Module 4: Unequal Probability Sampling

Probability Proportional to Size Sampling (PPS Sampling)

$\mathcal{U} = (U_1, \ldots, U_N)$ $\quad$ $y = (Y_1, \ldots Y_N)$ $\quad$ $Y_i \leftrightarrow U_i$

$\mathcal{U}$: Population

y: Study Variable

$x = (x_1, \ldots, x_N)$: Auxiliary Variables

x_i: Auxiliary variable associated with Unit U_i, measuring size (or weight) of Unit U_i.

Example (1) $\mathcal{U}$: Population of N agricultural holdings

y: Agricultural production in N agricultural holdings

x: Areas of N agricultural holdings

(2) $\mathcal{U}$: Population of N business companies

y: Revenue of N business companies

x: Size of N business companies (size in terms of their presence on the market)

A sample of n units is selected with probability of U_i being included in the sample proportional to x_i. We assume that x_is are natural numbers.

Probability Proportional to Size Sampling With Replacement (PPSWR Sampling)

Cumulative Total Method for generating PPSWR Sample

Units	Sizes/weights	Cumulative Total	Associated Random Numbers
U_1	x_1	x_1	$1 \ldots x_1$
U_2	x_2	$x_1 + x_2$	$x_1+1 \ldots x_1+x_2$
$\vdots$			
U_i	x_i	$x_1 + x_2 \cdots + x_i$	$\sum_{j=1}^{i-1} x_j + 1 \ldots \sum_{j=1}^{i} x_j$
$\vdots$			
U_N	x_N	$\sum_{i=1}^{N} x_i = X$ (say)	$\sum_{j=1}^{N-1} x_j + 1, \ldots \sum_{j=1}^{N} x_j = X$

1) Draw a random number R from $\{1 \ldots X\}$. If

$$\sum_{j=1}^{i-1} X_j + 1 \leq R \leq \sum_{j=1}^{i} X_j$$

Select U_i, $i = 1, \ldots, N$.

2) Repeat above step n times independently to get a PPSWR sample of size n.

Clearly

$P_i = P(U_i \text{ is selected in any specific draw}) = \frac{X_i}{X}$, $i = 1 \ldots N$,

Where $X = \sum_{j=1}^{N} X_j$

Lahiri's Method for generating PPSWR Sample.

Let

$$X_0 = \text{Max}\{X_1, \ldots, X_N\}.$$

1) Draw a random number, say R_1, from $\{1, \ldots, N\}$ and select unit U_{R_1} provisionally

Draw another random number R_2 from $\{1, 2, \ldots, X_0\}$.

If $R_2 \leq X_{R_1}$, select U_{R_1} finally

If $R_2 > X_{R_1}$, no unit is selected

The above procedure is repeated until a unit is selected → A draw consisting of many steps

n times independently

2) Repeat (1) for selecting a sample of size n

Here each draw consists of a number of steps as described above

Theorem Under Lahiri's method

$P_i = Pr(U_i \text{ is selected in any specific draw})$

$= \frac{X_i}{X}, \quad i = 1 \dots N.$

Proof

$a = Pr(\text{Unit } U_i \text{ is selected in a given step})$

$= P(R_1 = i, R_2 \le X_i)$

$= P(R_1 = i)\, P(R_2 \le X_i)$

$= \frac{1}{N} \cdot \frac{X_i}{X_0}$

$Pr(\text{no unit is selected in a given step})$

$b = \sum_{i=1}^{N} P(R_1 = i, R_2 > X_i)$

$= \sum_{i=1}^{N} P(R_1 = i)\, P(R_2 > X_i)$

$= \sum_{i=1}^{N} \frac{1}{N}\left(1 - \frac{X_i}{X_0}\right)$

$= 1 - \frac{1}{N}\frac{X}{X_0}$

$P_i = P(U_i \text{ is selected in any specific draw})$

$= a + ba + b^2 a + \dots$

$= \frac{a}{1-b} = \frac{\frac{1}{N}\cdot\frac{X_i}{X_0}}{\frac{X}{NX_0}} = \frac{X_i}{X}, \quad i = 1 \dots N$

Unequal Probability Sampling.

UPS $<$ WR / WOR

PSWR Sampling Design

$\underline{P} = (P_1, \dots, P_N) \quad P_i \ge 0 \;\; i = 1(1)N \;\; \sum_{i=1}^{N} P_i = 1$

$P_i = Pr(\text{drawing } U_i \text{ in any specific draw}), \; i = 1 \dots N$

$P(\underline{\Delta}) = Pr(\underline{S} = \underline{\Delta}) = \begin{cases} P_{i_1} P_{i_2} \cdots P_{i_n}, & \text{if } \underline{\Delta} = (U_{i_1}, \dots, U_{i_n}) \\ & i_1 \dots i_n \in \{1 \dots N\} \\ 0, & \text{o.w.} \end{cases}$

Inclusion Probabilities

Define

$\pi_i = Pr(U_i \in S)$, $\pi_i^c = 1 - \pi_i = Pr(U_i \notin S)$, $i = 1, \ldots, N$

$\pi_{ij} = Pr(U_i \in S, U_j \in S)$, $\pi_{ij}^c = Pr(U_i \notin S, U_j \notin S)$, $i, j \in \{1, \ldots, N\}$, $i \neq j$

Lemma (i) $\pi_i^c = (1 - p_i)^n$, $\pi_i = 1 - (1 - p_i)^n$, $i = 1, \ldots, N$

(ii) $\pi_{ij}^c = (1 - p_i - p_j)^n$, $i, j \in \{1, \ldots N\}$, $i \neq j$

(iii) $\pi_{ij} = 1 - (1 - p_i)^n - (1 - p_j)^n + (1 - p_i - p_j)^n$, $i, j = 1, \ldots, N$, $i \neq j$

Estimation of Population Total and Population Mean

$X = \sum_{i=1}^{N} x_i$, $x_0 = \max_{1 \le i \le N} x_i$ (for PSSWR(n))

$p_i = Pr$(drawing U_i in a specific draw), $i = 1, \ldots, N$.

