Population uncertainty from Census DP

Armelle Duston

June 2025

1 Proposed Models

1.1 Exact Benchmarking

$$U_j|P^*_{M_j} \sim \mathrm{BM}_{\mathrm{exact}}(P^*_{M_j}) = \mathbb{I}\left(\sum_{i\in M_j} P^*_i = U_j \ \ \& \ \ P^*_i \in \mathbb{N}_0\right) \tag{1}$$

$$P^*_i|P_i \sim G(P_i, \tau) = \mathrm{Laplace}(P_i, \tau) \tag{2}$$

$$P_i \sim \mathrm{Poisson}(\exp(S)) \tag{3}$$

$$S \sim \mathrm{CAR}(\rho) : s_i|s_{-i} \sim \mathrm{N}\left(\rho\bar{s}_{\delta_i}, \frac{\sigma_s^2}{n_{\delta_i}}\right) \tag{4}$$

Things that are known:

- $P^*_i, i = \{1, \ldots, n\}$ (reported noisy/processed population counts)
- τ (related to Census's "privacy budget" parameter)
- $\delta_i, i = \{1, \ldots, n\}$ (pre-specified neighborhood structure)

Things that are not known and require prior:

- $U_j, j \in \{1, \ldots, J\}$ (values to which P_{M_j}'s are benchmarked, "prior" specified above)
- $P_i, i \in \{1, \ldots, n\}$ (true population counts, prior specified above)
- S (spatial random effect, prior specified above)
- ρ (spatial correlation), uniform(0,1) prior (?)
- σ_s^2 (spatial variance), gamma(α, β) prior (?)

Posterior Distribution

$$f(U, P, S, \rho, \sigma_s^2|P^*) = f(U|P^*)f(P^*|P)f(P|S)f(S|\rho, \sigma_s^2)f(\rho)f(\sigma_s^2)$$

1.2 Inexact Benchmarking

Everything is the same except

$$\eta U_j | P^*_{M_j} \sim \mathrm{BM}_{\mathrm{inexact}}(P^*_{M_j}) = \mathrm{Poisson}(\eta \sum_{i \in M_j} P^*_i)\, \mathbb{I}(P^*_i \in \mathbb{N}_0) \tag{5}$$

Where η is a pre-specified discripancy parameter which controls how far off the counts are allowed to be from the benchmarks

2 Simulation Study

2.1 Plan

1. Simulate data according to the model above including spatial correlation, noise injection, and benchmarking
2. Assume simple spatial model
3. Assume spatial model + noise injection
4. Assume spatial model + noise injection + benchmarking
5. Evaluate performance using
    - bias
    - variance
    - credible interval coverage
    - anything else?