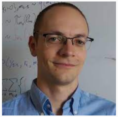
Robust Bayesian Inference for Simulator-based Models via the MMD Posterior Bootstrap

Charita Dellaporta [1], Jeremias Knoblauch [2],
Theodoros Damoulas [1], François-Xavier Briol[2]



Workshop on Functional Inference and Machine Intelligence (FIMI) , March 2024

[1]University of Warwick, [2]University College London

Outline

1. Motivation
    - Simulator-based models
    - Problem Setting
2. Background
    - Bayesian Nonparametric Learning (NPL)
3. Method
    - MMD posterior bootstrap
4. Theoretical Results
    - Generalisation Error
    - Frequentist consistency
    - Outliers robustness
5. Applications
6. Discussion & current work

Motivation

Simulator-based models

- Independent sampling is possible, but the likelihood is unavailable
- Model is usually at best a rough approximation of a complex physical or biological phenomenon
- It will most likely not capture all of the key characteristics of the underlying data generating process.

Two main problems:

1. Unavailability of the likelihood function
2. Model misspecification

Simulator-based models

- Independent sampling is possible, but the likelihood is unavailable
- Model is usually at best a rough approximation of a complex physical or biological phenomenon
- It will most likely not capture all of the key characteristics of the underlying data generating process.

Two main problems:

1. Unavailability of the likelihood function
2. Model misspecification

Simulator-based models

- Independent sampling is possible, but the likelihood is unavailable
- Model is usually at best a rough approximation of a complex physical or biological phenomenon
- It will most likely not capture all of the key characteristics of the underlying data generating process.

Two main problems:

1. Unavailability of the likelihood function
2. Model misspecification

Simulator-based models

- Independent sampling is possible, but the likelihood is unavailable
- Model is usually at best a rough approximation of a complex physical or biological phenomenon
- It will most likely not capture all of the key characteristics of the underlying data generating process.

Two main problems:

1. Unavailability of the likelihood function
2. Model misspecification

Problem setting

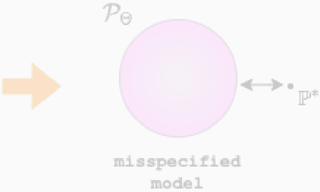
Simulator-based model family

$\boxed{\mathcal{P}_\Theta = \{\mathbb{P}_\theta : \theta \in \Theta\} \subseteq \mathcal{P}}$

$\mathbb{P}_\theta$ is associated with *simulator* function $G_\theta : \mathcal{U} \to \mathcal{X}$ and probability measure $\mathbb{U}$ in space $\mathcal{U}$

$$y := G_\theta(u) \sim \mathbb{P}_\theta, u \sim \mathbb{U}.$$

Observed i.i.d. data $x_{1:n} \sim \mathbb{P}^*$



Problem setting

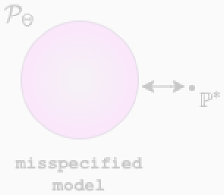
Simulator-based model family

$$\boxed{\mathcal{P}_\Theta = \{\mathbb{P}_\theta : \theta \in \Theta\} \subseteq \mathcal{P}}$$

$\mathbb{P}_\theta$ is associated with *simulator* function $G_\theta : \mathcal{U} \to \mathcal{X}$ and probability measure $\mathbb{U}$ in space $\mathcal{U}$

$$y := G_\theta(u) \sim \mathbb{P}_\theta, u \sim \mathbb{U}.$$

Observed i.i.d. data $x_{1:n} \sim \mathbb{P}^*$



Problem setting

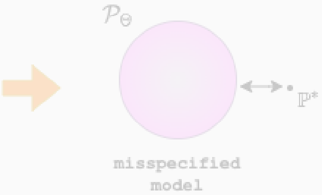
Simulator-based model family

$$\boxed{\mathcal{P}_\Theta = \{\mathbb{P}_\theta : \theta \in \Theta\} \subseteq \mathcal{P}}$$

$\mathbb{P}_\theta$ is associated with *simulator* function $G_\theta : \mathcal{U} \to \mathcal{X}$ and probability measure $\mathbb{U}$ in space $\mathcal{U}$

$$y := G_\theta(u) \sim \mathbb{P}_\theta, u \sim \mathbb{U}.$$

Observed i.i.d. data $x_{1:n} \sim \mathbb{P}^*$



Problem setting

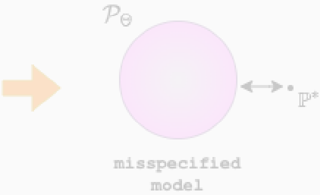
Simulator-based model family

$$\boxed{\mathcal{P}_\Theta = \{\mathbb{P}_\theta : \theta \in \Theta\} \subseteq \mathcal{P}}$$

$\mathbb{P}_\theta$ is associated with *simulator* function $G_\theta : \mathcal{U} \to \mathcal{X}$ and probability measure $\mathbb{U}$ in space $\mathcal{U}$

$$y := G_\theta(u) \sim \mathbb{P}_\theta, u \sim \mathbb{U}.$$

Observed i.i.d. data $x_{1:n} \sim \mathbb{P}^*$



Problem setting

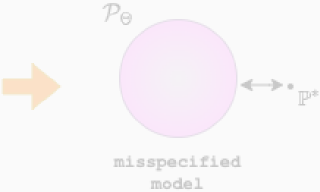
Simulator-based model family

$$\boxed{\mathcal{P}_\Theta = \{\mathbb{P}_\theta : \theta \in \Theta\} \subseteq \mathcal{P}}$$