$p_i = \frac{x_i}{X}$, $i = 1, \ldots, N$, under PPSWR(n) $\sum_{i=1}^{N} p_i = 1$, $p_i > 0$, $i = 1, \ldots, N$

Define

$$z_i = \frac{y_i}{p_i}, \quad i = 1, \ldots N$$

$$P(z) = \begin{pmatrix} z_1 & z_2 \ldots & z_N \\ p_1 & p_1 \ldots & p_N \end{pmatrix}$$

$$Y = \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} z_i p_i, \quad i = 1, \ldots N, \quad \bar{Y} = \frac{Y}{N} = \frac{1}{N} \sum_{i=1}^{N} z_i p_i$$

Let $(U_1^*, \ldots, U_n^*)$ be a PPSWR(n) sample from $U = (U_1 \ldots U_N)$ and let $(T_1, \ldots, T_n)$ be corresponding z_i's. Then $T_1, T_2, \ldots, T_n$ are iid rvs with $Pr(T_1 = z_i) = p_i$, $i = 1, \ldots N$

$$E(T_1) = \sum_{i=1}^{N} z_i p_i = Y; \quad V(T_1) = E((T_1 - Y)^2) = \sum_{i=1}^{N} (z_i - Y)^2 p_i.$$

Thus an unbiased estimator of Y is

$$\hat{Y}_{UPSWR} \quad \frac{1}{n} \sum_{i=1}^{n} T_i = \bar{T} = \frac{1}{n} \sum_{i \in S} \frac{y_i}{p_i}$$

$$\hat{Y}_{PPSWR} = \frac{X}{n} \sum_{i \in S} \frac{y_i}{x_i}$$

$$Var(\hat{Y}_{PPSWR}) = \frac{Var(T_1)}{n}$$

$$= \frac{1}{n} \sum_{i=1}^{N} (z_i - Y)^2 p_i$$

$$Var(\hat{Y}_{UPSWR}) = \frac{1}{n} \sum_{i=1}^{N} \left(\frac{y_i}{p_i} - Y\right)^2 p_i$$

$$Var(\hat{Y}_{PPSWR}) = \frac{X}{n} \sum_{i=1}^{N} \left(\frac{y_i}{x_i} - \frac{Y}{X}\right)^2 x_i$$

$$Var(\hat{Y}_{PPSWR}) = \frac{X}{n} \sum_{i=1}^{N} \left(\frac{y_i}{x_i} - \frac{Y}{X}\right)^2 x_i$$

$$V(\hat{Y}_{UPSWR}) = \frac{1}{n} \sum_{i=1}^{N} \left(\frac{y_i}{p_i} - Y\right)^2 p_i$$

$$= \frac{1}{n}\left[\sum_{i=1}^{N} \frac{y_i^2}{p_i} - Y^2\right]$$

Unbiased Estimation of $Var(\hat{Y}_{PPS})$

Note that $\hat{Y}_{PPS} = \bar{T}$

$$E(T_1) = Y$$

$$E(T_1^2) = \sum_{i=1}^{N} z_i^2 p_i = \sum_{i=1}^{N} \frac{y_i^2}{p_i}$$

$$\Rightarrow E\left(\frac{1}{n}\sum_{i=1}^{n} T_i^2\right) = \sum_{i=1}^{N} \frac{y_i^2}{p_i}$$

Let δ be an unbiased estimator of $Var(\hat{Y}_{PPS}) = \frac{Var(T_1)}{n}$. Then we may take

$$\delta = \frac{1}{n} \cdot \frac{1}{n-1} \sum_{i=1}^{n} (T_i - \bar{T})^2$$

$$= \frac{1}{n(n-1)} \sum_{i \in S} \left(\frac{y_i}{p_i} - \hat{Y}_{PPS}\right)^2$$

$$\hat{Var}(\hat{\bar{Y}}_{PPSWR}) = \frac{1}{n(n-1)} \sum_{i \in S} \left(\frac{y_i}{p_i} - \hat{\bar{Y}}_{PPS} \right)^2$$

Under PSWR(n),

Theorem. Suppose that $Cov\left(\frac{y_i^2}{x_i}, x_i\right) > 0$ (i.e. $\sum_{i=1}^{N} \frac{y_i^2}{x_i}(x_i - \bar{X}) > 0$). Then

$$V(\hat{\bar{Y}}_{PPSWR}) < V(\hat{\bar{Y}}_{SRSWR});$$

here $\bar{X} = \frac{X}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i$.

