Lecture 3. June 10 2019

Recap

- Last time, we described some of the mathematical underpinnings from population genetics, showed how they had influenced the field, and what they had led to
- The Ewens Sampling Formula paper (1972) started the field of inference in population genetics, and has many other uses in combinatorics and Bayesian non-parametric analysis
- We described an example from the world of computational cancer genomics, in which ABC could be used to attack what is in essence a combinatorics problem
- We alluded to the need for faster coalescent simulation approaches

Regression-based methods

Motivation – 1

The idea is to replace the hard cut-off in the ABC with a soft version that makes use of all of the observations.

Motivation – 1

The idea is to replace the hard cut-off in the ABC with a soft version that makes use of all of the observations.

In the summary statistic approach care has to be taken to choose $\rho(S, S')$.

For example, if $S = (S_1, \ldots, S_m)$ is an m-dimensional summary, and

$$\rho(S, S') = ||S' - S|| = \sqrt{\sum_{i=1}^{m} (S'_i - S_i)^2}$$

then accepting whenever $\rho \leq \epsilon$ treats the values of S' equally, regardless of the value of ρ.

Motivation – 1

The idea is to replace the hard cut-off in the ABC with a soft version that makes use of all of the observations.

In the summary statistic approach care has to be taken to choose $\rho(S, S')$.

For example, if $S = (S_1, \ldots, S_m)$ is an m-dimensional summary, and

$$\rho(S, S') = ||S' - S|| = \sqrt{\sum_{i=1}^{m} (S'_i - S_i)^2}$$

then accepting whenever $\rho \leq \epsilon$ treats the values of S' equally, regardless of the value of ρ. Beaumont et al (2002) added

- smooth weighting
- regression adjustment

The method is insensitive to the value of ϵ, and allows m to be large.

Motivation – 2

We start from M observations (θ_i, S_i), where each θ_i is an independent draw from the prior $\pi(\cdot)$ and S_i is the set of summary statistics generated when $\theta = \theta_i$.

We standardize the coordinates of S_i to have equal variances

Motivation – 2

We start from M observations (θ_i, S_i), where each θ_i is an independent draw from the prior $\pi(\cdot)$ and S_i is the set of summary statistics generated when $\theta = \theta_i$.

We standardize the coordinates of S_i to have equal variances

Now the posterior is

$$f(\theta|S) = \frac{f(\theta, S)}{f(S)}$$

so to estimate the left-hand side we could estimate the joint density and the marginal likelihood, and evaluate at $S = s_0$, the data.

Motivation – 3

The (θ_i, S_i) are a sample from the joint law, and the rejection method is just one way to estimate the conditional law when $S = s_0$; those with small values of $||S_i - s_0||$ are the ones to use.

Motivation – 3

The (θ_i, S_i) are a sample from the joint law, and the rejection method is just one way to estimate the conditional law when $S = s_0$; those with small values of $||S_i - s_0||$ are the ones to use.

This can be improved by

- weighting the θ_i according to $\rho(S_i, s_0)$
- adjusting the θ_i by local-linear regression

Motivation – 3

The (θ_i, S_i) are a sample from the joint law, and the rejection method is just one way to estimate the conditional law when $S = s_0$; those with small values of $||S_i - s_0||$ are the ones to use.

This can be improved by

- weighting the θ_i according to $\rho(S_i, s_0)$
- adjusting the θ_i by local-linear regression

Imagine that we have

$$\theta_i = m(s_i) + \epsilon := \alpha + (s_i - s_0)^T \beta + \epsilon_i, i = 1, 2, \ldots, M \tag{3}$$

where the ϵ_i are uncorrelated $(0, \sigma^2)$ rvs

Motivation – 4

When $s_i = s_0$, θ_i is drawn from a distribution with mean

$$\mathbb{E}(\theta|S = s_0) = \alpha$$

The least-squares estimator of (α, β) minimizes

$$\sum_{i=1}^{M}(\theta_i - \alpha - (s_i - s_0)^T\beta)^2$$

so that

$$(\hat{\alpha}, \hat{\beta}) = (X^T X)^{-1} X^T \theta,$$

where X is the design matrix

Motivation – 5

$$X = \begin{pmatrix} 1 & s_{11} - s_{01} & \dots & s_{1m} - s_{0m} \\ \vdots & \vdots & & \vdots \\ 1 & s_{M1} - s_{01} & \dots & s_{Mm} - s_{0m} \end{pmatrix}$$