$\mathbb{P}_\theta$ is associated with *simulator* function $G_\theta : \mathcal{U} \to \mathcal{X}$ and probability measure $\mathbb{U}$ in space $\mathcal{U}$

$$y := G_\theta(u) \sim \mathbb{P}_\theta, u \sim \mathbb{U}.$$

Observed i.i.d. data $x_{1:n} \sim \mathbb{P}^*$



Problem setting

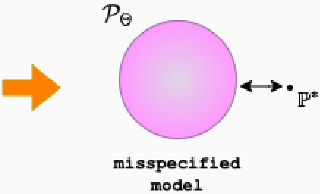
Simulator-based model family

$$\boxed{\mathcal{P}_\Theta = \{\mathbb{P}_\theta : \theta \in \Theta\} \subseteq \mathcal{P}}$$

$\mathbb{P}_\theta$ is associated with *simulator* function $G_\theta : \mathcal{U} \to \mathcal{X}$ and probability measure $\mathbb{U}$ in space $\mathcal{U}$

$$y := G_\theta(u) \sim \mathbb{P}_\theta, u \sim \mathbb{U}.$$

Observed i.i.d. data $x_{1:n} \sim \mathbb{P}^*$



Robust LFI

- Bayesian inference via uncertainty coming from the data generating distribution while preserving robustness
  1. Approximate Bayesian Computation (ABC) for ex. Bernton et al. (2019)
  2. Generalized Bayesian Inference (GBI) for ex. Pacchiardi and Dutta (2021), Schmon et al. (2020)

Robust LFI

- Bayesian inference via uncertainty coming from the data generating distribution while preserving robustness

    1. Approximate Bayesian Computation (ABC) for ex. Bernton et al. (2019)
    2. Generalized Bayesian Inference (GBI) for ex. Pacchiardi and Dutta (2021), Schmon et al. (2020)

Robust LFI

- Bayesian inference via uncertainty coming from the data generating distribution while preserving robustness

  1. Approximate Bayesian Computation (ABC) for ex. Bernton et al. (2019)
  2. Generalized Bayesian Inference (GBI) for ex. Pacchiardi and Dutta (2021), Schmon et al. (2020)

Background

Bayesian Nonparametric Learning (NPL) framework (Lyddon et al., 2018; Fong et al., 2019)

1. Place a nonparametric prior *directly* on the data-generating mechanism $\mathbb{P}^*$:

$$\mathbb{P} \sim DP(\alpha, \mathbb{F}), \quad \mathbb{P}|x_{1:n} \sim DP\left(\alpha', \mathbb{F}'\right)$$

   where

$$\alpha' = \alpha + n, \quad \mathbb{F}' := \frac{\alpha}{\alpha+n}\mathbb{F} + \frac{n}{n+\alpha}\mathbb{P}_n, \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$$

2. For a loss function $l(x, \theta)$ *propagate uncertainty* from $\mathbb{P}^*$ to the parameter of interest θ through

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta\in\Theta} \mathbb{E}_{X\sim\mathbb{P}^*}[l(X, \theta)]}$$

3. The push-forward measure $(\theta_l^*)_\#(\mathrm{DP}(\alpha', \mathbb{F}'))$ gives a posterior on Θ denoted by Π_{NPL}.

Bayesian Nonparametric Learning (NPL) framework (Lyddon et al., 2018; Fong et al., 2019)

1. Place a nonparametric prior *directly* on the data-generating mechanism $\mathbb{P}^*$:

$$\mathbb{P} \sim DP(\alpha, \mathbb{F}), \quad \mathbb{P}|x_{1:n} \sim DP\left(\alpha', \mathbb{F}'\right)$$

   where

$$\alpha' = \alpha + n, \quad \mathbb{F}' := \tfrac{\alpha}{\alpha+n}\mathbb{F} + \tfrac{n}{n+\alpha}\mathbb{P}_n, \quad \mathbb{P}_n = \tfrac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$

2. For a loss function $l(x, \theta)$ *propagate uncertainty* from $\mathbb{P}^*$ to the parameter of interest θ through

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta\in\Theta} \mathbb{E}_{X\sim\mathbb{P}^*}[l(X, \theta)]}$$

3. The push-forward measure $(\theta_l^*)_\#(\mathrm{DP}(\alpha', \mathbb{F}'))$ gives a posterior on Θ denoted by Π_{NPL}.