Proof. We have

$$V(\hat{\bar{Y}}_{PPSWR}) = \frac{1}{n} \left[\sum_{i=1}^{N} \frac{y_i^2}{p_i} - \bar{Y}^2 \right]$$

$$= \frac{1}{n} \left[X \sum_{i=1}^{N} \frac{y_i^2}{x_i} - \bar{Y}^2 \right]$$

$$V(\hat{\bar{Y}}_{SRSWR}) = N^2 \frac{\sigma^2}{n}$$

$$= \frac{1}{n} \left[N \sum_{i=1}^{N} y_i^2 - \bar{Y}^2 \right]$$

Thus

$$V(\hat{\bar{Y}}_{PPSWR}) < V(\hat{\bar{Y}}_{SRSWR})$$

$$\Leftrightarrow \quad X \sum_{i=1}^{N} \frac{y_i^2}{x_i} < N \sum_{i=1}^{N} y_i^2$$

$$\Leftrightarrow \quad \sum_{i=1}^{N} \left(y_i^2 - \frac{X}{N} \frac{y_i^2}{x_i} \right) > 0$$

$$\Leftrightarrow \quad \sum_{i=1}^{N} y_i^2 \left(1 - \frac{\bar{X}}{x_i} \right) > 0$$

$$\Leftrightarrow \quad \sum_{i=1}^{N} \frac{y_i^2}{x_i} (x_i - \bar{X}) > 0$$

$$\Leftrightarrow \quad Cov\left(\frac{y_i^2}{x_i}, x_i \right) > 0.$$

Gain due to UPSWR against SRSWR

We have

$$\hat{V}(\hat{Y}_{UPSWR}) = \frac{1}{n(n-1)}\left[\sum_{i\in s}\frac{y_i^2}{p_i^2} - n\hat{Y}^2_{PPS}\right] = 8$$

$$V(\hat{Y}_{SRSWR}) = \frac{1}{n}\left[N\sum_{i=1}^{N} y_i^2 - Y^2\right]$$

$$= \frac{N}{n^2}E\left(\sum_{i\in s}\frac{y_i^2}{p_i}\right) - \frac{1}{n}Y^2 \quad \left(E\left(\frac{1}{n}\sum_{i\in s}\frac{y_i^2}{p_i}\right) = \sum_{i=1}^{N} y_i^2\right)$$

$$= \frac{N}{n^2}E\left(\sum_{i\in s}\frac{y_i^2}{p_i}\right) - \frac{1}{n}\left[E\left(\hat{Y}^2_{PPS}\right) + V(\hat{Y}_{PPS})\right]$$

$$= \frac{N}{n^2}E\left(\sum_{i\in s}\frac{y_i^2}{p_i}\right) - \frac{1}{n}\left[E(\hat{Y}^2_{PPS}) - E(\hat{V}(\hat{Y}_{PPS}\right]$$

$$= E\left[\frac{1}{n^2}\left\{N\sum_{i\in s}\frac{y_i^2}{p_i} - n\hat{Y}^2_{PPS}\right\} + \frac{1}{n}\hat{V}(\hat{Y}_{PPS}\right]$$

$$= E\left[\hat{V}(\hat{Y}_{SRSWR})\right]$$

Estimated gain under PPSWR is

$$\frac{N}{n^2}\sum_{i\in s}\frac{y_i^2}{p_i} - \frac{1}{n}\hat{Y}^2_{PPS} + \frac{1}{n}\hat{V}(\hat{Y}_{UPSWR}) - \hat{V}(\hat{Y}_{UPSWR})$$

$$= \frac{N}{n^2}\sum_{i\in s}\frac{y_i^2}{p_i} - \frac{1}{n}\hat{Y}^2_{UPSWR}\frac{n-1}{n}\left[\frac{1}{n(n-1)}\left\{\sum_{i\in s}\frac{y_i^2}{p_i^2} - n\hat{Y}^2_{UPSWR}\right\}\right]$$

$$= \frac{1}{n^2}\left[N\sum_{i\in s}\frac{y_i^2}{p_i} - \sum_{i\in s}\frac{y_i^2}{p_i^2}\right]$$

$$= \frac{1}{n^2}\left[\sum_{i\in s}\left(N - \frac{1}{p_i}\right)\frac{y_i^2}{p_i}\right]$$

$$= \frac{1}{n^2}\left[\sum_{i\in s}\left(Np_i - 1\right)\frac{y_i^2}{p_i}\right] = \hat{G}_{UPSWR/SRSWR}$$

$\hat{G}_{UPSWR/SRSWR}$ → Gain of UPSWR over SRSWR

Unequal Probability Sampling Without Replacement (UPSWOR)

$\mathcal{U} = (U_1, \ldots, U_N)$, $\underline{p} = (p_1, \ldots, p_N)'$: given vector

$p_i \geq 0$ $i = 1(1)N$, $\sum_{i=1}^{N} p_i = 1$.

Sampling Scheme

(1) Draw one unit from $\mathcal{U}$ s.t. probability of selecting unit U_i is p_i, $i = 1 \ldots N$.
Let U_{i_1} be the unit selected. Do not replace it back.

(2) Draw one unit from remaining $(N-1)$ units such that probability of selecting unit U_i $(i \neq i_1)$ $\propto p_i$.

i.e. $p_{i|i_1} = P(\text{selecting unit } U_i \mid \text{unit } U_{i_1} \text{ was drawn in first draw})$

$$= \frac{p_i}{1 - p_{i_1}}$$

Let the selected unit be U_{i_2}.

$\vdots$

(3) Let $U_{i_1}, \ldots, U_{i_{n-1}}$ be the units selected in the first $(n-1)$ draws. At the nth draw select one unit from the remaining $(N-n+1)$ units such that probability of selecting unit U_i $(i \neq i_1 \neq i_2 \neq \cdots \neq i_{n-1})$ is $\propto p_i$, i.e.

