$$

where E_p denotes expectation operator w.r.t the sampling design p corresponding to $p(s)$ above and

$$d_{ij} = E_p\left[(b_{si} I_{si} - 1)(b_{sj} I_{sj} - 1)\right].$$

Rao (1979) proved the following result on MSE of a HLE of Y:

- Theorem 1. Let there exist constants $w_i (>0)$ independent of $\underset{\sim}{Y}$, such that
  "$\forall i \in U \; z_i = \frac{y_i}{w_i}$ is a ~~contant~~ constant"
  $\Longrightarrow M(t_s) = 0$.
  Then, $M(t_s) = -\sum\sum_{i<j}^{N} d_{ij} w_i w_j (z_i - z_j)^2$, uniformly in $\underset{\sim}{Y}$.

This theorem was proved in our class.

By considering fixed effective sample size design and $\pi_i > 0$ $\forall i$, one can show that the Horwitz Thompson estimator $\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$ satisfies Rao's condition (with $w_i = \pi_i$).

In this case, one has $d_{ij} = \dfrac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$

and from unbiasedness of $\hat{y}_{HT}$ and Rao's theorem, we get

$$Var(\hat{y}_{HT}) = \sum\sum_{i<j}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right)^2 \quad \text{(Yates and Grundy, 1953)}$$

We get an unbiased estimator of this variance (assuming $\forall i, j \; \pi_{ij} > 0$) as $\displaystyle\sum\sum_{i<j \in s} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2$

We wish to provide another unbiased estimator of $Var(\hat{y}_{HT})$ when Rao's condition is NOT satisfied.

- Theorem 2. Let there exist $w_i (\neq 0)$ independent of $\underset{\sim}{Y}$ for $i \in U$. If $z_i = \dfrac{y_i}{w_i}$ then

$$M(t_s) = - \sum\sum_{i<j}^{N} d_{ij} w_i w_j (z_i - z_j)^2 + \sum_{i=1}^{N} \frac{Y_i^2}{w_i} \cdot \alpha_i$$

where $\alpha_i = \displaystyle\sum_{j=1}^{N} d_{ij} w_j$.

proof:
$$\begin{aligned}
M(t_s) &= \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} Y_i Y_j \\
&= \sum\sum_{i \neq j} d_{ij} Y_i Y_j + \sum_{i=1}^{N} Y_i^2 \cdot d_{ii} \\
&= \sum\sum_{i \neq j} d_{ij} Y_i Y_j - \sum_{i=1}^{N} \frac{Y_i^2}{w_i} (\alpha_i - d_{ii} w_i) + \sum_{i=1}^{N} \frac{Y_i^2}{w_i} \cdot \alpha_i \\
&= \sum\sum_{i \neq j} d_{ij} Y_i Y_j - \sum_{i=1}^{N} \frac{Y_i^2}{w_i} \left[\sum_{j=1}^{N} d_{ij} w_j - d_{ii} w_i \right] + \sum_{i=1}^{N} \frac{Y_i^2}{w_i} \cdot \alpha_i \\
&= \sum\sum_{i \neq j} d_{ij} Y_i Y_j - \sum_{i=1}^{N} \frac{Y_i^2}{w_i} \left[\sum_{j \neq i} d_{ij} w_j \right] + \sum_{i=1}^{N} \frac{Y_i^2}{w_i} \cdot \alpha_i \\
&= -\frac{1}{2} \sum\sum_{i \neq j} d_{ij} w_i w_j \left[\frac{Y_i^2}{w_i^2} + \frac{Y_j^2}{w_j^2} - 2 \frac{Y_i Y_j}{w_i w_j} \right] + \sum_{i=1}^{N} \frac{Y_i^2}{w_i} \cdot \alpha_i \\
&= - \sum\sum_{i<j}^{N} d_{ij} w_i w_j (z_i - z_j)^2 + \sum_{i=1}^{N} \frac{Y_i^2}{w_i} \cdot \alpha_i \quad \blacksquare
\end{aligned}$$

