ON DOUBLE SAMPLING FOR PPS ESTIMATION

By Des Raj

Statistical Service of Greece, Athens

1. Introduction. Usually the double sampling technique is used for purposes of stratification or for ratio or regression estimation. There may, however, be situations in which it is considered important to select the sample units with probability proportionate to some measure of size x, information on which is not readily available but could be collected at moderate cost for a fairly large sample. For example, in sampling a list of agricultural holders it may be considered very desirable to select holdings with probability proportionate to area. In this case one may select a fairly large random sample of holdings from the list, obtain information on areas by interviewing farmers, select a small sample of holdings with probability proportionate to area and collect information on the desired characteristics from this small sample. The units selected may be subsampled if they are large. The purpose of this paper is to develop theory for this technique when the second sample is a subsample of the first (Case 1) or it is selected independently (Case 2). For related work, references at the end may be consulted.

2. Theory for Case 1. Using the notation from Cochran (1963), an unbiased estimate of the population total is given by

$$\hat{Y} = (N/n')(x'/n) \sum_1^n (y_i/x_i) \tag{1}$$

where x' is the total of the variate x for the initial sample and $\sum$ denotes summation over the second sample. The initial sample of size n' is selected with equal probabilities without replacement and the second sample is selected with pps with replacement. Using theorems on conditional expectations and variances (Raj (1956) p. 271) we have

$$\begin{aligned} V_1 E_2(\hat{Y}) &= N^2(n'^{-1} - N^{-1})(N-1)^{-1} \sum_1^N (y_i - \bar{Y})^2 \\ &= N(N-n')n'^{-1}S_y^2 \end{aligned} \tag{2}$$

$$\begin{aligned} E_1 V_2(\hat{Y}) &= \frac{N}{N-1} \frac{n'-1}{n'} \frac{1}{n} \left(X \sum_1^N \frac{y_i^2}{x_i} - Y^2 \right) \\ &= \frac{N}{N-1} \frac{n'-1}{n'} V_{\text{pps}} \end{aligned} \tag{3}$$

so that the variance of $\hat{Y}$ is given by

$$V(\hat{Y}) = N(N-1)^{-1}(n'-1)n'^{-1}V_{\text{pps}} + N(N-n')n'^{-1}S_y^2 \tag{4}$$

Received 30 July 1963.