$p_{i|i_1 \ldots i_{n-1}} = P(\text{selecting } U_i \mid U_{i_1}, \ldots U_{i_{n-1}} \text{ are drawn in first } n-1 \text{ draws})$

$$= \frac{p_i}{1 - \sum_{j=1}^{n-1} p_{i_j}}$$

Let the selected unit be U_{i_n}

Take $\underline{s} = (U_{i_1}, \ldots, U_{i_n})$

We have

$\{p_i\}_{i=1}^N$: Selection probability set of 1st draw

$\{p_{j|i}\}_{j(\neq i)=1}^N$: Conditional probability set of 2nd draw given that U_i is selected at first draw.

$$p_{j|i} = \frac{p_j}{1-p_i}, \quad j=1,\dots,N, \quad j \neq i$$

$\{p_{i_n|i_1\dots i_{n-1}}\}_{i_n(\neq i_1 \neq \dots \neq i_{n-1})=1}^N$: Conditional probability set of nth draw given that $n-1$ units $U_{i_1}, \dots, U_{i_{n-1}}$ are selected in first $(n-1)$ draws.

$$p_{i_n|i_1\dots i_{n-1}} = \frac{p_{i_n}}{1-\sum_{j=1}^{n-1} p_{i_j}}$$

Probability Proportional to Size Without Replacement (PPSWOR) Sampling

M : a known auxiliary variable (size measure)

$x_i = M(U_i)$, $i=1,\dots,N$, $\underline{x} = (x_1, \dots, x_N)$: known size measures

$x_i \geq 0$, $i=1,\dots,N$, $p_i = \frac{x_i}{X}$, where $X = \sum_{i=1}^N x_i$.

$\underline{p} = (p_1, \dots, p_N)$: vector of normed size measures.

$p_{j|i} \propto x_j$, $j=1,\dots,N$, $j \neq i$ $\Rightarrow$ $p_{j|i} = \frac{x_j}{X - x_i} = \frac{p_j}{1-p_i}$, $j=1,\dots,N$, $j \neq i$.

$\propto x_{i_n}$, $i_n = 1,\dots,N$, $i_n \neq i_1 \neq \dots \neq i_{n-1}$ $\Rightarrow$ $p_{i_n|i_1\dots i_{n-1}} = \frac{x_{i_n}}{X - \sum_{j=1}^{n-1} x_{i_j}}$

$$= \frac{p_{i_n}}{1-\sum_{j=1}^{n-1} p_{i_j}},$$

$i_n = 1,\dots,N$, $i_1 \neq i_2 \neq \dots \neq i_{n-1}$

Method of Drawing Sample

Cumulative total method or Lahiri's method can be applied with obvious modifications.

Alternative Sampling Methods

(1) Draw units one by one as per the above probability scheme.

(2) Draw units one by one with replacement, such that the probability of selecting U_i at any draw is p_i, until n distinct units are obtained. If an unit appears more than once, keep it in the position where it first appeared and discard it from all other positions.

Sampling Design

$$P(\underset{\sim}{S} = (U_{i_1}, \ldots, U_{i_n})) = \begin{cases} p_{i_1} \dfrac{p_{i_2}}{1-p_{i_1}} \cdots \dfrac{p_{i_n}}{1-p_{i_1}-p_{i_2}-\cdots-p_{i_{n-1}}}, \\ 1 \le i_1 \ne i_2 \ne \cdots \ne i_n \le N \\ 0 \quad \text{o.w.} \end{cases}$$

$\underset{\sim}{S} = (S_1, \ldots, S_n)$: UPSWOR(n)

$P_r(S_1 = U_i) = p_i, \; i = 1, 2, \ldots, N$, $P_r(S_2 = U_j \mid S_1 = U_i) = \dfrac{p_j}{1-p_i}, \; i \ne j$, $P_r(S_3 = U_k \mid S_1 = U_i, S_2 = U_j) = \dfrac{p_k}{1-p_i-p_j}, \; k \ne i \ne j, \ldots$

Inclusion Probabilities

Special Case $n = 2$

$$\pi_i = P_r(U_i \in \underset{\sim}{S}) = p_i + \sum_{\substack{j=1 \\ j \ne i}}^{N} p_j \frac{p_i}{1-p_j}$$

$$\boxed{\pi_i = p_i \left[1 + \sum_{\substack{j=1 \\ j \ne i}}^{N} \frac{p_j}{1-p_j} \right], \quad i = 1, 2, \ldots, N}$$

$$\pi_{ij} = P((U_i, U_j)) + P((U_j, U_i))$$

$$= p_i \frac{p_j}{1-p_i} + p_j \frac{p_i}{1-p_j}$$

$$\boxed{\pi_{ij} = \frac{p_i p_j (2-p_i-p_j)}{(1-p_i)(1-p_j)}, \quad i, j = 1, \ldots, N, \; i \ne j}$$

General n

$$\begin{pmatrix} U_1 & \cdots & U_N \\ p_1 & \cdots & p_N \end{pmatrix}$$

$p_i = \frac{x_i}{X}$, $i = 1, \ldots N$ Under PP

First Draw

$$\begin{pmatrix} U_1 & \cdots & U_N \\ p_1 & \cdots & p_N \end{pmatrix} \rightarrow U_i$$

