Initial Microsatellite Model

1. Model Overview

It simulates the mutation-driven evolution of microsatellite loci within a proliferating cell population. The system tracks two variables per locus:

$$L_j(t) : \text{current microsatellite length change (delta length)}, \quad C_j(t) : \text{number of cells containing state } L_j.$$

2. Initial Conditions

At $t = 0$:

$$L_j(0) = 0, \quad C_j(0) = 1, \quad \forall j \in \{1, \ldots, N_{\text{loci}}\}$$

with:

$$N_{\text{loci}} = 1000.$$

All loci start in an unmutated state, with one cell per locus.

3. Model Parameters

$$r : \text{exponential growth rate},$$
$$\mu_i : \text{insertion rate per division},$$
$$\mu_d : \text{deletion rate per division},$$
$$\Delta t : \text{discrete simulation time step (fixed at 0.1)}.$$

Per-step probabilities:

$$p_i = \min(\mu_i \Delta t, 0.1), \quad p_d = \min(\mu_d \Delta t, 0.1), \quad p_s = 1 - p_i - p_d.$$

4. Clonal Expansion Dynamics

At time t, the population size follows exponential growth:

$$N(t) = e^{rt}.$$

The number of newly divided cells in the time interval $[t - \Delta t, t]$ is approximated as:

$$N_{\text{new}} = N(t) - N(t - \Delta t) = e^{rt} - e^{r(t - \Delta t)}.$$

Each locus j receives a share of the new cells proportional to its prevalence:

$$P_j(t) = \frac{C_j(t)}{\sum_{k=1}^{N_{\text{loci}}} C_k(t)}.$$

New cells assigned to locus $j \sim \text{Multinomial}(N_{\text{new}}, P_1(t), \ldots, P_{N_{\text{loci}}}(t)).$

Cell count updates as:

$$C_j(t + \Delta t) = C_j(t) + \text{assigned new cells}.$$

5. Mutation Dynamics

At each time step, mutations occur independently at each locus j. Define a mutation outcome random variable:

$$X_j(t) \sim \text{Categorical}(p_i, p_d, p_s),$$

Microsatellite length updates for each locus:

$$L_j(t + \Delta t) = L_j(t) + \mathbf{1}[X_j(t) = 0] - \mathbf{1}[X_j(t) = 1].$$

This process is stochastic and independent across loci.
After evolving until total time T:

$$L_j^{\text{final}} = L_j(T), \quad \forall j.$$

This forms the observed microsatellite length distribution.

7. Inference Framework (SMC-ABC)

The SMC-ABC inference algorithm estimates parameters $\theta = (\mu_i, \mu_d)$ given observed microsatellite lengths L^{obs}.

Step 1: Priors

$$\mu_i \sim \text{Uniform}(a_i, b_i), \quad \mu_d \sim \text{Uniform}(a_d, b_d).$$

Step 2: Simulation and Distance Computation

For each particle $\theta_k = (\mu_{i,k}, \mu_{d,k})$:

$$L^{(k)} = f_{\text{model}}(\theta_k), \quad D_k = W_1(L^{(k)}, L^{\text{obs}}),$$

where W_1 denotes the 1D Wasserstein distance.

Step 3: Rejection Based on ε

Accept all particles where $D_k < \varepsilon$. The adaptive threshold updates as:

$$\varepsilon_{t+1} = \text{median}\{D_k\} \cdot \eta, \quad 0 < \eta < 1 \quad (\text{typically } \eta = 0.9).$$

Step 4: Weighting and Resampling

Accepted particles are weighted using an exponential kernel:

$$w_k \propto e^{-D_k/\varepsilon_t}, \quad \sum_k w_k = 1.$$

Resampling is performed with replacement according to these weights:

$$\theta_k' \sim \text{Categorical}(w_1, \ldots, w_N),$$

with Gaussian jitter added:

$$\theta_k' = \theta_k' + \mathcal{N}(0, \sigma^2), \quad \mu_i, \mu_d > 0.001.$$