• Corollary 1. If Rao's condition does not hold, then the expression of variance of $\hat{y}_{HT}$ is given by

$$Var^*(\hat{y}_{HT}) = \underset{\substack{i<j \\ 1}}{\sum\sum}^{N} (\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 + \sum_{i=1}^{N} \frac{\alpha_i Y_i^2}{\pi_i}$$

where $\alpha_i = 1 + \dfrac{\sum_{j\neq i} \pi_{ij}}{\pi_i} - \sum_{j=1}^{N} \pi_j$

proof: For $\hat{y}_{HT}$, $\alpha_i = \sum_{j=1}^{N} d_{ij} w_j$

$$= \sum_{j=1}^{N} \left(\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\right) \cdot \pi_j$$

$$= \frac{\sum_{j=1}^{N} \pi_{ij}}{\pi_i} - \sum_{j=1}^{N} \pi_j$$

$$= 1 + \frac{\sum_{j\neq i} \pi_{ij}}{\pi_i} - \sum_{j=1}^{N} \pi_j .$$

The rest follows from Theorem 2 and unbiasedness of $\hat{y}_{HT}$.

• Corollary 2. An unbiased estimator of $Var^*(\hat{y}_{HT})$ is given by

$$\hat{V}^*(\hat{y}_{HT}) = \sum\sum_{i<j\in s} \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 + \sum_{i\in s} \frac{\alpha_i y_i^2}{\pi_i^2}$$

Solution of the Problem

Let E_R denote and V_R denote the expectation and variance operator w.r.t the randomisation device described in the problem.

Let E_p and V_p denote the expectation and variance operator w.r.t the sampling design.

We have $E_R(z_i) = c.y_i + \sum_{j=1}^{M} q_j x_j$

$$\Rightarrow E_R(r_i) = y_i \quad \text{where } r_i = \frac{z_i - \sum_{j=1}^{M} q_j x_j}{c}$$

Now we write $\underset{\sim}{Y} = (Y_1, \ldots, Y_i, \ldots, Y_N)$ and $\underset{\sim}{R} = (r_1, \ldots, r_i, \ldots, r_N)$ generically. We consider the estimator

$$t = t(s, \underset{\sim}{R}) = \frac{1}{N} \sum_{i \in s} \frac{r_i}{\pi_i} \quad \text{where } \pi_i = \sum_{s \ni i} p(s)$$

(This is valid as it is given that $\pi_i > 0 \ \forall i$)

Now, $E(t) = E_p E_R(t) = E_p E_R\left[\frac{1}{N} \sum_{i \in s} \frac{r_i}{\pi_i}\right]$

$$= \frac{1}{N} E_p\left[\sum_{i \in s} \frac{E_R(r_i)}{\pi_i}\right]$$

$$= \frac{1}{N} E_p\left[\sum_{i \in s} \frac{y_i}{\pi_i}\right]$$

$$= \frac{1}{N}.Y \text{ (from the unbiasedness of HT estimator)}$$

$$= \overline{Y}.$$

Hence, $t = \frac{1}{N} \sum_{i \in s} \frac{r_i}{\pi_i}$ is an UE for $\overline{Y}$.

We note that $E_R(t) = \frac{1}{N} \hat{y}_{HT}$

and $E_p(t) = \overline{R}$

Now, $V(t) = E_R V_p(t) + V_R E_p(t)$

$$= E_R V_p\left[\frac{1}{N}\sum_{i\in s}\frac{\pi_i}{\pi_i}\right] + V_R(\bar{R})$$

$$= \frac{1}{N^2}\cdot E_R \cdot V_p\left[\sum_{i\in s}\frac{\pi_i}{\pi_i}\right] + \frac{1}{N^2}\cdot V_R\left(\sum_{i=1}^{N} R_i\right)$$

$$= \frac{1}{N^2}\cdot E_R\, V_{ar}^*(\hat{\pi}_{HT}) + \frac{1}{N^2}\cdot\sum_{i=1}^{N} V_R(\pi_i)$$

(as π_i's are independent of each other)

$$= \frac{1}{N^2}\, E_R\left[E_p\left\{\hat{V}^*(\hat{\pi}_{HT})\right\}\right] + \frac{1}{N^2}\cdot\sum_{i=1}^{N} V_R(\pi_i)$$

(from Corollary 2)

$$= \frac{1}{N^2}\, E_R E_p\left[\hat{V}^*(\hat{\pi}_{HT})\right] + \frac{1}{N^2}\sum_{i=1}^{N} E_R(v_i)$$