2nd Draw

$$\begin{pmatrix} U_1 & \cdots & U_{i-1} & U_{i+1} & \cdots & U_N \\ p^{(1)}_{1|i} & \cdots & p^{(1)}_{i-1|i} & p^{(1)}_{i+1|i} & \cdots & p^{(1)}_{N|i} \end{pmatrix} \rightarrow U_j$$

$\binom{N}{1}$ Acts of Conditional probabilities

$$p^{(1)}_{j|i} = \frac{p_j}{1-p_i}, \quad j = 1, \ldots N, \quad j \neq i$$

Under PPS

$$p^{(1)}_{j|i} = \frac{x_j}{X - x_i}, \quad j = 1, \ldots N, \quad j \neq i$$

3rd Draw

$$\begin{pmatrix} U_1 & \cdots & U_{i-1} & U_{i+1} & \cdots & U_{j-1} & U_{j+1} & \cdots & U_N \\ p^{(2)}_{1|i,j} & \cdots & p^{(2)}_{i-1|i,j} & p^{(2)}_{i+1|i,j} & \cdots & p^{(2)}_{j-1|i,j} & p^{(2)}_{j+1|i,j} & \cdots & p^{(2)}_{N|i,j} \end{pmatrix}$$

$$p^{(2)}_{k|i,j} = \frac{p_k}{1-p_i-p_j}, \quad k = 1, \ldots N, \quad k \neq i, j$$

Under PPS

$$p^{(2)}_{k|i,j} = \frac{x_k}{X - x_i - x_j}, \quad k = 1, \ldots N, \quad k \neq i, j$$

$\rightarrow \binom{N}{2}$ Acts of Conditional probabilities

$$P(S = (U_{i_1}, \ldots, U_{i_n})) = p_{i_1} \times \frac{p_{i_2}}{1-p_{i_1}} \times \frac{p_{i_3}}{1-p_{i_1}-p_{i_2}} \times \cdots \times \frac{p_{i_n}}{1-p_{i_1}-\cdots-p_{i_{n-1}}},$$

$U_{i_j} \in \{U_1, \ldots U_N\}$, $U_{i_1} \neq U_{i_2} \neq \cdots \neq U_{i_n}$

For $\binom{N}{n} \cdot n = N(N-1)\cdots(N-n+1)$ samples we can compute these probabilities

$$\pi_i = Pr(U_i \in S) = \sum_{A : U_i \in A} Pr(S = A)$$

$$\pi_{ij} = Pr(U_i \in S, U_j \in S) = \sum_{A : U_i, U_j \in A} Pr(S = A)$$

Estimation of Population Total Y

Define

$$A_i = \begin{cases} 1, & \text{if } U_i \in S \\ 0, & \text{O.W.} \end{cases}, \quad i = 1, \ldots, N.$$

$E(A_i) = \pi_i$
$V(A_i) = \pi_i(1-\pi_i)$
$i = 1, 2, \ldots, N$

Let

$$L = \sum_{i \in S} c_i Y_i, \quad \text{for some constants } c_1, \ldots, c_N$$

$$= \sum_{i=1}^{N} A_i c_i Y_i$$

We would choose c_i's so that

$$E(L) = Y = \sum_{i=1}^{N} Y_i$$

$$\Rightarrow \sum_{i=1}^{N} E(A_i) c_i Y_i = \sum_{i=1}^{N} Y_i$$

$$\Rightarrow \sum_{i=1}^{N} \pi_i c_i Y_i = \sum_{i=1}^{N} Y_i$$

One may take $c_i = \frac{1}{\pi_i}$, $i = 1, \ldots, N$

$$\hat{Y}_{UPSWOR} = \sum_{i \in S} \frac{Y_i}{\pi_i}$$

→ Horvitz-Thompson Estimator

is an unbiased estimator of Y.

$$Var(\hat{Y}_{UPSWOR}) = V\left(\sum_{i=1}^{N} \frac{A_i Y_i}{\pi_i} \right)$$

$$= \sum_{i=1}^{N} V(A_i) \frac{Y_i^2}{\pi_i^2} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} Cov(A_i, A_j) \frac{Y_i Y_j}{\pi_i \pi_j}$$

$$= \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i}(1-\pi_i) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \frac{Y_i Y_j}{\pi_i \pi_j} (\pi_{ij} - \pi_i \pi_j)$$

$$Var(\hat{Y}_{UPSWOR}) = \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i}(1-\pi_i) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \frac{Y_i Y_j}{\pi_i \pi_j} (\pi_{ij} - \pi_i \pi_j)$$

$$Var(\hat{Y}_{PPSWOR}) = \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2}(1-\pi_i) E(A_i) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \frac{Y_i Y_j}{\pi_i \pi_j} \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_{ij}} E(A_i A_j)$$

$$= E\left[\sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2}(1-\pi_i)A_i + \sum_{i=1}^{N}\sum_{\substack{j=1 \\ i\neq j}}^{N} \frac{Y_i Y_j}{\pi_i \pi_j} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}} A_i A_j \right]$$

$$= E\left[\sum_{i\in S} \frac{Y_i^2}{\pi_i^2}(1-\pi_i) + \sum_{\substack{i\in S \\ i\neq j}} \sum_{j\in S} \frac{Y_i Y_j}{\pi_i \pi_j} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}} \right]$$