Bayesian Nonparametric Learning (NPL) framework (Lyddon et al., 2018; Fong et al., 2019)

1. Place a nonparametric prior *directly* on the data-generating mechanism $\mathbb{P}^*$:

$$\mathbb{P} \sim DP(\alpha, \mathbb{F}), \quad \mathbb{P}|x_{1:n} \sim DP\left(\alpha', \mathbb{F}'\right)$$

   where

$$\alpha' = \alpha + n, \quad \mathbb{F}' := \tfrac{\alpha}{\alpha+n}\mathbb{F} + \tfrac{n}{n+\alpha}\mathbb{P}_n, \quad \mathbb{P}_n = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \delta_{x_i}$$

2. For a loss function $l(x, \theta)$ *propagate uncertainty* from $\mathbb{P}^*$ to the parameter of interest θ through

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta \in \Theta} \mathbb{E}_{X \sim \mathbb{P}^*}[l(X, \theta)]}$$

3. The push-forward measure $(\theta_l^*)_{\#}(\mathrm{DP}(\alpha', \mathbb{F}'))$ gives a posterior on Θ denoted by Π_{NPL}.

Bayesian Nonparametric Learning (NPL) framework (Lyddon et al., 2018; Fong et al., 2019)

1. Place a nonparametric prior *directly* on the data-generating mechanism $\mathbb{P}^*$:

$$\mathbb{P} \sim DP(\alpha, \mathbb{F}), \quad \mathbb{P}|x_{1:n} \sim DP\left(\alpha', \mathbb{F}'\right)$$

   where

$$\alpha' = \alpha + n, \quad \mathbb{F}' := \tfrac{\alpha}{\alpha+n}\mathbb{F} + \tfrac{n}{n+\alpha}\mathbb{P}_n, \quad \mathbb{P}_n = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \delta_{x_i}$$

2. For a loss function $l(x, \theta)$ *propagate uncertainty* from $\mathbb{P}^*$ to the parameter of interest θ through

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta\in\Theta} \mathbb{E}_{X\sim\mathbb{P}^*}[l(X, \theta)]}$$

3. The push-forward measure $(\theta_l^*)_\#(\mathrm{DP}(\alpha', \mathbb{F}'))$ gives a posterior on Θ denoted by Π_{NPL}.

Push-forward measure

To obtain independent realisations from Π_{NPL} at iteration j:

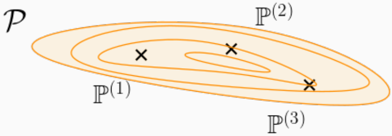
1. Sample $\mathbb{P}^{(j)}$ from the posterior DP;
2. Compute $\theta^{(j)} = \theta_l^*(\mathbb{P}^{(j)})$

Push-forward measure

To obtain independent realisations from Π_{NPL} at iteration j:

1. Sample $\mathbb{P}^{(j)}$ from the posterior DP;
2. Compute $\theta^{(j)} = \theta_l^*(\mathbb{P}^{(j)})$

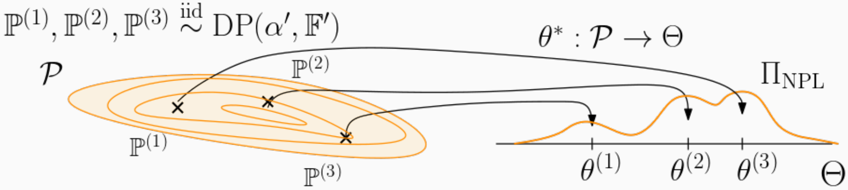
$$\mathbb{P}^{(1)}, \mathbb{P}^{(2)}, \mathbb{P}^{(3)} \overset{\text{iid}}{\sim} \mathrm{DP}(\alpha', \mathbb{F}')$$



Push-forward measure

To obtain independent realisations from Π_{NPL} at iteration j:

1. Sample $\mathbb{P}^{(j)}$ from the posterior DP;
2. Compute $\theta^{(j)} = \theta_l^*(\mathbb{P}^{(j)})$



Method

Distance-based loss function

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta\in\Theta} \mathbb{E}_{X\sim\mathbb{P}^*}[l(X,\theta)]}$$

$$\downarrow$$

$$\mathrm{D}(\mathbb{P}_\theta, \mathbb{P}^*)$$

- Use a distance-based loss between real and target distribution
- Target: $\theta_0 = \arg\inf_{\theta\in\Theta} D(\mathbb{P}^*, \mathbb{P}_\theta)$ for some distance function $D : \mathcal{P}\times\mathcal{P} \to \mathbb{R}_+$
- Minimisation of expected loss $\longrightarrow$ Minimum Distance Estimation (MDE), Parr and Schucany (1980)

$$\theta_D^*(\mathbb{P}^*) = \arg\inf_{\theta\in\Theta} D(\mathbb{P}^*, \mathbb{P}_\theta)$$

Distance-based loss function

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta\in\Theta} \mathbb{E}_{X\sim\mathbb{P}^*}[l(X,\theta)]}$$

$$\downarrow$$

$$\mathrm{D}(\mathbb{P}_\theta, \mathbb{P}^*)$$

- Use a distance-based loss between real and target distribution
- Target: $\theta_0 = \arg\inf_{\theta\in\Theta} D(\mathbb{P}^*, \mathbb{P}_\theta)$ for some distance function $D : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+$
- Minimisation of expected loss $\longrightarrow$ Minimum Distance Estimation (MDE), Parr and Schucany (1980)

$$\theta_D^*(\mathbb{P}^*) = \arg\inf_{\theta\in\Theta} D(\mathbb{P}^*, \mathbb{P}_\theta)$$

