Synthetic population generation: the hidden model in agent based models

K. Chapuis¹

¹ IRD, UMR 228 Espace-Dev

kevin.chapuis@gmail.com

Keywords

multi-layer synthetic population, Gama library, Gen*

Abstract

Almost 20 years after the creation of the KIDS acronyme [Edmonds and Hamilton, 2004] and after decades of descriptive ABM, the effort modelers put into the generation of a realistic population of agents remain limited [Chapuis and Taillandier, 2019]. If we consider it as a mandatory piece of expertise required to initialize a data-driven agent based social simulation, the generation of a synthetic population is in a strange methodological position, to say the least: many publications have proposed algorithms and libraries, discussing in extenso about a variety of methods and tools, but synthetic population generation is still a modeling blind spot for most ABM researchers. In fact, principles and challenges are often misunderstood [Chapuis et al. 2019], the tools are merely used, resulting in a start-from-scratch syndrome [Lovelace et al. 2015] and the synthetic population characteristics are rarely discussed when reporting ABMs (e.g. there is only one occurrence of "agent population" in the supplementary materials related to the initialization part of the latest updated ODD version publish in JASSS - Grimm et al. 2020). In this proposal, we try to illustrate the main difficulties a modeler can face when dealing with synthetic population generation using the creation of individuals into household agents in the Gama platform, as an example. Based on french demographic data, we detail a simple (yet complicated) process using a minimal synthetic reconstruction algorithm from the Gen* library, to build the population of individual agents, combined with a hierarchical sampling algorithm, written in gaml, to construct households. After a brief description of the approach, we identify three main limitations: the algorithm needs for new developments when new data are added or when data changes, the process requirement in terms of modeling effort to fill the gaps left by missing data, and the global approach complexity, including the necessity to conduct parameters sensitivity analysis and to build population quality assessment indicators. Based on this feedback from experience, we advocate for the use of generic platform extension tools to support the development of dedicated models to generate synthetic population, hence discovering the hidden model in models.

Additional material

https://github.com/ANRGenstar