Stratified resampling

Suzie Brown

22 December 2020

In this note I will show that our theorem also applies to stratified resampling. The calculations are similar to those for stochastic rounding, since in both cases there are only a small number of values that each ν_i can take, given w_i. In stratified resampling, the value of ν_i is almost surely restricted conditional on w_i:

$$\nu_i \mid w_i = \begin{cases} \lfloor N w_i \rfloor - 1 & \text{w.p. } p_{-1} \\ \lfloor N w_i \rfloor & \text{w.p. } p_0 \\ \lfloor N w_i \rfloor + 1 & \text{w.p. } p_1 \\ \lfloor N w_i \rfloor + 2 & \text{w.p. } p_2 \end{cases} \tag{1}$$

where $p_{-1} + p_0 + p_1 + p_2 = 1$.

Case $\lfloor N w_i \rfloor = 0$.

$$\begin{aligned} \mathbb{E}[(\nu_i)_2 \mid \mathcal{H}_t] &= p_{-1}(\lfloor N w_i \rfloor - 1)_2 + p_0(\lfloor N w_i \rfloor)_2 + p_1(\lfloor N w_i \rfloor + 1)_2 + p_2(\lfloor N w_i \rfloor + 2)_2 = 2p_2 \geq 0, \\ \mathbb{E}[(\nu_i)_3 \mid \mathcal{H}_t] &= p_{-1}(\lfloor N w_i \rfloor - 1)_3 + p_0(\lfloor N w_i \rfloor)_3 + p_1(\lfloor N w_i \rfloor + 1)_3 + p_2(\lfloor N w_i \rfloor + 2)_3 = 0. \end{aligned}$$

The other cases follow similarly.

Case $\lfloor N w_i \rfloor = 1$.

$$\begin{aligned} \mathbb{E}[(\nu_i)_2 \mid \mathcal{H}_t] &= 2p_1 + 6p_2, \\ \mathbb{E}[(\nu_i)_3 \mid \mathcal{H}_t] &= 6p_2 \leq 2p_1 + 6p_2. \end{aligned}$$

Case $\lfloor N w_i \rfloor = 2$.

$$\begin{aligned} \mathbb{E}[(\nu_i)_2 \mid \mathcal{H}_t] &= 2p_0 + 6p_1 + 24p_2, \\ \mathbb{E}[(\nu_i)_3 \mid \mathcal{H}_t] &= 6p_1 + 24p_2 \leq 2p_0 + 6p_1 + 24p_2. \end{aligned}$$

Case $\lfloor N w_i \rfloor \geq 3$.

$$\begin{aligned} \mathbb{E}[(\nu_i)_2 \mid \mathcal{H}_t] &= p_{-1}(\lfloor N w_i \rfloor - 1)_2 + p_0(\lfloor N w_i \rfloor)_2 + p_1(\lfloor N w_i \rfloor + 1)_2 + p_2(\lfloor N w_i \rfloor + 2)_2 \\ &\geq p_{-1} + p_0 + p_1 + p_2 = 1, \\ \mathbb{E}[(\nu_i)_3 \mid \mathcal{H}_t] &= p_{-1}(\lfloor N w_i \rfloor - 1)_3 + p_0(\lfloor N w_i \rfloor)_3 + p_1(\lfloor N w_i \rfloor + 1)_3 + p_2(\lfloor N w_i \rfloor + 2)_3 \\ &\leq (p_{-1} + p_0 + p_1 + p_2)(\lfloor N w_i \rfloor + 2)_3 \leq (a^2 + 2)_3. \end{aligned}$$

Putting these together,

$$\begin{aligned} \sum_{i=1}^N \mathbb{E}[(\nu_i)_2 \mid \mathcal{H}_t] &\geq (2p_1 + 6p_2)|\{i : \lfloor N w_i \rfloor = 1\}| + (2p_0 + 6p_1 + 24p_2)|\{i : \lfloor N w_i \rfloor = 2\}| + |\{i : \lfloor N w_i \rfloor \geq 3\}|, \\ \sum_{i=1}^N \mathbb{E}[(\nu_i)_3 \mid \mathcal{H}_t] &\leq (2p_1 + 6p_2)|\{i : \lfloor N w_i \rfloor = 1\}| + (2p_0 + 6p_1 + 24p_2)|\{i : \lfloor N w_i \rfloor = 2\}| + (a^2 + 2)_3|\{i : \lfloor N w_i \rfloor \geq 3\}|. \end{aligned}$$

We have (ignoring for the moment the possibility of dividing by zero):

$$
\begin{aligned}
&\frac{(N)_2}{(N)_3}\frac{\sum_{i=1}^N \mathbb{E}[(\nu_i)_3 \mid \mathcal{H}_t]}{\sum_{i=1}^N \mathbb{E}[(\nu_i)_2 \mid \mathcal{H}_t]} \\
&\qquad \leq \frac{(N)_2}{(N)_3}\frac{2p_1+6p_2)|\{i : \lfloor Nw_i \rfloor = 1\}| + (2p_0+6p_1+24p_2)|\{i : \lfloor Nw_i \rfloor = 2\}| + (a^2+2)_3|\{i : \lfloor Nw_i \rfloor \geq 3\}|}{(2p_1+6p_2)|\{i : \lfloor Nw_i \rfloor = 1\}| + (2p_0+6p_1+24p_2)|\{i : \lfloor Nw_i \rfloor = 2\}| + |\{i : \lfloor Nw_i \rfloor \geq 3\}|} \\
&\qquad \leq \frac{1}{N-2}\left(\frac{(2p_1+6p_2)|\{i : \lfloor Nw_i \rfloor = 1\}| + (2p_0+6p_1+24p_2)|\{i : \lfloor Nw_i \rfloor = 2\}|}{(2p_1+6p_2)|\{i : \lfloor Nw_i \rfloor = 1\}| + (2p_0+6p_1+24p_2)|\{i : \lfloor Nw_i \rfloor = 2\}|} + \frac{(a^2+2)_3|\{i : \lfloor Nw_i \rfloor \geq 3\}|}{|\{i : \lfloor Nw_i \rfloor \geq 3\}|}\right) \\
&\qquad = \frac{1}{N-2}(1+(a^2+2)_3) \leq \frac{1}{N-2}(a^2+2)^3 =: b_N.
\end{aligned}
$$

Some, but not all, of the denominator terms may be equal to zero, in which case the numerator and denominator should really be modified before taking the ratio, but this will never actually cause a problem. In any case the terms that are left in the ratio yield a bound less than the b_N defined here. (If no terms are left then both sides are equal to zero and we may set b_N arbitrarily.)

To apply the theorem, it yet remains to prove the finite time scale condition, or otherwise include it in the statement of the corollary. I expect a similar argument to that used for stochastic rounding should work here. We will require a similar condition to ensure the weights are bounded away from $(1, \dots, 1)/N$ since, like stochastic rounding, stratified resampling is degenerate under equal weights.