Distance-based loss function

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta\in\Theta} \underset{\displaystyle\big\downarrow \atop \displaystyle \mathrm{D}(\mathbb{P}_\theta, \mathbb{P}^*)}{\mathbb{E}_{X\sim\mathbb{P}^*}[l(X,\theta)]}}$$

- Use a distance-based loss between real and target distribution
- Target: $\theta_0 = \arg\inf_{\theta\in\Theta} D(\mathbb{P}^*, \mathbb{P}_\theta)$ for some distance function $D : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+$
- Minimisation of expected loss $\longrightarrow$ Minimum Distance Estimation (MDE), Parr and Schucany (1980)

$$\theta_D^*(\mathbb{P}^*) = \arg\inf_{\theta\in\Theta} D(\mathbb{P}^*, \mathbb{P}_\theta)$$

Proposal: Maximum Mean Discrepancy (MMD) based loss

- Proposal: *Maximum Mean Discrepancy (MMD)*
- Map θ^* now corresponds to a minimum MMD estimator as in Briol et al. (2019); Chérief-Abdellatif and Alquier (2022)
- *Integral Probability Metrics (IPMs), Müller (1997)*:

$$D(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{X}} f(x)\mathbb{P}(dx) - \int_{\mathcal{X}} f(x)\mathbb{Q}(dx) \right|$$

- **Maximum Mean Discrepancy (MMD)**: Restrict $\mathcal{F}$ to a unit *Reproducing Kernel Hilbert space* (RKHS), $\mathcal{H}_k$, defined through a symmetric, positive definite kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, with associated norm $\| \cdot \|_{\mathcal{H}_k}$ and inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k} : \mathcal{H}_k \times \mathcal{H}_k \to \mathbb{R}$.
- Reproducing property:

$$f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}_k}$$

- Functions in $\mathcal{H}_k$ have the form $f(x) = \sum_{i=1}^{d} c_i k(x, x_i)$ for some $c_i \in \mathbb{R}$

Proposal: Maximum Mean Discrepancy (MMD) based loss

- Proposal: *Maximum Mean Discrepancy (MMD)*
- Map θ^* now corresponds to a minimum MMD estimator as in Briol et al. (2019); Chérief-Abdellatif and Alquier (2022)
- *Integral Probability Metrics (IPMs), Müller (1997)*:

$$D(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{X}} f(x)\mathbb{P}(dx) - \int_{\mathcal{X}} f(x)\mathbb{Q}(dx) \right|$$

- **Maximum Mean Discrepancy (MMD)**: Restrict $\mathcal{F}$ to a unit *Reproducing Kernel Hilbert space* (RKHS), $\mathcal{H}_k$, defined through a symmetric, positive definite kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, with associated norm $\|\cdot\|_{\mathcal{H}_k}$ and inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k} : \mathcal{H}_k \times \mathcal{H}_k \to \mathbb{R}$.
- Reproducing property:

$$f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}_k}$$

- Functions in $\mathcal{H}_k$ have the form $f(x) = \sum_{i=1}^{d} c_i k(x, x_i)$ for some $c_i \in \mathbb{R}$

Proposal: Maximum Mean Discrepancy (MMD) based loss

- Proposal: *Maximum Mean Discrepancy (MMD)*
- Map θ^* now corresponds to a minimum MMD estimator as in Briol et al. (2019); Chérief-Abdellatif and Alquier (2022)
- *Integral Probability Metrics (IPMs), Müller (1997)*:

$$D(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{X}} f(x)\mathbb{P}(dx) - \int_{\mathcal{X}} f(x)\mathbb{Q}(dx) \right|$$

- **Maximum Mean Discrepancy (MMD)**: Restrict $\mathcal{F}$ to a unit *Reproducing Kernel Hilbert space* (RKHS), $\mathcal{H}_k$, defined through a symmetric, positive definite kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, with associated norm $\| \cdot \|_{\mathcal{H}_k}$ and inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k} : \mathcal{H}_k \times \mathcal{H}_k \to \mathbb{R}$.
- Reproducing property:

$$f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}_k}$$

- Functions in $\mathcal{H}_k$ have the form $f(x) = \sum_{i=1}^{d} c_i k(x, x_i)$ for some $c_i \in \mathbb{R}$

Maximum Mean Discrepancy (MMD)

- So MMD is defined as:

$$MMD(\mathbb{P},\mathbb{Q}) = \sup_{\substack{f\in\mathcal{H}_k,\\ \|f\|_{\mathcal{H}_k}\leq 1}} \left| \int_{\mathcal{X}} f(x)\mathbb{P}(dx) - \int_{\mathcal{X}} f(x)\mathbb{Q}(dx) \right|$$

- The MMD between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ can be expressed as

$$\begin{aligned} MMD^2(\mathbb{P},\mathbb{Q}) := & \int_{\mathcal{X}}\int_{\mathcal{X}} k(x,y)\mathbb{P}(dx)\mathbb{P}(dy) - 2\int_{\mathcal{X}}\int_{\mathcal{X}} k(x,y)\mathbb{P}(dx)\mathbb{Q}(dy) \\ & + \int_{\mathcal{X}}\int_{\mathcal{X}} k(x,y)\mathbb{Q}(dx)\mathbb{Q}(dy) \end{aligned}$$

