MTH 432M: Introduction to Sampling Theory
Assignment No. 4: Unequal Probability Sampling

1. Let $P_1, \ldots, P_N$ be non-negative constants such that $\sum_{i=1}^{N} P_i = 1$. Under $PPSWR(n, P_1, \ldots, P_N)$, show that

$$\mathrm{Var}(\hat{Y}_{PPSWR}) = \frac{1}{n} \sum_{1 \leqslant i < j \leqslant N} \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j} \right)^2 P_i P_j$$

2. Consider a population of N units. Let V_1 be the variance of $\hat{Y}_{PPSWR}$ under $PPSWR(n, P_1, \ldots, P_N)$ and V_2 be the variance of $\hat{Y}_{PPSWOR}$ $(= \hat{Y}_{HT})$ under $PPSWOR(n, P_1, \ldots, P_N)$ with $\pi_i = nP_i, i = 1, \ldots, N$. Show that $V_2 \leqslant V_1$ iff $\pi_{ij} > \frac{n-1}{n}\pi_i\pi_j, \forall\ i \neq j$. Hence compare the variances of $\hat{Y}_{SRSWR}$ and $\hat{Y}_{SRSWR}$ based on random samples of size n each.

3. In the context of Problem 2 above, show that there exists a $PPSWOR(n, P_1, \ldots, P_N)$ with $\pi_i = nP_i, i = 1, \ldots, N$.

4. For the sample size $n = 2$, show that $\mathrm{Var}(\hat{Y}_{DS}) \leqslant \mathrm{Var}(\hat{Y}_{PPSWR})$ where the same weights $\underline{P} = (P_1, \ldots, P_N)$ are used under the $PPSWR$ and $PPSWOR$. Can the above result be generalized to a general sample size n?

5. Suppose that $Y_i = \alpha + \beta X_i,\ i = 1, \ldots, N$, where α and β $(\neq 0)$ are fixed constants. Show that $\mathrm{Var}(\hat{Y}_{PPSWR}) > \mathrm{Var}(\hat{Y}_{SRSWR})$ if

$$\frac{\beta^2}{\alpha^2} < \frac{\overline{X} - \tilde{X}}{\tilde{X}\sigma_X^2},$$

where $\tilde{X}$ is the harmonic mean of X_is.

6. Consider a populations $\mathcal{U} = (U_1, \ldots, U_N)$ of N units. A sample of size n is drawn according to the Midzuno SRSWOR sampling scheme: in the first draw a unit is selected such that the probability of selecting unit U_i is P_i $(\sum_{i=1}^{N} P_i = 1)$. The remaining $n-1$ units are selected with SRSWOR. Show that

   (i) $\pi_i = \frac{N-n}{N-1}P_i + \frac{n-1}{N-1},\ i = 1, \ldots, N;$

   (ii) $\pi_{ij} = \frac{n-1}{N-1}\left[\frac{N-n}{N-2}(P_i + P_j) + \frac{n-2}{N-2}\right], i, j = 1, \ldots, N,\ i \neq j;$

   (iii) Show that the Yates-Grundy estimator is non-negative.

7. Consider a population $\mathcal{U} = (U_1, U_2, U_3, U_4, U_5)$ of five units with study variables $\mathcal{Y} = (Y_1, Y_2, Y_3, Y_4, Y_5) = (1, 1, 2, 2, 3)$. A sample of size $n = 2$ is drawn according to the following sampling design:

$$P(s_1, s_2)) = \Pr((S_1, S_2) = (s_1, s_2)) = \begin{cases} \frac{1}{2}, & \text{if } (s_1, s_2) = (U_1, U_2) \\ \frac{1}{6}, & \text{if } (s_1, s_2) \in \{(U_3, U_4), (U_3, U_5), (U_4, U_5)\ \} \\ 0, & \text{otherwise} \end{cases}$$

   (a) Calculate the first and second order inclusion probabilities and find the probability distribution of the Horvitz-Thompson estimator $\hat{Y}_{HT}$;

   (b) Find the probability distribution of $\hat{Y}_{HT}^2$ and verify if it is an unbiased estimator of Y^2. Find the variance of $\hat{Y}_{HT}^2$.

   (c) Is it a $PPSWOR(2, P_1, \ldots, P_5)$ design for some P_is?

   (d) Let $P_i, i = 1, \ldots, 5$, denote the probability that unit U_i is selected in the first draw. Consider the Des Raj estimator $\hat{Y}_{DS}$ based on $PPSWOR(2, P_1, \ldots, P_5)$. Find the probability distribution, mean and variance of $\hat{Y}_{DS}$. Is $\hat{Y}_{DS}$ an unbiased estimator of Y?

   (e) Compare the performances of of $\hat{Y}_{HT}$ and $\hat{Y}_{DS}$ discussed above.

8. Consider a population $\mathcal{U} = (U_1, \ldots, U_N)$ of N units from which a sample of $n = 2$ units is drawn according to the following sampling scheme. The first unit of the sample is drawn in a manner that the probability of selecting U_i is $p_i, i = 1, \ldots, N$ $(\sum_{i=1}^{N} p_i = 1)$, and the second unit is selected in a manner that if unit U_i is drawn in the first draw then the probability of drawing any unit $U_j, j \neq i$, is proportional to q_j $(\sum_{i=1}^{N} q_i = 1)$, where $p_i = \sum_{j=1, j\neq i}^{N} p_j \frac{q_i}{1-q_j}, i = 1, \ldots, N$. Prove that

   (a) $\pi_i = 2p_i, i = 1, \ldots, N$;

   (b) $\mathrm{Var}(\sum_{i\in \underline{S}} \frac{Y_i}{2p_i}) \leqslant \mathrm{Var}(\hat{Y}_{PPSWR})$.