Weekly Report (3.1-3.7)*

Yanpei Cai†

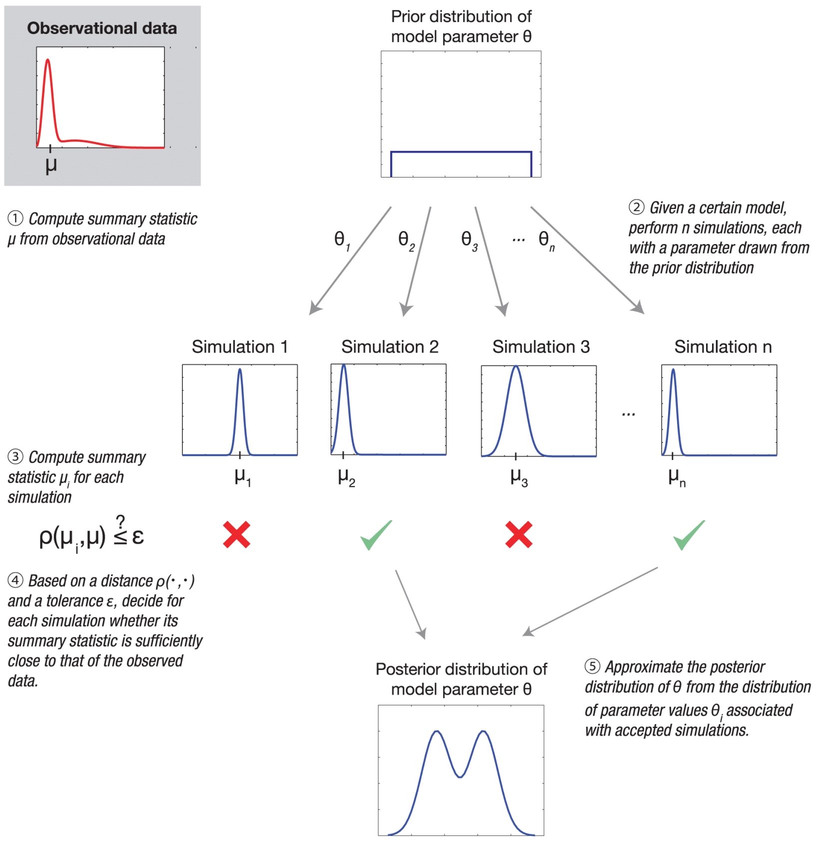
March 7, 2025

During the past week, I continued to read Michael's paper but spend most of the time on improving understanding of the background knowledge (especially through his course Probabilistic Modelling and Reasoning). The following is a summary of the work that has been done during the past week and some possible measures for the plan for next week.

Summary

Simulator-based Statistical Models

- The probability density (mass) function $p_{\boldsymbol{y}|\boldsymbol{\theta}}$ is defined implicitly by the simulator (more details to be added).



Question: Which step corresponds to the computation in Bayesian Inference?

†Artificial Intelligence and its Applications Institute, School of Informatics, The University of Edinburgh
Email Address: yanpei.cai@ed.ac.uk
*This is a summary of the works was done in the past week.

Likelihood-free Inference of Simulator-based Statistical Models

As per introduced in the last section, likelihood-free inference (LFI) refers to a set of statistical methods used to estimate parameters of a model without explicitly computing the likelihood function. This is particularly useful when the likelihood is intractable (p.s. in agent-based models, the likelihood function too complex to compute analytically). Instead of explicitly evaluating the likelihood, LFI methods use simulations to compare observed data with simulated data generated from the model.

General Strategy

- **Simulation-Based Approach**: Since the likelihood function is intractable, LFI generates synthetic data from the model using different parameter values.
- **Comparison to Observed Data**: The simulated data is compared to the observed data. Discrepancy is obtained directly/indirectly (e.g. distance metrics, summary statistics, etc).
- **Parameter Estimation**: The best-fitting parameters are inferred by implementing different metrics, such as rejection sampling, approximate Bayesian Computation (ABC), Kernel Density Estimation, etc.

