Population synthesis as scenario generation for simulation-based planning under uncertainty

Joel Dyer, Arnau Quera-Bofarull, Nicholas Bishop, J. Doyne Farmer, Anisoara Calinescu, Michael Wooldridge

University of Oxford

{joel.dyer, arnau.quera-bofarull, nicholas.bishop}@cs.ox.ac.uk

Overview

- Agent-based models (ABMs) rely on synthetic populations.
- Existing methods for generating populations leverage real-world datasets of agent-level attributes.
- Instead we treat synthetic population generation as **scenario generation**:

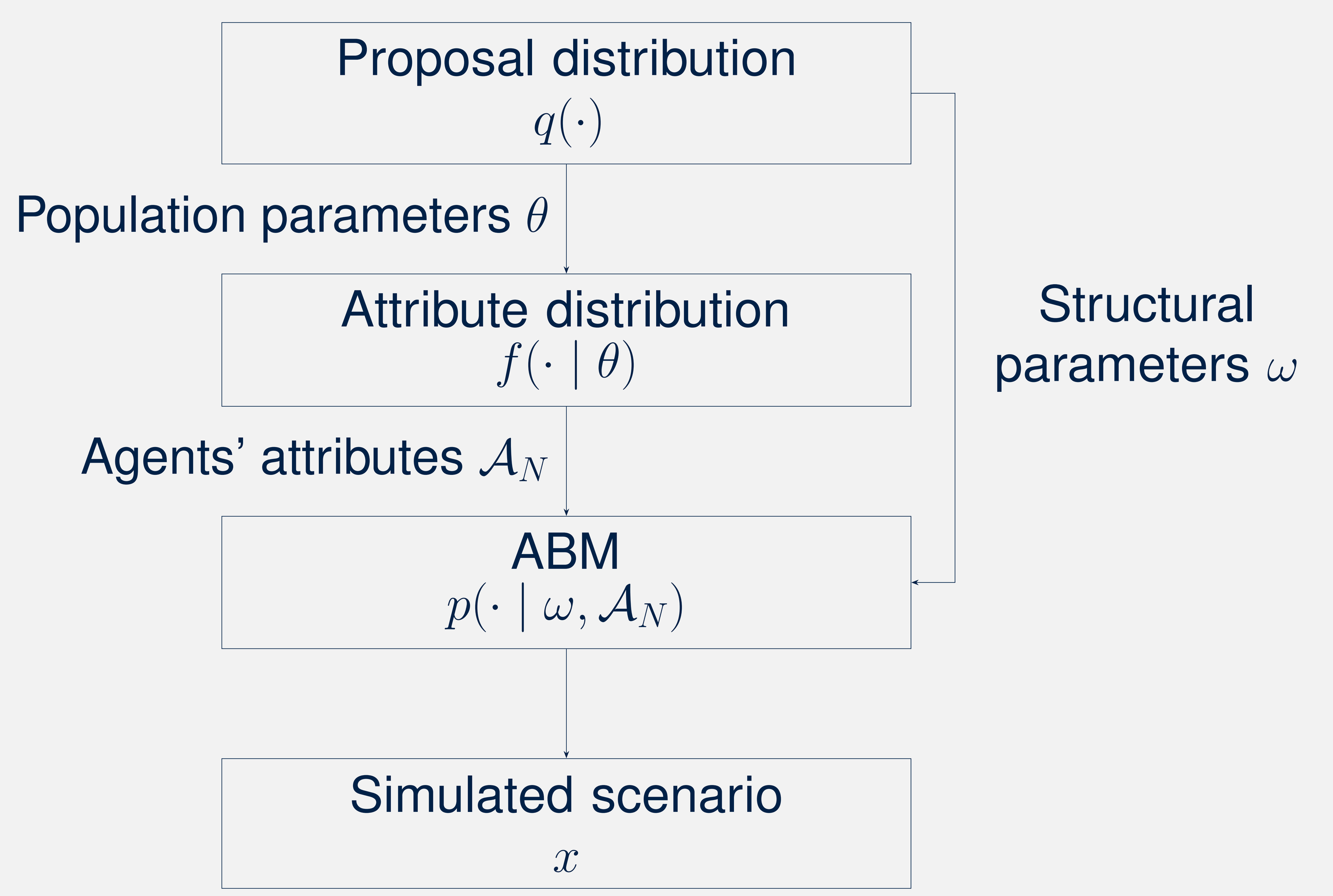
  "Under the assumption that the model is correct, what might the population in this system need to look like in order to realise a user-specified scenario?"

