Approximate Bayesian Computation (ABC) for Infectious Disease Modelling

Reading Notes on Minter & Retkute (Epidemics 2019)

1 Why ABC?

Lots of epidemic models (SIR/SEIR variants, spatial kernels, under–reporting, seasonality) have likelihoods that are ugly or basically impossible to write down. Idea: **Approximate Bayesian Computation** skips the likelihood. Instead:

$$\theta \sim p(\theta) \implies \text{simulate data } D^{\text{sim}} \implies \text{keep } \theta \text{ if } \rho(S^{\text{sim}}, S^{\text{obs}}) \leq \epsilon.$$

If $\epsilon \to 0$ and summaries are sufficient, we approximate the true posterior.

2 Core Ingredients

- Prior $p(\theta)$: e.g. transmission rates, reporting probability, mixing exponent α.
- Simulator: TSIR / spatial kernel model, etc.
- Summary stats $S(\cdot)$: could just be the full time series; often we compress to peak size, peak week, cumulative incidence, etc.
- Distance ρ: usually something like sum of squared errors on log counts.
- Tolerance ϵ: fixed for rejection ABC, or a decreasing schedule $\epsilon_1 > \cdots > \epsilon_G$ for SMC.

3 Algorithms (Quick Tour)

3.1 Rejection ABC

Loop: sample θ, simulate, compute distance, accept if below ϵ. Pros: dead simple and embarrassingly parallel. Cons: once ϵ is small, acceptance rate tanks.

3.2 MCMC–ABC

Metropolis–Hastings where likelihood ratio replaced by indicator $\mathbb{I}\{\rho \leq \epsilon\}$. Better than pure rejection but can get sticky if proposal is bad.

3.3 SMC–ABC (the good one)

Maintain a particle population. For generation g use a tighter tolerance ϵ_g. Resample/move particles using a kernel K; weights:

$$w_g^{(i)} \propto \frac{p(\theta_g^{(i)})}{\sum_j w_{g-1}^{(j)} K(\theta_g^{(i)} - \theta_{g-1}^{(j)})}.$$

Adapt ϵ_g via a quantile so ESS stays reasonable. Local (nearest–neighbour) covariance matrices can cut simulations (slightly lower ESS but often worth it).

4 Model Examples in the Paper

- **Synthetic SIR/TSIR**: sanity check; shows SMC posterior tighter for same compute.
- **Measles (TSIR)**: $I_{t+1} \sim \text{Poisson}(\beta_t S_t I_t^{\alpha}/N)$, $S_{t+1} = S_t + B_t - I_{t+1}$. Seasonal β_t recovered; α posterior shrinks nicely under SMC.
- **Plant virus (spatial)**: distance–based kernel. Local covariance SMC cuts simulations by $\sim 40\%$ vs global.

5 Practical Tips (what I should remember)

- Good summaries matter. Bad summaries $\Rightarrow$ biased posterior even if ϵ tiny.
- Start with a big ϵ (e.g. accept $\approx$ 30–50%) then shrink using quantiles of distances.
- Kernel covariance: weighted sample covariance $\times$ scale factor. Try local covariance for speed.
- Monitor: acceptance rate, Effective Sample Size (ESS), posterior predictive checks.
- Parallelise simulations! SMC usually saves an order of magnitude runs compared to vanilla rejection.

6 Main Takeaways

SMC–ABC dominates simple rejection in efficiency. Nearest–neighbour (local) kernels reduce the number of expensive simulations further, with only mild ESS loss. ABC framework is flexible enough for temporal + spatial infectious disease models.

7 How This Connects to My TSIR–PINN

Treat neural network parameters (or spline coefficients) of $\beta(t)$ as θ. Simulator: forward TSIR with Poisson observation. Summaries: full weekly log incidence or reduced set {peak week, peak height, cumulative cases}. Plan: prototype rejection ABC on a single state, then move to SMC–ABC for joint states. Posterior predictive draws give uncertainty bands to overlay on current PINN forecasts.