Thus

$$\hat{V}(\hat{Y}_{UPSWOR}) = \sum_{i\in S} \frac{Y_i^2}{\pi_i^2}(1-\pi_i) + \sum_{\substack{i,j\in S \\ i\neq j}} \sum \frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}$$

is an unbiased estimator of $V(\hat{Y}_{PPSWOR})$

Alternate Expression for $V(\hat{Y}_{PPSWOR})$

$$V(\hat{Y}_{UPSWOR}) = \sum\sum_{1\le i<j\le N} (\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2$$

Alternate Unbiased Estimator of $V(\hat{Y}_{PPSWOR})$

$$\hat{V}_{YG}(\hat{Y}_{UPSWOR}) = \sum_{i\in S}\sum_{\substack{j\in S \\ i<j}} \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}$$

→ Yates-Grundy Estimator

In general, Not same as $\hat{V}(\hat{Y}_{PPSWOR})$. For SRSWOR($n$) ($\pi_i = \frac{n}{N}$, $i=1,\dots N$) the two estimators are the same)

Des Raj Estimator of Y

First Consider $n=2$. Let $\underline{s} = (s_1, s_2)$ be the selected sample.

Define

$$(v_1, v_2) = \begin{cases} \left(\frac{Y_i}{P_i}, Y_i + \frac{1-P_i}{P_j}Y_j\right), & \text{if } (s_1, s_2) = (u_i, u_j) \\ & i,j = 1,\dots N \quad i\neq j \\ 0, & \text{o.w} \end{cases}$$

Let

$$D = \frac{V_1 + V_2}{2}$$

$$\hat{\bar{Y}}_{DS} = D = \frac{1}{2}\left[\frac{1+P_i}{P_i} Y_i + \frac{1-P_i}{P_j} Y_j \right], \quad \text{if } (S_1, S_2) = (U_i, U_j), \quad i,j = 1 \cdots N, \ i \neq j$$

Theorem (i) $E(V_i) = \bar{Y}$ $\quad i = 1, 2$

(ii) $E(D) = \bar{Y}$.

(iv) $Cov(V_1, V_2) = 0$

(iii) $\hat{V}(D) = D^2 - V_1 V_2$ is an unbiased estimator of $V(D)$.

$\downarrow$

$\hat{V}(\hat{\bar{Y}}_{DS}) =$

Proof (i) $E(V_1) = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ (\neq i)}}^{N} \frac{Y_i}{P_i} P(S_1 = U_i, S_2 = U_j)$

$$= \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \frac{Y_i}{P_i} P_i \frac{P_j}{1-P_i} = \sum_{i=1}^{N} Y_i = \bar{Y}$$

Alt

$$E(V_1) = \sum_{i=1}^{N} E(V_1 | S_1 = U_i) P_i = \sum_{i=1}^{N} \frac{Y_i}{P_i} P_i = \bar{Y}$$

$$E(V_2) = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \left(Y_i + \frac{1-P_i}{P_j} Y_j \right) P_i \frac{P_j}{1-P_i}$$

$$= \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \left(\frac{Y_i P_i P_j}{1-P_i} + Y_j P_j \right)$$

$$= \sum_{i=1}^{N} \left[Y_i P_i + P_i \sum_{j \neq i} Y_j \right]$$

$$= \sum_{i=1}^{N} Y_i P_i + \sum_{i=1}^{N} P_i (\bar{Y} - Y_i)$$

$$= \bar{Y} \qquad \left(\sum_{i=1}^{N} P_i = 1 \right)$$

Alt

$$E(V_2 | V_1 = \frac{Y_i}{P_i}) = \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(Y_i + \frac{1-P_i}{P_j} Y_j \right) P(S_2 = U_j | S_1 = U_i)$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(Y_i + \frac{1-P_i}{P_j} Y_j \right) \frac{P_j}{1-P_i}$$

$$= Y_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} Y_j = \bar{Y}$$

$$E(V_2 | V_1) = \bar{Y}$$

$$\Rightarrow E(V_2) = E(E(V_2 | V_1)) = \bar{Y}$$

(ii) $E(V_1 V_2) = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ (\neq i)}}^{N} \frac{Y_i}{P_i} \left(Y_i + \frac{1-P_i}{P_j} Y_j \right) P_i \frac{P_j}{1-P_i}$

$$= \sum_{i=1}^{N} \left[Y_i^2 + \sum_{j \neq i} Y_j \right] = \bar{Y}^2$$

Alt

$$E(V_1 V_2) = E(E(V_1 V_2 | V_1))$$

$$= \bar{Y} E(V_1) = \bar{Y}^2$$

$Cov(V_1, V_2) = 0$

(iii) $E(D^2 - V_1 V_2)$

$= E(D^2) - \bar{Y}^2$

Also $E(D) = \frac{1}{2}[E(V_1) + E(V_2)] = \bar{Y}$

$E(D^2 - V_1 V_2) = V(D)$

Expression for $V(D)$

$D = \frac{V_1 + V_2}{2}$

$V(\hat{\bar{Y}}_{PS}) = V(D) = \frac{1}{4}[V(V_1) + V(V_2)]$ $(Cov(V_1, V_2) = 0)$

Result 1 (a) $V(V_1) = \sum_{i=1}^{N} P_i \left(\frac{Y_i}{P_i} - \bar{Y}\right)^2 = \sum_{i=1}^{N}\sum_{j>i}^{N} \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2 P_i P_j$

(b) $V(V_2) = \sum_{i=1}^{N}\sum_{j>i}^{N} \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2 (1 - P_i - P_j) P_i P_j$