Exact Inference v.s. Approximate Inference

- **Exact Inference**: Posterior distribution is obtained by the retained simulated data via rejection sampling. **NOTE:** Only valid for discrete random variables. (more details to be added)
  IDEA: The posterior is obtained by conditioning $p_{\boldsymbol{\theta}|\boldsymbol{y}}(\boldsymbol{\theta}|\boldsymbol{y}_o)$ on $\boldsymbol{y} = \boldsymbol{y}_o$, as

  $$p_{\boldsymbol{\theta}|\boldsymbol{y}}(\boldsymbol{\theta}|\boldsymbol{y}_o) = \frac{p_{\boldsymbol{y}|\boldsymbol{\theta}}(\boldsymbol{y}_o|\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{p_{\boldsymbol{y}}(\boldsymbol{y}_o)} = \frac{p_{\boldsymbol{y}|\boldsymbol{\theta}}(\boldsymbol{y}_o \wedge \boldsymbol{\theta})}{p_{\boldsymbol{y}}(\boldsymbol{y}_o)}$$

  Given tuples $(\boldsymbol{\theta}_i, \boldsymbol{y}_i)$ where

  - $\boldsymbol{\theta}_i \sim p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ (iid from the prior)
  - $\boldsymbol{y}_i = g(\omega_i, \boldsymbol{\theta}_i)$ (obtained by running the simulator),

  retain only those where $\boldsymbol{y}_i = \boldsymbol{y}_o$. The $\boldsymbol{\theta}_i$ from the retained tuples are samples from the posterior $p_{\boldsymbol{\theta}|\boldsymbol{y}}(\boldsymbol{\theta}|\boldsymbol{y}_o)$.

  Approximate Inference:

  - Reduce the dimensionality of the data to some features or summary statistics.
  - Allow the discrepancy between the simulated data and the observed data to be small enough instead of to be exactly zero. **REASON:** The probability of the discrepancy to be zero can becomes smaller and smaller as the dimension of the data increases, the posterior is therefore not well-described by the accepted samples.

Examples

- **Approximate Bayesian Computation (ABC)**: e.g. Rejection ABC (more details to be added)
- **Synthetic Likelihood**: Synthetic likelihood function is obtained using a Gaussian approximation on summary statistics (e.g. if the summary statistics is obtained via averaging then the likelihood function is approximated Gaussian via CLT). It explicitly estimates a likelihood but in a simplified manner.
- **Kernal Density Estimation**: (more details to be added)

Discussion

There are few difficulties of generalising likelihood-free inference:

- Assessment of the discrepancy between the simulated data and the observed data.
- Less efficiency due to the simulations for large datasets.
- Lack of knowledge about the relation between discrepancy and model parameters

Formal Definition of a Kernel Function

In non-parametric statistics, a **kernel function** $K : \mathbb{R} \to \mathbb{R}$ is a real-valued function that satisfies the following properties:

- **Integrability (Normalization Condition):**

  $$\int_{-\infty}^{\infty} K(x)\, dx = 1$$

  This ensures that the kernel properly normalizes density estimates.

- **Symmetry:**

  $$K(x) = K(-x), \quad \forall x \in \mathbb{R}$$

  This ensures that the function is centered and does not introduce bias.

- **Non-Negativity (for many kernels, but not always required):**

  $$K(x) \geq 0, \quad \forall x \in \mathbb{R}$$

  Some kernels (e.g., certain wavelet kernels) may take negative values, but most commonly used kernels are non-negative.

- **Finite Variance (Second-Order Moment Condition):**

  $$\int_{-\infty}^{\infty} x^2 K(x)\, dx < \infty$$

  This ensures that the kernel has finite spread.

A kernel is often used in a **scaled** form, incorporating a bandwidth parameter h, which controls the level of smoothness:

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$$

where $h > 0$ is the bandwidth.

Kernel Density Estimation (KDE)

Given a sample $X_1, X_2, \ldots, X_n$, the **kernel density estimator** (Parzen-Rosenblatt estimator) is given by:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where

- $\hat{f}(x)$ is the estimated probability density function (PDF),
- h is the bandwidth parameter,
- $K(\cdot)$ is the kernel function.