- It can be estimated for example using a U-statistic as in Gretton et al. (2008):

$$\begin{aligned} MMD^2_{k,U}(\mathbb{P}^m,\mathbb{Q}^n) = & \tfrac{1}{m(m-1)}\sum_{i\neq j} k(y_i,y_j) - \tfrac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} k(x_i,y_j) \\ & + \tfrac{1}{n(n-1)}\sum_{i\neq j} k(x_i,x_j) \end{aligned}$$

where $\{y_j\}_{j=1}^m \overset{iid}{\sim} \mathbb{P}$ and $\{x_i\}_{i=1}^n \overset{iid}{\sim} \mathbb{Q}$.

$$\longrightarrow \theta^*(\mathbb{P}^*) = \arg\inf_{\theta\in\Theta} \mathrm{MMD}^2_{k,U}(\mathbb{P}^*,\mathbb{P}_\theta)$$

Maximum Mean Discrepancy (MMD)

- So MMD is defined as:

$$MMD(\mathbb{P},\mathbb{Q}) = \sup_{\substack{f\in\mathcal{H}_k,\\ \|f\|_{\mathcal{H}_k}\leq 1}} \left| \int_{\mathcal{X}} f(x)\mathbb{P}(dx) - \int_{\mathcal{X}} f(x)\mathbb{Q}(dx) \right|$$

- The MMD between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ can be expressed as

$$\begin{aligned} MMD^2(\mathbb{P},\mathbb{Q}) := & \int_{\mathcal{X}}\int_{\mathcal{X}} k(x,y)\mathbb{P}(dx)\mathbb{P}(dy) - 2\int_{\mathcal{X}}\int_{\mathcal{X}} k(x,y)\mathbb{P}(dx)\mathbb{Q}(dy) \\ & + \int_{\mathcal{X}}\int_{\mathcal{X}} k(x,y)\mathbb{Q}(dx)\mathbb{Q}(dy) \end{aligned}$$

- It can be estimated for example using a U-statistic as in Gretton et al. (2008):

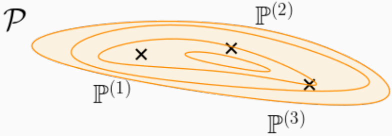
$$\begin{aligned} MMD^2_{k,U}(\mathbb{P}^m,\mathbb{Q}^n) = & \tfrac{1}{m(m-1)}\sum_{i\neq j} k(y_i,y_j) - \tfrac{2}{nm}\sum_{i=1}^n\sum_{j=1}^m k(x_i,y_j) \\ & + \tfrac{1}{n(n-1)}\sum_{i\neq j} k(x_i,x_j) \end{aligned}$$

where $\{y_j\}_{j=1}^m \overset{iid}{\sim} \mathbb{P}$ and $\{x_i\}_{i=1}^n \overset{iid}{\sim} \mathbb{Q}$.

$$\longrightarrow \boxed{\theta^*(\mathbb{P}^*) = \arg\inf_{\theta\in\Theta} \mathrm{MMD}^2_{k,U}(\mathbb{P}^*,\mathbb{P}_\theta)}$$

MMD Posterior Bootstrap

$\mathbb{P}^{(1)}, \mathbb{P}^{(2)}, \mathbb{P}^{(3)} \overset{\text{iid}}{\sim} \mathrm{DP}(\alpha', \mathbb{F}')$



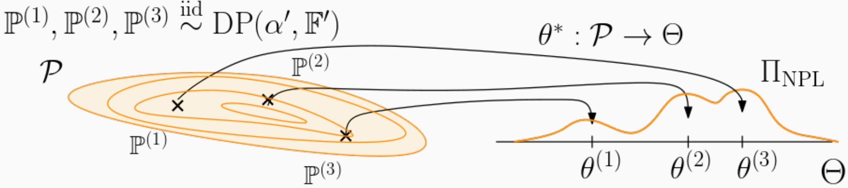
1. Draw $\mathbb{P}^{(j)} \sim DP(\alpha', \mathbb{F}')$
2. Obtain $\theta^{(j)} := \theta^*(\mathbb{P}^{(j)}) = \arg\min_{\theta \in \Theta} \mathrm{MMD}^2_{k,U}(\mathbb{P}^{(j)}, \mathbb{P}_\theta)$

$$\tilde{x}^{(j)}_{1:T} \overset{\text{iid}}{\sim} \mathbb{F}, \quad (w^{(j)}_{1:n}, \tilde{w}^{(j)}_{1:T}) \sim \mathrm{Dir}\left(1, \ldots, 1, \frac{\alpha}{T}, \ldots, \frac{\alpha}{T}\right).$$

$$\mathbb{P}^{(j)} = \sum_{i=1}^{n} w^{(j)}_i \delta_{x_i} + \sum_{k=1}^{T} \tilde{w}^{(j)}_k \delta_{\tilde{x}^{(j)}_k} \sim \hat{\nu}. \tag{1}$$

where $\hat{\nu}$ denotes the probability measure on $\mathcal{P}$ defined by (1).