Then, from (3)

$$\theta_i^* = \theta_i - (s_i - s_0)^T \hat{\beta}$$

form an approximate random sample from $f(\theta|s_0)$

Motivation – 5

$$X = \begin{pmatrix} 1 & s_{11} - s_{01} & \dots & s_{1m} - s_{0m} \\ \vdots & \vdots & & \vdots \\ 1 & s_{M1} - s_{01} & \dots & s_{Mm} - s_{0m} \end{pmatrix}$$

Then, from (3)

$$\theta_i^* = \theta_i - (s_i - s_0)^T \hat{\beta}$$

form an approximate random sample from $f(\theta|s_0)$

Note that $\mathbb{E}(\theta|s_0) = \hat{\alpha} = M^{-1} \sum \theta_i^*$

Regression method – 1

We can improve things by using weighted regression. Replace the minimization objective with

$$\sum_{i=1}^{M}(\theta_i - \alpha - (s_i - s_0)^T\beta)^2\, K_\epsilon(||s_i - s_0||) \tag{4}$$

One choice is the Epanechnikov kernel

$$K_\epsilon(t) = \frac{3}{2\epsilon}\left(1 - \left(\frac{t}{\epsilon}\right)^2\right)\ \mathbb{1}(t \leq \epsilon);$$

$\int_0^\epsilon K_\epsilon(t)dt = 1$. Now we get

$$(\hat{\alpha}, \hat{\beta}) = (X^T W X)^{-1} X^T W \theta,$$

Regression method – 2

where $W = \mathrm{diag}\{K_\epsilon(||s_i - s_0||)\}$
We now have

$$\mathbb{E}(\theta|s_0) = \hat{\alpha} = \frac{\sum \theta_i^* K_\epsilon(||s_i - s_0||)}{\sum K_\epsilon(||s_i - s_0||)}$$

Regression method – 2

where $W = \mathrm{diag}\{K_\epsilon(||s_i - s_0||)\}$
We now have

$$\mathbb{E}(\theta|s_0) = \hat{\alpha} = \frac{\sum \theta_i^* K_\epsilon(||s_i - s_0||)}{\sum K_\epsilon(||s_i - s_0||)}$$

Our weighted sample from the posterior is given by (θ_i^*, w_i), with

$$w_i = \frac{K_\epsilon(||s_i - s_0||)}{\sum K_\epsilon(||s_i - s_0||)}, i = 1, \ldots, M$$

Regression method – 2

where $W = \mathrm{diag}\{K_\epsilon(||s_i - s_0||)\}$
We now have

$$\mathbb{E}(\theta|s_0) = \hat{\alpha} = \frac{\sum \theta_i^* K_\epsilon(||s_i - s_0||)}{\sum K_\epsilon(||s_i - s_0||)}$$

Our weighted sample from the posterior is given by (θ_i^*, w_i), with

$$w_i = \frac{K_\epsilon(||s_i - s_0||)}{\sum K_\epsilon(||s_i - s_0||)}, i = 1, \ldots, M$$

This discussion concerns linear adjustment. Can generalize to other regression models by replacing $\alpha + (s_i - s_0)^T\beta$ in (4) by an appropriate $m(s_i)$.

Regression method – 3: Special cases

- For local-constant regression, we set $\beta = 0$
- Then if $K_\epsilon(\cdot)$ is replaced by

$$I_\epsilon(t) = \epsilon^{-1}\,\mathbb{1}(t \leq \epsilon)$$

get

$$\hat{\alpha} = \frac{\sum \theta_i I_\epsilon(||s_i - s_0||)}{\sum I_\epsilon(||s_i - s_0||)},$$

which is the rejection method estimate

Choice of ϵ

- Set ϵ to be a quantile, P_ϵ, of the empirical distribution of simulated values of $||s_i - s_0||$
- The choice of ϵ involves, as ever, a trade-off between bias and variance:
  - As $\epsilon \uparrow$, you use more observations, so less variance . . .
  - . . . but more bias

Note: as $\epsilon \downarrow 0$, both rejection and regression methods are equivalent

Implementation: the *abc* package in R implements these and many other methods.