Proof $V(V_1) = E\left((V_1 - \bar{Y})^2\right)$

$= \sum_{i=1}^{N}\sum_{j\neq i}^{N} \left(\frac{Y_i}{P_i} - 1\right)^2 P_i \frac{P_j}{1 - P_i}$

$= \sum_{i=1}^{N} \left(\frac{Y_i}{P_i} - \bar{Y}\right)^2 P_i$

$= \sum_{i=1}^{N} \frac{Y_i^2}{P_i} - \bar{Y}^2$

$= \sum_{i=1}^{N} \frac{Y_i^2}{P_i} - \left(\sum_{i=1}^{N} Y_i^2 + \sum\sum_{i\neq j} Y_i Y_j\right)$

$= \sum_{i=1}^{N} \frac{Y_i^2}{P_i}(1 - P_i) - \sum_{i\neq j}^{N}\sum^{N} Y_i Y_j$

$= \sum_{i\neq j}^{N}\sum^{N} \frac{P_j}{P_i} Y_i^2 - \sum_{i\neq j}^{N}\sum^{N} Y_i Y_j$

$= \sum\sum_{i\neq j} \left(P_i P_j \frac{Y_i^2}{P_i^2} - \sum\sum_{i\neq j} Y_i Y_j\right)$

$= \frac{1}{2}\sum_{i\neq j}^{N}\sum^{N} \left[P_i P_j \frac{Y_i^2}{P_i} + P_i P_j \frac{Y_j^2}{P_j^2} - 2 Y_i Y_j\right]$

16/27

$$= \frac{1}{2} \sum_{i \neq j}^{N} \sum^{N} \left(\frac{\gamma_i^2}{p_i^2} + \frac{\gamma_j^2}{p_j^2} - 2 \frac{\gamma_i}{p_i} \frac{\gamma_j}{p_j} \right) p_i p_j$$

$$= \frac{1}{2} \sum_{i \neq j}^{N} \sum^{N} \left(\frac{\gamma_i}{p_i} - \frac{\gamma_j}{p_j} \right)^2 p_i p_j$$

(b) $V(V_2) = V(E(V_2|V_1)) + E(V(V_2|V_1))$

$= V(\gamma_1 + E(V(V_2|V_1)))$

$= E(V(V_2|V_1))$

$$V(V_2 | V_1 = \frac{\gamma_i}{p_i}) = E((V_2 - \gamma)^2 | V_1 = \frac{\gamma_i}{p_i})$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\gamma_i + \frac{1-p_i}{p_j} \gamma_j - \gamma \right)^2 \frac{p_j}{1-p_i}$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\frac{\gamma_j}{p_{j|i}} - (\gamma - \gamma_i) \right) p_{j|i} ,$$

where $p_{j|i} = \frac{p_j}{1-p_i}$, $j \neq i$. Note that $\sum_{j \neq i} p_{j|i} = 1$ and

$\sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{\gamma_j}{p_{j|i}} \cdot p_{j|i} = \sum_{j \neq i} \gamma_j = \gamma - \gamma_i$. As in the proof of (a) above

$$V(V_2 | V_1 = \frac{\gamma_i}{p_i}) = \sum_{\substack{j=1 \\ j<j', j,j' \neq i}}^{N} \sum_{j'=1}^{N} \left(\frac{\gamma_j}{p_{j|i}} - \frac{\gamma_{j'}}{p_{j'|i}} \right)^2 p_{j|i} p_{j'|i}$$

$$= \sum_{\substack{j=1 \\ j<j', j,j' \neq i}}^{N} \sum_{j'=1}^{N} \left(\frac{\gamma_j}{p_j} - \frac{\gamma_{j'}}{p_{j'}} \right) p_j p_{j'} .$$

$$V(V_2) = E(V_2(V_2|V_1))$$

$$= \sum_{i=1}^{N} V(V_2 | V_1 = \frac{\gamma_i}{p_i}) P(V_1 = \frac{\gamma_i}{p_i})$$

$$= \sum_{i=1}^{N} \left(\sum_{\substack{j=1 \\ j<j', j,j' \neq i}}^{N} \sum_{j'=1}^{N} \left(\frac{\gamma_j}{p_j} - \frac{\gamma_{j'}}{p_{j'}} \right)^2 p_j p_{j'} \right) p_i$$

$$= \sum_{\substack{j=1 \\ 1<j'}}^{N} \sum_{j'=1}^{N} \left(\frac{\gamma_j}{p_j} - \frac{\gamma_{j'}}{p_{j'}} \right)^2 p_j p_{j'} \left(\sum_{\substack{i=1 \\ i \neq j, j'}}^{N} p_i \right)$$

$$= \sum_{j=1}^{N} \sum_{\substack{j'=1 \\ j<j'}}^{N} \left(\frac{Y_j}{P_j} - \frac{Y_{j'}}{P_{j'}} \right)^2 P_j P_{j'} (1 - P_j - P_{j'})$$

$$= \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i<j}}^{N} \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j} \right)^2 P_i P_j (1 - P_i - P_j).$$

Theorem $V(\hat{Y}_{DS}) = V(D) = \frac{1}{4} \sum_{\substack{i=1 \\ i<j}}^{N} \sum_{j=1}^{N} P_i P_j (2 - P_i - P_j) \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j} \right)^2$

Proof $V(\hat{Y}_{DS}) = V(D) = V\left(\frac{V_1 + V_2}{2}\right)$

$= \frac{1}{4} [V(V_1) + V(V_2)]$ $\quad (Cov(V_1, V_2) = 0)$

$= \frac{1}{4} \sum_{\substack{i=1 \\ i<j}}^{N} \sum_{j=1}^{N} \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j} \right)^2 (2 - P_i - P_j) P_i P_j,$

Remark: For $N \geq 3$,

$V(V_2) \leq V(V_1)$

and the equality holds iff $\left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j} \right)^2 = 0, \ \forall \ i \neq j$

i.e. iff $\frac{Y_i}{P_i} = c, \ i = 1 \cdots N,$ for some constant c.