MMD Posterior Bootstrap



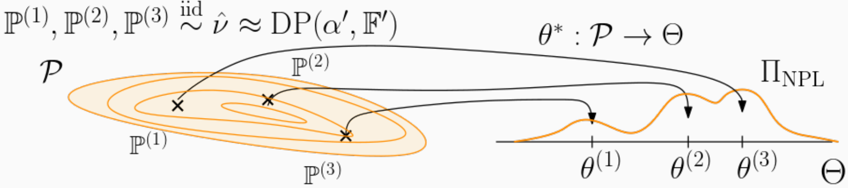
1. Draw $\mathbb{P}^{(j)} \sim DP(\alpha', \mathbb{F}')$
2. Obtain $\theta^{(j)} := \theta^*(\mathbb{P}^{(j)}) = \arg\min_{\theta \in \Theta} \mathrm{MMD}^2_{k,U}(\mathbb{P}^{(j)}, \mathbb{P}_\theta)$

$$\tilde{x}^{(j)}_{1:T} \overset{\text{iid}}{\sim} \mathbb{F}, \quad (w^{(j)}_{1:n}, \tilde{w}^{(j)}_{1:T}) \sim \mathrm{Dir}\left(1, \ldots, 1, \frac{\alpha}{T}, \ldots, \frac{\alpha}{T}\right).$$

$$\mathbb{P}^{(j)} = \sum_{i=1}^{n} w^{(j)}_i \delta_{x_i} + \sum_{k=1}^{T} \tilde{w}^{(j)}_k \delta_{\tilde{x}^{(j)}_k} \sim \hat{\nu}. \tag{1}$$

where $\hat{\nu}$ denotes the probability measure on $\mathcal{P}$ defined by (1).

MMD Posterior Bootstrap



1. Draw $\mathbb{P}^{(j)} \sim DP(\alpha', \mathbb{F}')$
2. Obtain $\theta^{(j)} := \theta^*(\mathbb{P}^{(j)}) = \arg\min_{\theta \in \Theta} \mathrm{MMD}^2_{k,U}(\mathbb{P}^{(j)}, \mathbb{P}_\theta)$

$$\tilde{x}^{(j)}_{1:T} \overset{\text{iid}}{\sim} \mathbb{F}, \quad (w^{(j)}_{1:n}, \tilde{w}^{(j)}_{1:T}) \sim \mathrm{Dir}\left(1, \ldots, 1, \frac{\alpha}{T}, \ldots, \frac{\alpha}{T}\right).$$

$$\mathbb{P}^{(j)} = \sum_{i=1}^{n} w^{(j)}_i \delta_{x_i} + \sum_{k=1}^{T} \tilde{w}^{(j)}_k \delta_{\tilde{x}^{(j)}_k} \sim \hat{\nu}. \tag{1}$$

where $\hat{\nu}$ denotes the probability measure on $\mathcal{P}$ defined by (1).

Posterior Bootstrap with the MMD

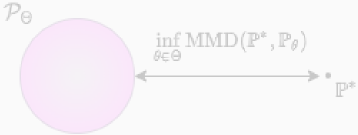
Algorithm 1: MMD Posterior Bootstrap

```
input: $x_{1:n}$, $T$, $B$, $\alpha$, $\mathbb{F}$, $\mathbb{U}$, $G_\theta$.
for $j \leftarrow 1$ to $B$ do
    Sample $\tilde{x}^{(j)}_{1:T} \overset{\text{iid}}{\sim} \mathbb{F}$ and
    $(w^{(j)}_{1:n}, \tilde{w}^{(j)}_{1:T}) \sim \text{Dir}\left(1, \ldots, 1, \frac{\alpha}{T}, \ldots, \frac{\alpha}{T}\right)$.
    Set $\mathbb{P}^{(j)} = \sum_{i=1}^{n} w^{(j)}_i \delta_{x_i} + \sum_{k=1}^{T} \tilde{w}^{(j)}_k \delta_{\tilde{x}^{(j)}_k}$.
    Obtain $\theta^{(j)} = \theta^*(\mathbb{P}^{(j)})$ using numerical optimisation.
return Posterior bootstrap sample $\theta^{(1:B)}$
```

Theoretical Results

Generalisation error

- **Assumption:** $\sup_{x,x' \in \mathcal{X}} |k(x,x')| < \infty$
- W.L.O.G. $|k(x,x')| \leq 1 \quad \forall \quad x,x' \in \mathcal{X}$
- $\nu := DP(\alpha', \mathbb{F}')$



Generalisation error

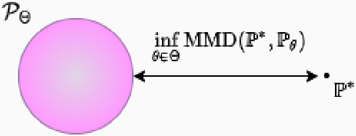
- **Assumption:** $\sup_{x,x' \in \mathcal{X}} |k(x,x')| < \infty$
- W.L.O.G. $|k(x,x')| \leq 1 \quad \forall \quad x, x' \in \mathcal{X}$
- $\nu := DP(\alpha', \mathbb{F}')$

$$\mathbb{E}_{x_{1:n} \overset{\text{iid}}{\sim} \mathbb{P}^*} \left[\mathbb{E}_{\mathbb{P} \sim \nu} \left[\text{MMD}(\mathbb{P}^*, \mathbb{P}_{\theta^*(\mathbb{P})}) \right] \right]$$

