Square GPS Sampling Algorithm

27 December 2016

Patrick D. Emond

Algorithm

Once a town has been selected for sampling the following algorithm is performed on that town:

1. determine width of squares based on user input "number of squares per side"*
2. select random point in town and identify square, S, the point falls into
3. make a set of all unselected buildings, B, which fall in square S
4. if B is empty then go back to step 2 (if this happens 1000 times then quit with an error)
5. select a random building, b, in set B
6. add all individuals** in all households in building b to the sample
7. if number of individuals in the sample is less than the sample size return to step 2

* "number of squares per side" is a parameter defined by the user when running the sampler and corresponds to how many squares there are along any one axis for all towns (eg: 10 would mean that every town is divided into $10 \times 10 = 100$ squares total since there are always exactly the same number of squares along the X and Y axes)

** selected individuals may be restricted to children only, or one individual per household based on input parameters defined by the user

Sampling Weights

Probability of selecting town i is:

$$P_i = P\big(\text{select town i}\big) = \frac{n_i}{N}$$

where n_i = population of town i, and N = total population. Strictly, it's number of buildings or households, but we ignore that, and assume the number of people per household in the different towns is constant.

Probability of selecting square $\overset{\square}{j}$ is:

$$P_{\underset{j}{\square} \bigvee i} = P\big(\text{select square } \overset{\square}{j} \text{ within town } i\big) = \frac{n_s}{S}$$

where n_s = number of squares chosen within the town (including empty squares), and S = number of squares in the town. The number of squares chosen will vary depending on the size of squares and population density.

Probability of selecting household k in square $\overset{\square}{j}$ is:

$$P_{k \bigvee \overset{\square}{j}} = P\big(\text{select household } k \text{ within square } \overset{\square}{j}\big) = \frac{1}{n_{\underset{ij}{\square}}}$$

where $n_{\underset{ij}{\square}}$ = number of households in square $\overset{\square}{j}$.

Therefore the probability of sampling household k is:

$$P_{\underset{ijk}{\square}} = P\big(\text{select household } \overset{\square}{ijk}\big) = P_i \times P_{\underset{j}{\square} \bigvee i} \times P_{k \bigvee \overset{\square}{j}}$$

Making the weight for household $ij\overset{\square}{k}$ equal to:

$$\begin{aligned} w_{ij\overset{\square}{k}} &= \left[P_{ij\overset{\square}{k}}\right]^{-1} \\ &= \left[P_i \times P_{j\overset{\square}{\bigvee} i} \times P_{k\overset{\square}{\bigvee} j}\right]^{-1} \\ &= \frac{n_{ij\overset{\square}{}} NS}{n_i n_s} \end{aligned}$$

where $n_{ij\overset{\square}{}}$ = number of households in square $\overset{\square}{j}$, n_i = population of town i, n_s = number of squares chosen within the town, S = number of squares in the town and N = total population.

If all eligible people in the household are chosen, then this is the weight for all those individuals. When only one person per household is selected, let n_{kl} be the number of eligible people in the household (e.g.: the number of children or number of adults):

$$P_{l\bigvee k} = P\big(\text{select person}l\text{within household}k\big) = \frac{1}{n_{kl}}$$

Therefore the probability of sampling an individual l is:

$$P(\text{select individiual}) = P_{ijkl\overset{\square}{}} = P_{l\bigvee k} \times P_{ijk\overset{\square}{}}$$

Making the weight for individual $ijk\overset{\square}{l}$ equal to:

$$w_{ijkl\overset{\square}{}} = \frac{n_{ij\overset{\square}{}} n_{kl} NS}{n_i n_s}$$