Step 5: Iteration

Repeat for a fixed number of SMC iterations n_{iter}. Posterior distributions are approximated empirically from particles.
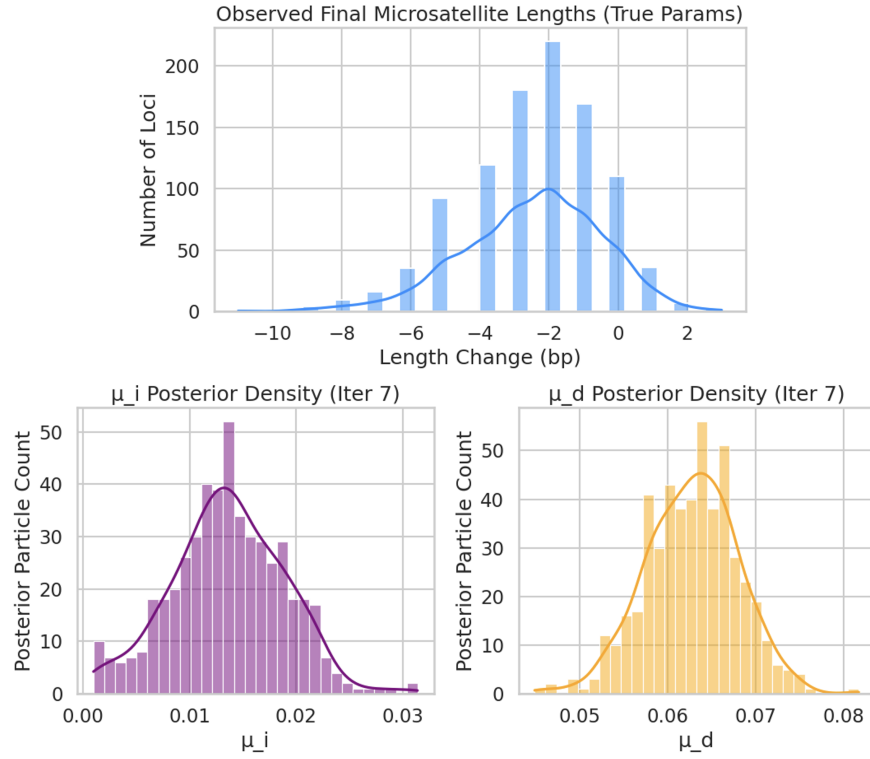
Results

Figure 1: True parameters: $\mu_i = 0.015, \mu_d = 0.085$

Then replaced $r * t$ with D_{eff} as number of effective cell division in the same model and infer that over uniform distribution along with other two parameters. Result of that is:

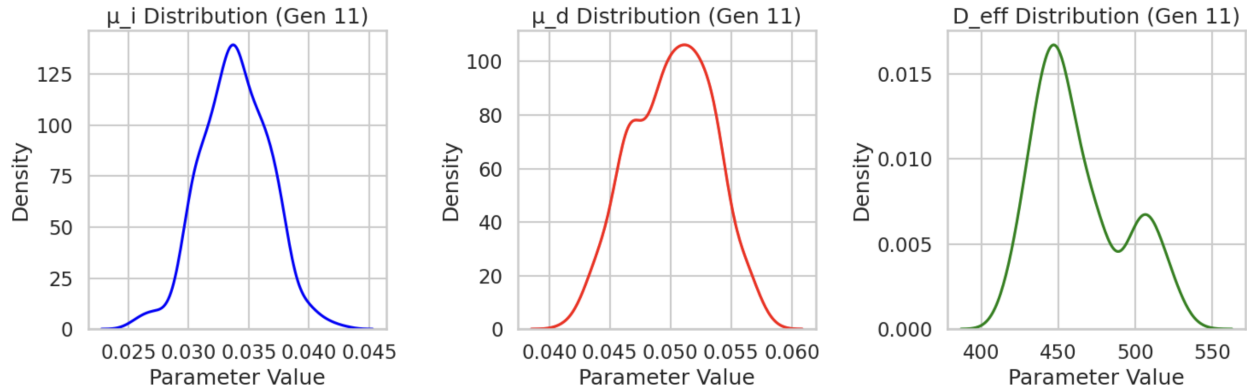Figure 2: Posterior Parameter Distributions of the Final Inference Iteration.
True: $\mu_i = 0.02, \mu_d = 0.03, D_{eff} = 750$ Estimated: $\mu_i = 0.0339, \mu_d = 0.0501, D_{eff} = 463.23$