Generalisation error

- **Assumption:** $\sup_{x,x' \in \mathcal{X}} |k(x,x')| < \infty$
- W.L.O.G. $|k(x,x')| \leq 1 \quad \forall \quad x, x' \in \mathcal{X}$
- $\nu := DP(\alpha', \mathbb{F}')$

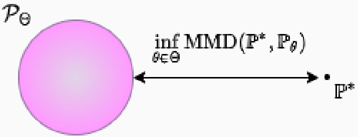
$$\inf_{\theta \in \Theta} \text{MMD}(\mathbb{P}^*, \mathbb{P}_\theta) \quad \leq \quad \mathbb{E}_{x_{1:n} \overset{\text{iid}}{\sim} \mathbb{P}^*} \left[\mathbb{E}_{\mathbb{P} \sim \nu} \left[\text{MMD}(\mathbb{P}^*, \mathbb{P}_{\theta^*(\mathbb{P})}) \right] \right]$$



Generalisation error

- **Assumption:** $\sup_{x,x' \in \mathcal{X}} |k(x,x')| < \infty$
- W.L.O.G. $|k(x,x')| \leq 1 \quad \forall \quad x,x' \in \mathcal{X}$
- $\nu := DP(\alpha', \mathbb{F}')$

$$\begin{aligned} 0 \quad &\leq \quad \overbrace{\mathbb{E}_{x_{1:n} \overset{\text{iid}}{\sim} \mathbb{P}^*}\left[\mathbb{E}_{\mathbb{P}\sim\nu}\left[\text{MMD}(\mathbb{P}^*, \mathbb{P}_{\theta^*(\mathbb{P})})\right]\right] - \inf_{\theta\in\Theta} \text{MMD}(\mathbb{P}^*, \mathbb{P}_\theta)}^{\text{Generalisation Error}} \\ &\leq \quad \frac{2}{\sqrt{n}} + \frac{2}{\sqrt{\alpha+n+2}} + \frac{4\alpha}{\alpha+n} \end{aligned}$$



- Rate agrees with results in Chérief-Abdellatif and Alquier (2022)

Posterior consistency

- For standard Bayesian inference with posterior measure Π_n defined on Θ directly, for any $M_n \to +\infty$ such that $M_n n^{-\frac{1}{2}} \to 0$:

$$\Pi_n \left(\theta \in \Theta : \mathrm{MMD}(\mathbb{P}_\theta, \mathbb{P}^*) > \inf_{\theta \in \Theta} \mathrm{MMD}(\mathbb{P}_\theta, \mathbb{P}^*) + \tfrac{M_n}{n^{1/2}} \right) \xrightarrow{n \to \infty} 0 \tag{2}$$

- In our case:

$$\nu \left(\mathbb{P} \in \mathcal{P} : \mathrm{MMD}(\mathbb{P}_{\theta^*(\mathbb{P})}, \mathbb{P}^*) > \inf_{\theta \in \Theta} \mathrm{MMD}(\mathbb{P}_\theta, \mathbb{P}^*) + \frac{M_n}{n^{1/2}} \right) \xrightarrow{n \to \infty} 0.$$

Posterior consistency

- For standard Bayesian inference with posterior measure Π_n defined on Θ directly, for any $M_n \to +\infty$ such that $M_n n^{-\frac{1}{2}} \to 0$:

$$\Pi_n \left(\theta \in \Theta : \mathrm{MMD}(\mathbb{P}_\theta, \mathbb{P}^*) > \inf_{\theta \in \Theta} \mathrm{MMD}(\mathbb{P}_\theta, \mathbb{P}^*) + \tfrac{M_n}{n^{1/2}} \right) \xrightarrow{n \to \infty} 0 \qquad (2)$$

- In our case:

$$\nu \left(\mathbb{P} \in \mathcal{P} : \mathrm{MMD}(\mathbb{P}_{\theta^*(\mathbb{P})}, \mathbb{P}^*) > \inf_{\theta \in \Theta} \mathrm{MMD}(\mathbb{P}_\theta, \mathbb{P}^*) + \frac{M_n}{n^{1/2}} \right) \xrightarrow{n \to \infty} 0.$$

Robustness to outliers

- Suppose $\mathbb{P}^* = (1-\epsilon)\tilde{\mathbb{P}} + \epsilon\,\mathbb{Q}$
- Then

$$\mathbb{E}_{x_{1:n} \overset{\text{iid}}{\sim} \mathbb{P}^*}\left[\mathbb{E}_{\mathbb{P}\sim\hat{\nu}}\left[\mathrm{MMD}\left(\tilde{\mathbb{P}}, \mathbb{P}_{\theta^*(\mathbb{P})}\right)\right]\right]$$

$$\leq \inf_{\theta\in\Theta} \mathrm{MMD}(\tilde{\mathbb{P}}, \mathbb{P}_\theta) + 4\epsilon + \frac{2}{\sqrt{n}} + \frac{2}{\sqrt{\alpha+n+2}} + \frac{4\alpha}{\alpha+n}$$