Des Raj Estimator For General n

Let $P_i > 0, \ \forall \ i = 1 \cdots N$. For observed sample $\underline{s} = (u_{i_1}, \cdots, u_{i_n})$, define

$V_1 = \frac{Y_{i_1}}{P_{i_1}}, \quad V_2 = Y_{i_1} + \frac{Y_{i_2}}{P_{i_2}} (1 - P_{i_1}), \quad \cdots$

$V_j = Y_{i_1} + \cdots + Y_{i_{j-1}} + \frac{Y_{i_j}}{P_{i_j}} \left(1 - \sum_{r=1}^{j-1} P_{i_r}\right), \quad j = 3 \cdots n.$

and

$$\hat{Y}_{DS} = \frac{1}{n} \sum_{i=1}^{n} V_i = \bar{V}, \text{ say}.$$

Theorem (i) $E(V_j) = \gamma, \quad j = 1, \dots N$

and

$$E(\hat{\gamma}_{DS}) = \gamma .$$

(ii) $Cov(V_j, V_k) = 0, \quad \forall \ 1 \le j < k \le n$

(iii) $$\hat{V}(\hat{\gamma}_{DS}) = \frac{1}{n(n-1)} \sum_{j=1}^{n} (V_j - \hat{\gamma}_{DS})^2$$

Proof (i) for $i_1 \neq i_2 \neq \dots \neq i_{j-1}$

$$E(V_j \mid V_1 = \frac{\gamma_{i_1}}{p_{i_1}}, \ V_2 = \gamma_{i_1} + \frac{\gamma_{i_2}}{p_{i_2}}(1 - p_{i_1}), \dots, V_{j-1} = \gamma_{i_1} + \dots + \gamma_{i_{j-2}} + \frac{\gamma_{i_{j-1}}}{p_{i_{j-1}}} \times \left(1 - \sum_{\nu=1}^{j-2} p_{i_\nu}\right))$$

$$= \sum_{\substack{i_j = 1 \\ i_j \neq i_1 \ i_2 \dots \ i_{j-1}}}^{N} \left(\gamma_{i_1} + \dots + \gamma_{i_{j-1}} + \frac{\gamma_{i_j}}{p_{i_j}} (1 - p_{i_1} - \dots - p_{i_{j-1}}) \right) \frac{p_{i_j}}{1 - p_{i_1} - \dots - p_{i_{j-1}}}$$

$$= \gamma_{i_1} + \dots + \gamma_{i_{j-1}} + \sum_{\substack{i_j = 1 \\ i_j \neq i_1 \dots i_{j-1}}}^{N} \gamma_{i_j}$$

$$= \gamma .$$

$$\Rightarrow \quad E(V_j) = \gamma, \quad j = 1, \dots n$$

(ii) for $1 \le j < k \le N$

$$E(V_j V_k) = E\left(E(V_j V_k \mid V_1 \dots V_{k-1})\right)$$

$$= E\left(V_j \, E(V_k \mid V_1 \dots V_{k-1})\right)$$

$$= E(V_j \gamma) = \gamma E(V_j) = \gamma^2$$

$$\Rightarrow \quad Cov(V_j, V_k) = E(V_j V_k) - E(V_j) E(V_k) = 0$$

(iii) $$E\left(\sum_{j=1}^{n} (V_j - \hat{\gamma}_{DS})^2\right) = E\left(\sum_{j=1}^{n} (V_j - \bar{V})^2\right)$$

$$= E\left(\sum_{j=1}^{n} V_j^2\right) - n E(\bar{V}^2)$$

$$= \sum_{j=1}^{n} \left[V(V_j) + \gamma^2 \right] - n \left(V(\hat{\gamma}_{DS}) + \gamma^2 \right)$$

$$= \sum_{j=1}^{n} V(V_j) - n\, V(\hat{\theta}_{DS})$$

Also

$$V(\hat{\theta}_{DS}) = V(\bar{V})$$

$$= \frac{1}{n^2} \sum_{j=1}^{n} V(V_j) \qquad (\mathrm{Cov}(V_j, V_k) = 0, \; j \neq k)$$

Therefore

$$E\left(\sum_{j=1}^{n} (V_j - \hat{\theta}_{DS})^2 \right) = n^2 V(\hat{\theta}_{DS}) - n\, V(\hat{\theta}_{DS})$$

$$\Rightarrow \quad V(\hat{\theta}_{DS}) = E\left[\frac{1}{n(n-1)} \sum_{j=1}^{n} (V_j - \hat{\theta}_{DS})^2 \right]$$

$$\Rightarrow \quad \hat{V}(\hat{\theta}_{DS}) = \frac{1}{n(n-1)} \sum_{j=1}^{n} (V_j - \hat{\theta}_{DS})^2$$