Bayesian Optimisation for Likelihood Free Inference

Make model parameterisation go brrr

Jacob Cumming

University of Melbourne

April 2024

Notation

- Model is considered a (random) function $f(\boldsymbol{\theta})$ that maps $\boldsymbol{\theta}$ (a vector of parameters) to a model output, that can be transformed into $\mathbf{X}$, that has the same shape as:
- $\mathbf{X}_{\text{obs}}$, a vector of outputs given to us usually in the forms of summary statistics (incidence, prevalence, hospitalisations etc).

Parameter inference would become easy if we had...

- An explicit form for the likelihood: $\mathcal{L}(\theta|\mathbf{X}_{\text{obs}}) := \Pr(\mathbf{X}_{\text{obs}}|\theta)$

Parameter inference would become easy if we had...

- An explicit form for the likelihood: $\mathcal{L}(\theta|\mathbf{X}_{\text{obs}}) := \Pr(\mathbf{X}_{\text{obs}}|\theta)$
- Or even $\mathcal{L}(\boldsymbol{\theta}|S(\mathbf{X}_{\text{obs}})) := \Pr(S(\mathbf{X}_{\text{obs}})|\boldsymbol{\theta})$, where $S(\mathbf{X}_{\text{obs}})$ is a (vector of) summary statistic(s)

Parameter inference would become easy if we had...

- An explicit form for the likelihood: $\mathcal{L}(\theta|\mathbf{X}_{\text{obs}}) := \Pr(\mathbf{X}_{\text{obs}}|\theta)$
- Or even $\mathcal{L}(\boldsymbol{\theta}|S(\mathbf{X}_{\text{obs}})) := \Pr(S(\mathbf{X}_{\text{obs}})|\boldsymbol{\theta})$, where $S(\mathbf{X}_{\text{obs}})$ is a (vector of) summary statistic(s)
- $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|S(\mathbf{X}_{\text{obs}}))$

Parameter inference would become easy if we had...

- An explicit form for the likelihood: $\mathcal{L}(\theta|\mathbf{X}_{\text{obs}}) := \Pr(\mathbf{X}_{\text{obs}}|\theta)$
- Or even $\mathcal{L}(\boldsymbol{\theta}|S(\mathbf{X}_{\text{obs}})) := \Pr(S(\mathbf{X}_{\text{obs}})|\boldsymbol{\theta})$, where $S(\mathbf{X}_{\text{obs}})$ is a (vector of) summary statistic(s)
- $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|S(\mathbf{X}_{\text{obs}}))$
- $\Pr(\boldsymbol{\theta}|S(\mathbf{X}_{\text{obs}})) \propto \Pr(S(\mathbf{X}_{\text{obs}})|\boldsymbol{\theta}) \Pr(\boldsymbol{\theta})$

The Sad Truth

- As models become more complicated, explicit likelihoods don't exist (think agent based models).

A Standard Bayesian Solution

- Approximate Bayesian Computation (ABC)
  1. Sample from prior
  2. Run model
  3. Accept or reject parameters run based on how well $\mathbf{X}$ 'matches' $\mathbf{X}_{\text{obs}}$.

What is 'matches'

- Discrepency function $D : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
  - Can be a norm such as $||S(\mathbf{X}) - S(\mathbf{X}_{\text{obs}})||_p := (\sum_{i=1}^{d} |S(\mathbf{X}) - S(\mathbf{X}_{\text{obs}})|^p)^{1/p}$

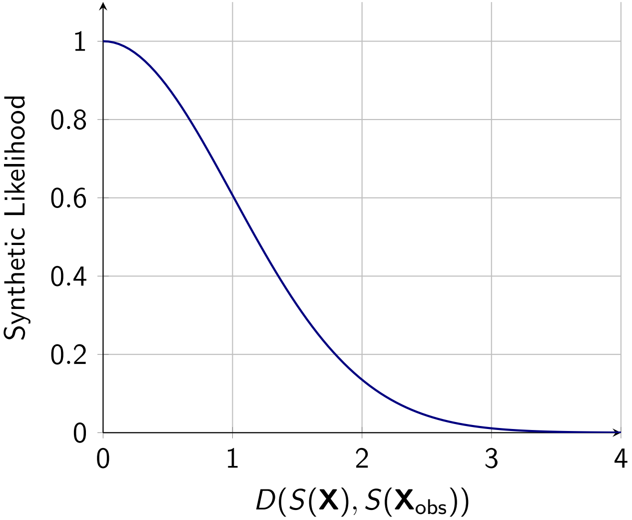
What is 'matches'

- Discrepency function $D : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
  - Can be a norm such as $||S(\mathbf{X}) - S(\mathbf{X}_{\text{obs}})||_p := (\sum_{i=1}^{d} |S(\mathbf{X}) - S(\mathbf{X}_{\text{obs}})|^p)^{1/p}$
  - Care should be taken to rescale $S(\mathbf{X}_{\text{obs}})$ and $S(\mathbf{X})$ appropriately (ie via a covariance matrix).