Heteroscedastic regression

There is a literature on heteroscedastic models. The regression equation is

$$\theta = m(s) + \sigma(s)\xi$$

where $\sigma(s)$ is the square root of conditional variance of θ given s, and ξ is the residual. Using $\hat{}$ to denote estimator, the adjusted observations are

$$\begin{aligned} \theta_i^* &= \hat{m}(s_0) + \hat{\sigma}(s_0)\hat{\xi}_i \\ &= \hat{m}(s_0) + \frac{\hat{\sigma}(s_0)}{\hat{\sigma}(s_i)}(\theta_i - \hat{m}(s_i)) \end{aligned}$$

Reference: Blum MG (2018) Chapter 3 in *Handbook of Approximate Bayesian Computation*, CRC Press

Markov Chain Monte Carlo methods

MCMC – 1

The idea is to construct an ergodic Markov chain that has $f(\theta|\mathcal{D})$ as its stationary distribution, in the case that normalising constants cannot be computed. Here is Hastings' (*Biometrika*,1970) classic method:

1. Now at θ
2. Propose move to θ' according to $q(\theta \to \theta')$
3. Calculate the Hastings ratio

$$h = \min\left(1, \frac{\mathbb{P}(\mathcal{D} \mid \theta')\pi(\theta')q(\theta' \to \theta)}{\mathbb{P}(\mathcal{D} \mid \theta)\pi(\theta)q(\theta \to \theta')}\right)$$

4. Accept θ' with probability h, else return θ

MCMC – 2

There are more things to check:

- Is the chain ergodic?
- Does it mix well?
- Is the chain stationary?
- Burn in?
- Diagnostics of the run (no free lunches) – see `coda` package in R for example

MCMC – 3

MCMC in evolutionary genetics setting

- Small tweaks in the biology often translate into huge changes in algorithm
- Long development time
- All the usual problems with convergence
- *Almost all the effort goes into evaluation of likelihood*

ABC-MCMC – 1

Here is an ABC version (Marjoram et al, *PNAS*, 2003)

1. Now at θ
2. Propose a move to θ' according to $q(\theta \to \theta')$
3. Generate $\mathcal{D}'$ using θ'
4. If $\mathcal{D}' = \mathcal{D}$, go to next step, else return θ
5. Calculate

$$h = h(\theta, \theta') = \min\left(1, \frac{\pi(\theta')q(\theta' \to \theta)}{\pi(\theta)q(\theta \to \theta')}\right)$$

6. Accept θ' with probability h, else return θ

ABC-MCMC – 2

Lemma: The stationary distribution of the chain is, indeed, $f(\theta|\mathcal{D})$.

Proof: In class . . .

ABC-MCMC – 3

Here is the practical version, for data $\mathcal{D}$, summary statistics S

4′. If $\rho(\mathcal{D}', \mathcal{D}) \leq \epsilon$, go to next step, otherwise return θ

ABC-MCMC – 3

Here is the practical version, for data $\mathcal{D}$, summary statistics S

4′. If $\rho(\mathcal{D}', \mathcal{D}) \leq \epsilon$, go to next step, otherwise return θ

4″. If $\rho(S', S) \leq \epsilon$, go to next step, otherwise return θ

for some suitable metric ρ and approximation level ϵ

ABC-MCMC – 3

Here is the practical version, for data $\mathcal{D}$, summary statistics S

4′. If $\rho(\mathcal{D}', \mathcal{D}) \leq \epsilon$, go to next step, otherwise return θ

4″. If $\rho(S', S) \leq \epsilon$, go to next step, otherwise return θ

for some suitable metric ρ and approximation level ϵ

Observations now from $f(\theta \mid \rho(\mathcal{D}', \mathcal{D}) \leq \epsilon)$ or $f(\theta \mid \rho(S', S) \leq \epsilon)$

Variations on a theme – 1

There have been many variants on the theme. For example, one might use multiple simulations from a given θ to get a better estimate of the likelihood. This is known as the pseudo-marginal method (Beaumont, *Genetics*, 2003; Tavaré et al. *PNAS*, 2003; Andrieu & Roberts *Ann Statist*, 2009). The idea is to simulate pairs of data points $(\theta, \hat{\mathbb{P}}(\mathcal{D}|\theta))$:

2′. If at θ' simulate B values of $\mathcal{D}'$, and use these to estimate $\mathbb{P}(\mathcal{D}|\theta')$ via

$$\hat{\mathbb{P}}(\mathcal{D}|\theta') = \frac{1}{B}\sum_{j=1}^{B} \mathbb{1}(\mathcal{D}' = \mathcal{D})$$

Variations on a theme – 1

There have been many variants on the theme. For example, one might use multiple simulations from a given θ to get a better estimate of the likelihood. This is known as the pseudo-marginal method (Beaumont, *Genetics*, 2003; Tavaré et al. *PNAS*, 2003; Andrieu & Roberts *Ann Statist*, 2009). The idea is to simulate pairs of data points $(\theta, \hat{\mathbb{P}}(\mathcal{D}|\theta))$:

2′. If at θ' simulate B values of $\mathcal{D}'$, and use these to estimate $\mathbb{P}(\mathcal{D}|\theta')$ via

$$\hat{\mathbb{P}}(\mathcal{D}|\theta') = \frac{1}{B}\sum_{j=1}^{B} \mathbb{1}(\mathcal{D}' = \mathcal{D})$$

3′. If this is 0, stay at θ; else

Variations on a theme – 2

4′. Accept θ' and $\hat{\mathbb{P}}(\mathcal{D}|\theta')$ with probability

$$h = \min\left(1, \frac{\hat{\mathbb{P}}(\mathcal{D} \mid \theta')\pi(\theta')q(\theta' \to \theta)}{\hat{\mathbb{P}}(\mathcal{D} \mid \theta)\pi(\theta)q(\theta \to \theta')}\right)$$

else stay at θ.

Variations on a theme – 3

- Convergence an issue?
- These methods can often be started at stationarity, so no burn-in
- If the underlying probability model is complex, simulating data will often not lead to acceptance. Thus need update for parts of the probability model (data augmentation)
- There are versions with varying ϵ; see Bortot et al (*JASA*, 2007) for example
- There are now many hybrid versions of these approaches (e.g. ABC-within-Gibbs)

Variations on a theme – 4

Bortot et al (*JASA*, 2007) have a nice way to include ϵ in the ABC-MCMC approach

1. Now at (θ, ϵ)

Variations on a theme – 4

Bortot et al (*JASA*, 2007) have a nice way to include ϵ in the ABC-MCMC approach

1. Now at (θ, ϵ)
2. Propose a move to (θ', ϵ') according to $q((\theta, \epsilon) \rightarrow (\theta', \epsilon'))$

Variations on a theme – 4

Bortot et al (*JASA*, 2007) have a nice way to include ϵ in the ABC-MCMC approach

1. Now at (θ, ϵ)
2. Propose a move to (θ', ϵ') according to $q((\theta, \epsilon) \rightarrow (\theta', \epsilon'))$
3. Generate $\mathcal{D}'$ from model with parameters θ'

Variations on a theme – 4

Bortot et al (*JASA*, 2007) have a nice way to include ϵ in the ABC-MCMC approach

1. Now at (θ, ϵ)
2. Propose a move to (θ', ϵ') according to $q((\theta, \epsilon) \to (\theta', \epsilon'))$
3. Generate $\mathcal{D}'$ from model with parameters θ'
4. Calculate

$$h = \min\left(1, \frac{\pi(\theta')\pi(\epsilon')q((\theta', \epsilon') \to (\theta, \epsilon))}{\pi(\theta)\pi(\epsilon)q((\theta, \epsilon) \to (\theta', \epsilon'))} \mathbb{1}(\rho(S', S) \leq \epsilon)\right)$$

Variations on a theme – 4

Bortot et al (*JASA*, 2007) have a nice way to include ϵ in the ABC-MCMC approach

1. Now at (θ, ϵ)
2. Propose a move to (θ', ϵ') according to $q((\theta, \epsilon) \to (\theta', \epsilon'))$
3. Generate $\mathcal{D}'$ from model with parameters θ'
4. Calculate

$$h = \min\left(1, \frac{\pi(\theta')\pi(\epsilon')q((\theta', \epsilon') \to (\theta, \epsilon))}{\pi(\theta)\pi(\epsilon)q((\theta, \epsilon) \to (\theta', \epsilon'))} \mathbb{1}(\rho(S', S) \leq \epsilon)\right)$$