(where v_i is an unbiased estimator of $V_i = V_R(\pi_i)$)

$$= \frac{1}{N^2}\, E\left[\hat{V}^*(\hat{\pi}_{HT})\right] + \frac{1}{N^2}\sum_{i=1}^{N} E_R(v_i)$$

$$= \frac{1}{N^2}\, E\left[\hat{V}^*(\hat{\pi}_{HT})\right] + \frac{1}{N^2}\cdot E\left[\sum_{i\in s}\frac{v_i}{\pi_i}\right]$$

$$= E\left[\frac{1}{N^2}\left\{\hat{V}^*(\hat{\pi}_{HT}) + \sum_{i\in s}\frac{v_i}{\pi_i}\right\}\right]$$

So, ~~we have to~~ it is enough to find an unbiased estimator v_i of $V_i = V_R(\pi_i)$, in order to produce an unbiased estimator of $V(t)$.

$$\text{Now, } V_R(r_i) = \frac{1}{c^2} \cdot V_R(z_i)$$

$$= \frac{1}{c^2} \cdot \left[E_R(z_i^2) - E_R^2(z_i) \right]$$

$$= \frac{1}{c^2} \left[c \cdot y_i^2 + \sum_{j=1}^{M} q_j x_j^2 - \left(c \cdot y_i + \sum_{j=1}^{M} q_j x_j \right)^2 \right]$$

$$= \frac{1}{c^2} \left[c(1-c) y_i^2 - 2c \cdot y_i \cdot \sum_{j=1}^{M} q_j x_j + \sum_{j=1}^{M} q_j x_j^2 - \left(\sum_{j=1}^{M} q_j x_j \right)^2 \right]$$

$$= \alpha \cdot y_i^2 + \beta \cdot y_i + \gamma$$

where $\alpha = \frac{1-c}{c}$, $\beta = -\frac{2}{c} \cdot \sum_{j=1}^{M} q_j x_j$

$$\gamma = \frac{1}{c^2} \left[\sum_{j=1}^{M} q_j x_j^2 - \left(\sum_{j=1}^{M} q_j x_j \right)^2 \right].$$

So, $V_i = V_R(r_i) = \alpha \cdot y_i^2 + \beta \cdot y_i + \gamma$.

Let $v_i = \dfrac{\alpha \cdot r_i^2 + \beta \cdot r_i + \gamma}{1+\alpha}$

$$\text{Then } E_R(v_i) = \frac{\alpha \cdot E_R(r_i^2) + \beta \cdot E_R(r_i) + \gamma}{1+\alpha}$$

$$= \frac{\alpha \cdot \{ V_i + y_i^2 \} + \beta \cdot y_i + \gamma}{1+\alpha}$$

$$= \frac{\alpha \cdot V_i + \alpha \cdot y_i^2 + \beta \cdot y_i + \gamma}{1+\alpha}$$

$$= \frac{\alpha \cdot V_i + V_i}{1+\alpha} = V_i \, .$$

Hence, an unbiased estimator of $V(t)$ is given by

$$\frac{1}{N^2} \left[\sum\sum_{i<j \in s} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) \left(\frac{r_i}{\pi_i} - \frac{r_j}{\pi_j} \right)^2 + \sum_{i \in s} \frac{\alpha_i r_i^2}{\pi_i^2} + \sum_{i \in s} \frac{v_i}{\pi_i} \right]$$

Note that, the above estimator covers both the cases

(i) $\nu(s)$ = fixed for all samples $s \in \mathcal{S}$

$\downarrow$

distinct units in s

(ii) $\nu(s)$ not fixed.

When $\nu(s)$ = fixed for all samples $s \in \mathcal{S}$ $= n$ (let)

then we have $\sum_{j=1}^{N} \pi_j = n$ and $\sum_{j \neq i}^{N} \pi_{ij} = (n-1)\pi_i$

Hence, in this case $\alpha_i = 1 + \dfrac{\sum_{j\neq i} \pi_{ij}}{\pi_i} - \sum_{j=1}^{N} \pi_j$

$$= 1 + \frac{(n-1)\pi_i}{\pi_i} - n$$

$$= 1 + (n-1) - n = 0$$

and the expression of the above variance estimator becomes less complicated.