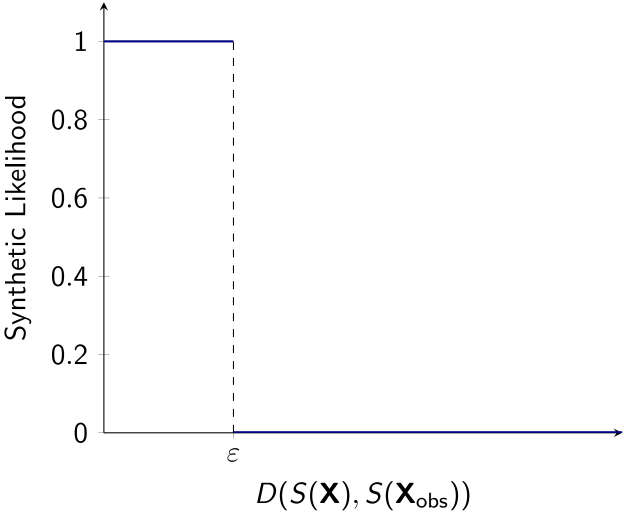
What is 'matches'

- Discrepency function $D : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$
  - Can be a norm such as $||S(\mathbf{X}) - S(\mathbf{X}_{\text{obs}})||_p := (\sum_{i=1}^{d} |S(\mathbf{X}) - S(\mathbf{X}_{\text{obs}})|^p)^{1/p}$
  - Care should be taken to rescale $S(\mathbf{X}_{\text{obs}})$ and $S(\mathbf{X})$ appropriately (ie via a covariance matrix).
- $D(S(\mathbf{X}), S(\mathbf{X}_{\text{obs}}))$, gives acceptance probability of $\boldsymbol{\theta}$.

Acceptance Probability



Attempt 2



Synthetic Likelihood
1
0.8
0.6
0.4
0.2
0
ε
$D(S(\mathbf{X}), S(\mathbf{X}_{\text{obs}}))$

Overall Idea of my Research

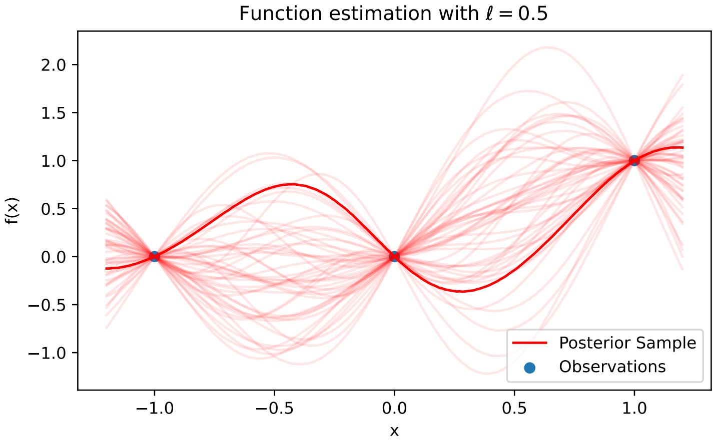
- What if we could 'predict' discrepency values we hadn't seen before?

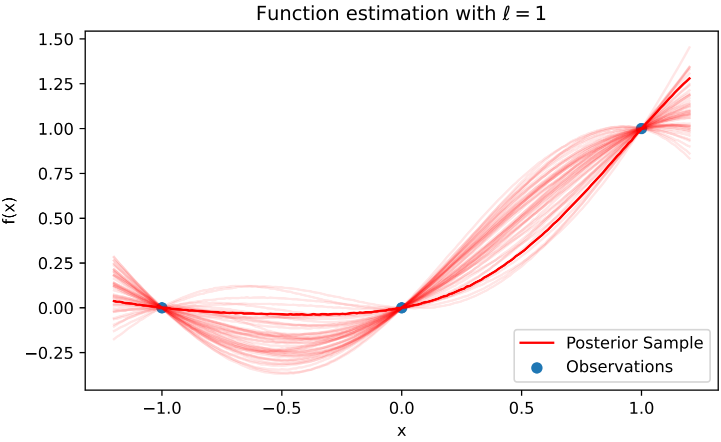
Overall Idea of my Research

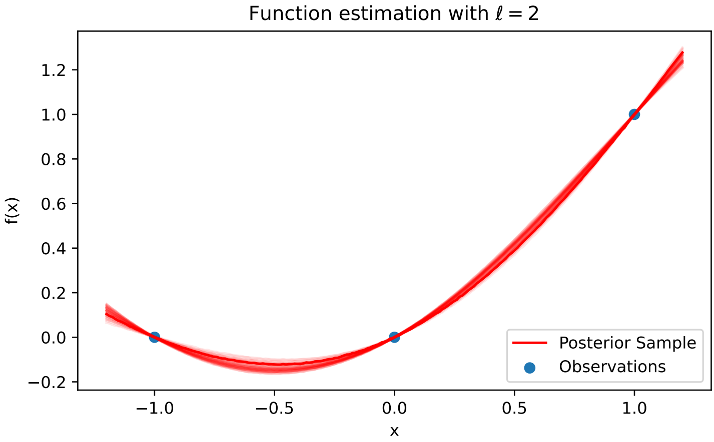
- What if we could 'predict' discrepency values we hadn't seen before?
- For parameters 'close' to parameters we've already tried it should be easy.

Gaussian Processes

- A class of random functions
- Common examples - Brownian motion, Ornstein Uhlenbeck process
- Model the mean discrepency using one of these (kriging)


Function estimation with $\ell = 0.5$
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
f(x)
−1.0
−0.5
0.0
0.5
1.0
x
Posterior Sample
Observations


Function estimation with $\ell = 1$
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
f(x)
−1.0
−0.5
0.0
0.5
1.0
x
Posterior Sample
Observations


Function estimation with $\ell = 2$
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
f(x)
−1.0
−0.5
0.0
0.5
1.0
x
Posterior Sample
Observations

Sample frame title