5. Accept (θ', ϵ') with probability h, else return (θ, ϵ)

Variations on a theme – 5

- The idea is to run the chain with typical values of ϵ being small

Variations on a theme – 5

- The idea is to run the chain with typical values of ϵ being small
- Filter the series $\{(\theta_i, \epsilon_i)\}$ by restricting to $\{i : \epsilon_i < \epsilon_T\}$ after accepting values from the chain

Variations on a theme – 5

- The idea is to run the chain with typical values of ϵ being small
- Filter the series $\{(\theta_i, \epsilon_i)\}$ by restricting to $\{i : \epsilon_i < \epsilon_T\}$ after accepting values from the chain

  These values provide an estimate of $f(\theta|\mathcal{D})$ from values of $f(\theta|\rho \leq \epsilon_T)$ with weights given by $\pi(\epsilon)$

Variations on a theme – 5

- The idea is to run the chain with typical values of ϵ being small
- Filter the series $\{(\theta_i, \epsilon_i)\}$ by restricting to $\{i : \epsilon_i < \epsilon_T\}$ after accepting values from the chain

  These values provide an estimate of $f(\theta|\mathcal{D})$ from values of $f(\theta|\rho \leq \epsilon_T)$ with weights given by $\pi(\epsilon)$
- The authors use priors of the form $\epsilon \sim \mathrm{Exp}(\tau)$

Sequential Monte Carlo methods

SMC – 1

Repetitive sampling from the prior does not seem sensible, and various approaches have been designed to deal with this. Here is a version from Beaumont et al (*Biometrika*, 2009)

SMC – 1

Repetitive sampling from the prior does not seem sensible, and various approaches have been designed to deal with this. Here is a version from Beaumont et al (*Biometrika*, 2009)

The aim is to

- perform weighted resampling of the points already drawn
- shrink ϵ as you go

SMC – 1

Repetitive sampling from the prior does not seem sensible, and various approaches have been designed to deal with this. Here is a version from Beaumont et al (*Biometrika*, 2009)

The aim is to

- perform weighted resampling of the points already drawn
- shrink ϵ as you go

In what follows,

- $\mathcal{D}_0$ are the observed data
- $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_T$ are given

SMC – 2

1. For iteration $t = 1$,

SMC – 2

1. For iteration $t = 1$,
    1 For $i = 1, 2, \ldots, N$

        Simulate $\theta_i^{(1)} \sim \pi(\theta)$ and $\mathcal{D}$ from the model with parameter $\theta_i^{(1)}$ until $\rho(\mathcal{D}, \mathcal{D}_0) < \epsilon_1$

SMC – 2

1. For iteration $t = 1$,
    1 For $i = 1, 2, \ldots, N$

        Simulate $\theta_i^{(1)} \sim \pi(\theta)$ and $\mathcal{D}$ from the model with parameter $\theta_i^{(1)}$ until $\rho(\mathcal{D}, \mathcal{D}_0) < \epsilon_1$

        Set $w_i^{(1)} = 1/N$

SMC – 2

1. For iteration $t = 1$,

   1 For $i = 1, 2, \ldots, N$

   Simulate $\theta_i^{(1)} \sim \pi(\theta)$ and $\mathcal{D}$ from the model with parameter $\theta_i^{(1)}$ until $\rho(\mathcal{D}, \mathcal{D}_0) < \epsilon_1$

   Set $w_i^{(1)} = 1/N$

   Calculate $\tau_2^2 =$ twice the empirical variance of the $\{\theta_i^{(1)}\}$

2 For iteration $t = 2, \ldots, T$,

2 For iteration $t = 2, \ldots, T$,

For $i = 1, 2, \ldots, N$, repeat

Choose θ_i^* from the $\theta_j^{(t-1)}$s with probability $w_j^{(t-1)}$

Generate $\theta_i^{(t)} \sim \mathrm{N}(\theta_i^*, \tau_t^2)$, and $\mathcal{D}$ from the model with parameter $\theta_i^{(t)}$

until $\rho(\mathcal{D}, \mathcal{D}_0) < \epsilon_t$

2 For iteration $t = 2, \ldots, T$,

For $i = 1, 2, \ldots, N$, repeat

Choose θ_i^* from the $\theta_j^{(t-1)}$s with probability $w_j^{(t-1)}$

Generate $\theta_i^{(t)} \sim \mathrm{N}(\theta_i^*, \tau_t^2)$, and $\mathcal{D}$ from the model with parameter $\theta_i^{(t)}$

until $\rho(\mathcal{D}, \mathcal{D}_0) < \epsilon_t$

Set

$$w_i^{(t)} \propto \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^{N} w_j^{(t-1)} \phi(\tau_t^{-1}(\theta_i^{(t)} - \theta_j^{(t-1)})}$$

