A closer examination of two-stage cluster sampling assuming equal-size clusters

The relationships between two-stage cluster sampling and other methods are more easily seen if we simplify matters by assuming equal-size clusters. Thus, lets assume that all clusters contain the same number, $\overline{M}$, of elements, and that equal size samples of size m, are taken from each cluster. In this case, the estimator of μ simplifies:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij},$$

which is the overall sample average. The variance expression also simplifies:

$$\begin{aligned}
\widehat{V}(\widehat{\mu}) &= \left(\frac{N-n}{N}\right)\left(\frac{1}{\overline{M}^2}\right)\frac{s_b^2}{n} + \frac{1}{nN\overline{M}^2}\sum_{i=1}^{n} M_i^2\left(\frac{M_i - m_i}{M_i}\right)\frac{s_i^2}{m_i} \\
&= \left(\frac{N-n}{N}\right)\left(\frac{1}{n\overline{M}^2}\right)\frac{\sum_{i=1}^{n}\left(M_i\overline{y}_i - \overline{M}\widehat{\mu}\right)^2}{n-1} + \frac{1}{nN\overline{M}^2}\sum_{i=1}^{n}\overline{M}^2\left(\frac{\overline{M}-m}{\overline{M}}\right)\frac{s_i^2}{m} \\
&= \left(\frac{N-n}{N}\right)\left(\frac{1}{n\overline{M}^2}\right)\frac{\sum_{i=1}^{n}\overline{M}^2\left(\overline{y}_i - \widehat{\mu}\right)^2}{n-1} + \left(\frac{\overline{M}-m}{\overline{M}}\right)\left(\frac{1}{N}\right)\frac{\sum_{i=1}^{n} s_i^2/n}{m} \\
&= \left(\frac{N-n}{N}\right)\frac{MSB}{nm} + \left(\frac{\overline{M}-m}{\overline{M}}\right)\left(\frac{1}{N}\right)\frac{MSW}{m},
\end{aligned}$$

where

$$MSB = \frac{m\sum_{i=1}^{n}\left(\overline{y}_i - \widehat{\mu}\right)^2}{n-1}, \text{ and } MSW = \sum_{i=1}^{n} s_i^2/n.$$

Here are some conclusions from this expression:

1. If N is large, then $\widehat{V}(\widehat{\mu}) \approx MSB/(nm)$ and the variance just depends on the cluster means $\overline{y}_i$. Thus, using an easy method for sampling within clusters, like systematic random sampling, will not usually present any problems.

2. If $m = \overline{M}$, then two-stage cluster sampling simplifies to one-stage cluster sampling.

3. If $n = N$, then $\widehat{V}(\widehat{\mu}) = \left(\frac{\overline{M}-m}{\overline{M}}\right)\frac{MSW}{mn}$, which is the same expression as used in stratified random sampling, with $n = N$ strata and m observations sampled from each.

Thus we can now see that two-stage cluster sampling is a very general method that includes single-stage cluster sampling and stratified random sampling as special cases.