Limitations of Existing Approaches

- Rely on individual-level datasets - **often unavailable**!
- Perform population synthesis upfront - **population is not informed by the behaviour of the ABM!**

Our Approach

- Sample ABM **structural parameters** ω and **population parameters** $\boldsymbol{\theta}$ from a **proposal distribution** $q(\cdot)$:
$$(\omega, \theta) \sim q(\cdot).$$
- Sample agent attributes $\mathcal{A}_N$ from an **attribute distribution** with parameter θ:
$$\mathcal{A}_N \sim f(\cdot \mid \theta).$$
- **Forward-simulate** ABM with $\mathcal{A}_N$ and ω to generate output $\mathbf{x}$:
$$\mathbf{x} \sim p(\cdot \mid \omega, \mathcal{A}_N).$$



Challenge: Choosing a Proposal Distribution

- Main challenge lies in specifying the proposal distribution q.
- **Idea:** Allow the modeller to define a **loss function** $\ell : \mathcal{X} \to [0, \infty)$ describing proximity of an outcome to a desired scenario.
- We can lift this loss to a loss over structural and population parameters:
$$\mathcal{L}(\omega, \theta) = \mathbb{E}_{p(x|\omega,\theta)}[h_\epsilon(\ell(x))],$$
where h_ϵ is method-dependent function parameterised by $\epsilon > 0$.

Method 1: Threshold-based Sampling (TBS)

- Let h_ϵ be a probability kernel function, before letting
$$q(\omega, \theta) \propto \mathcal{L}(\omega, \theta)$$
- **Example:** Setting
$$h_\epsilon(\cdot) \propto \mathbb{I}(\cdot \leq \epsilon)$$
corresponds to
$$q(\omega, \theta) \propto \mathbb{P}(\{\ell(x) \leq \epsilon \mid x \sim p(\cdot \mid \omega, \theta)\}).$$
That is, the pair (ω, θ) is down-weighted if the probability it produces scenarios within an ϵ-ball of a desired scenario is low.
- Smoother choices include $h_\epsilon \propto \exp(- \cdot / \epsilon)$.
- We may sample from q in a **Monte Carlo fashion**.
  - In our experiments we use **sequential Monte Carlo sampling (TBS-SMC).**
- Hyperparameter ϵ controls the variance of q.

Method 2: Variational Optimisation (VO)

- Set h_ϵ to the identity function and consider a **parameterised family** of proposal distributions $\mathcal{Q} = \{q(\cdot \mid \phi) \mid \phi \in \Phi\}$.
  - In our experiments we parameterise with a **normalising flow (VO-NF).**
- Solve the resulting **variational optimisation problem**:
$$q = \arg\min_{\phi \in \Phi} \left\{ \mathbb{E}_{\omega,\theta \sim q(\omega,\theta|\phi)} \left[\mathcal{L}(\omega, \theta)\right] - \gamma \cdot \mathbb{H}(q(\cdot \mid \phi)) \right\},$$
where $\mathbb{H}$ is the differential entropy and $\gamma \geq 0$ is a hyperparameter.
- Setting $\gamma = 0$ causes q to **collapse** into a degenerate distribution whose mass is concentrated on pairs (ω, θ) that minimise $\mathcal{L}$.
- Larger γ encourages greater **diversity**.
- We may estimate q through **stochastic gradient descent**.

Experiments: Axtell's Model of Firms

- Model agents moving between firms across time $t \in [0, 1]$.
- Agent n works with some effort level $e_n^t \in [0, 1]$ at time t.
- Each agent reevaluates their situation at an agent-specific rate ρ_n.
- Each agent also maintains an agent-specific parameter $v_n \in [0, 1]$ describing their preference for leisure vs income.
- When reevaluating, agents decide between
  - adjusting their effort level,
  - moving to an existing firm,
  - or starting a new firm.

Experiments: Designing a Loss

- **Question**: *Can an initially hardworking population become lazy over time?*
- We choose the following loss function that measures the difference between the average effort of agents at the beginning and end of the time horizon:
$$\ell(x) = \frac{1}{N} \sum_{n=1}^{N} (e_n^1 - e_n^0 + 1)$$
- Choose the following attribute distribution:
$$f(e_n^0, \nu_n, \rho_n \mid \theta) = \text{Beta}(e_n^0 \mid \varepsilon_a, \varepsilon_b) \cdot \text{Beta}(\nu_n \mid g_a, g_b) \cdot \text{Gamma}(\rho_n \mid \varrho_a, \varrho_b)$$

Results

- Both methods easily outperform a uniform proposal.
- Also, both methods provide insight into the properties of an ideal population:
  - Agents must **strongly prefer leisure** over income
  - and must **frequently reevaluate** their position!