2 For iteration $t = 2, \ldots, T$,

For $i = 1, 2, \ldots, N$, repeat

Choose θ_i^* from the $\theta_j^{(t-1)}$s with probability $w_j^{(t-1)}$

Generate $\theta_i^{(t)} \sim \mathrm{N}(\theta_i^*, \tau_t^2)$, and $\mathcal{D}$ from the model with parameter $\theta_i^{(t)}$

until $\rho(\mathcal{D}, \mathcal{D}_0) < \epsilon_t$

Set

$$w_i^{(t)} \propto \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^{N} w_j^{(t-1)} \phi(\tau_t^{-1}(\theta_i^{(t)} - \theta_j^{(t-1)})}$$

Calculate $\tau_{t+1}^2 =$ twice the empirical variance of the $\{\theta_i^{(t)}\}$

SMC – 4

In the previous slide, $\mathrm{N}(\mu, \sigma^2)$ denotes the Normal distribution with mean μ and variance σ^2, and $\phi(x)$ is the standard Normal density function.

Notes:

- Equally well, we can replace $\mathcal{D}$ and $\mathcal{D}_0$ with $S(\mathcal{D})$ and $S(\mathcal{D}_0)$

SMC – 4

In the previous slide, $\mathrm{N}(\mu, \sigma^2)$ denotes the Normal distribution with mean μ and variance σ^2, and $\phi(x)$ is the standard Normal density function.

Notes:

- Equally well, we can replace $\mathcal{D}$ and $\mathcal{D}_0$ with $S(\mathcal{D})$ and $S(\mathcal{D}_0)$
- Can also replace the step

$$\theta_i^{(t)} \sim \mathrm{N}(\theta_i^*, \tau_t^2)$$

with

$$\theta_i^{(t)} \sim \mathrm{K}(\cdot|\theta_i^*, \tau_t^2)$$

where K need not be a Gaussian kernel, but could be a t distribution for example

SMC – 5: Rationale

- In step [1] we simulate from the prior and keep the N closest points (so this is rejection-ABC). This set of points is drawn (roughly) from the posterior

SMC – 5: Rationale

- In step [1] we simulate from the prior and keep the N closest points (so this is rejection-ABC). This set of points is drawn (roughly) from the posterior
- Next, fit a density with kernel K

$$q(\theta) := \sum_{j=1}^{N} w_j^{(t-1)} \, K(\theta | \theta_j^{(t-1)}, \tau_t^2)$$

around the points, and resample values from this

SMC – 5: Rationale

- In step [1] we simulate from the prior and keep the N closest points (so this is rejection-ABC). This set of points is drawn (roughly) from the posterior
- Next, fit a density with kernel K

$$q(\theta) := \sum_{j=1}^{N} w_j^{(t-1)} K(\theta|\theta_j^{(t-1)}, \tau_t^2)$$

  around the points, and resample values from this
- then reduce the tolerance, ϵ

SMC – 5: Rationale

- In step [1] we simulate from the prior and keep the N closest points (so this is rejection-ABC). This set of points is drawn (roughly) from the posterior
- Next, fit a density with kernel K

$$q(\theta) := \sum_{j=1}^{N} w_j^{(t-1)} \, K(\theta | \theta_j^{(t-1)}, \tau_t^2)$$

  around the points, and resample values from this
- then reduce the tolerance, ϵ
- simulate data and their summary statistics

SMC – 5: Rationale

- In step [1] we simulate from the prior and keep the N closest points (so this is rejection-ABC). This set of points is drawn (roughly) from the posterior
- Next, fit a density with kernel K

$$q(\theta) := \sum_{j=1}^{N} w_j^{(t-1)} \, K(\theta | \theta_j^{(t-1)}, \tau_t^2)$$

  around the points, and resample values from this
- then reduce the tolerance, ϵ
- simulate data and their summary statistics
- weight each point by $\pi(\cdot)/q(\cdot)$ to allow for the fact that the points are not samples from the prior