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 is constant.

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

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

where n_s = number of squares chosen within the town, and S = number of squares in the town. Since all towns always have the same number of squares we do not have to include $P_{\overset{\square}{j} \bigvee i}$ when determining sampling weights.

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:

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

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

$$\begin{aligned} w_{\overset{\square}{ijk}} &= \frac{1}{P_{\overset{\square}{ijk}}} \\ &= \frac{1}{P_i \times P_{k\bigvee\overset{\square}{j}}} \quad \Big(\text{since} P_{\overset{\square}{j}\bigvee i} \text{ is identical in all towns}\Big) \\ &= \frac{n_{\overset{\square}{ij}}}{n_i} N \end{aligned}$$