Robustness to outliers

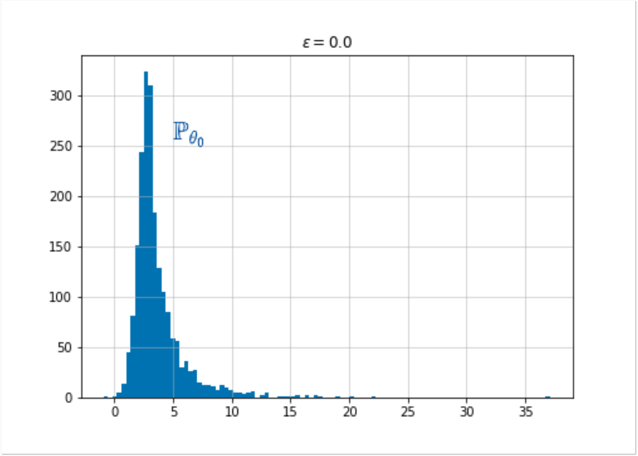
- Suppose $\mathbb{P}^* = (1-\epsilon)\tilde{\mathbb{P}} + \epsilon\,\mathbb{Q}$
- Then

$$\begin{aligned}&\mathbb{E}_{x_{1:n} \overset{\text{iid}}{\sim} \mathbb{P}^*}\left[\mathbb{E}_{\mathbb{P}\sim\hat{\nu}}\left[\mathrm{MMD}\left(\tilde{\mathbb{P}}, \mathbb{P}_{\theta^*(\mathbb{P})}\right)\right]\right]\\ &\leq \inf_{\theta\in\Theta} \mathrm{MMD}(\tilde{\mathbb{P}}, \mathbb{P}_\theta) + 4\epsilon + \frac{2}{\sqrt{n}} + \frac{2}{\sqrt{\alpha+n+2}} + \frac{4\alpha}{\alpha+n}\end{aligned}$$

Applications

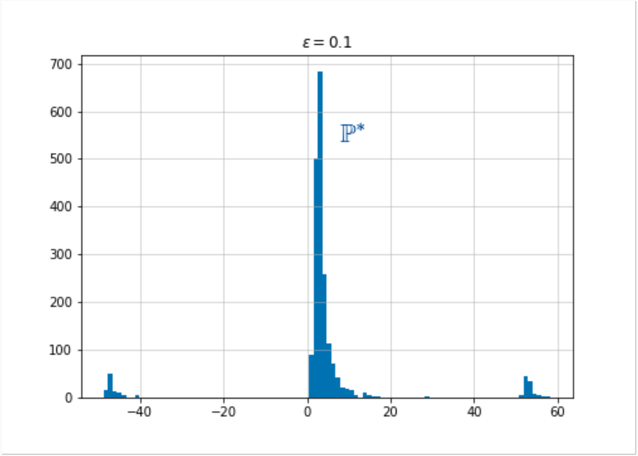
Example: Contaminated G-and-k distribution model

$\mathbb{P}^* = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{Q}$ where $\mathbb{P}_{\theta_0}$ denotes the G-and-k distribution with $\theta_0 = (3, 1, 1, -\log(2))$, and $\mathbb{Q}$ is the shifted distribution $\mathbb{Q} = \mathbb{P}_{\theta_0} \pm 50$



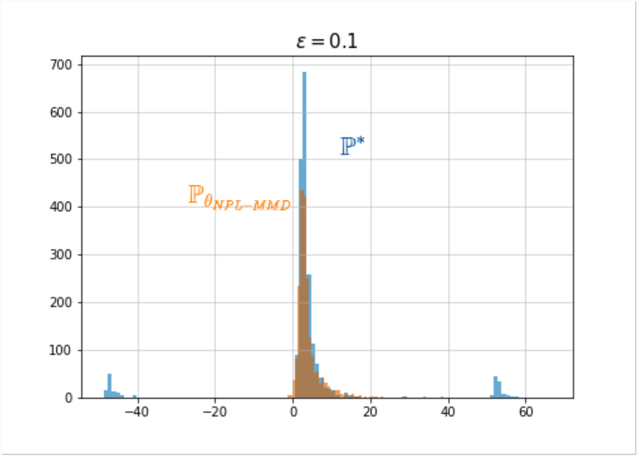
Example: Contaminated G-and-k distribution model

$\mathbb{P}^* = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{Q}$ where $\mathbb{P}_{\theta_0}$ denotes the G-and-k distribution with $\theta_0 = (3, 1, 1, -\log(2))$, and $\mathbb{Q}$ is the shifted distribution $\mathbb{Q} = \mathbb{P}_{\theta_0} \pm 50$



Example: Contaminated G-and-k distribution model

$\mathbb{P}^* = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{Q}$ where $\mathbb{P}_{\theta_0}$ denotes the G-and-k distribution with $\theta_0 = (3, 1, 1, -\log(2))$, and $\mathbb{Q}$ is the shifted distribution $\mathbb{Q} = \mathbb{P}_{\theta_0} \pm 50$



Example: Contaminated G-and-k distribution model

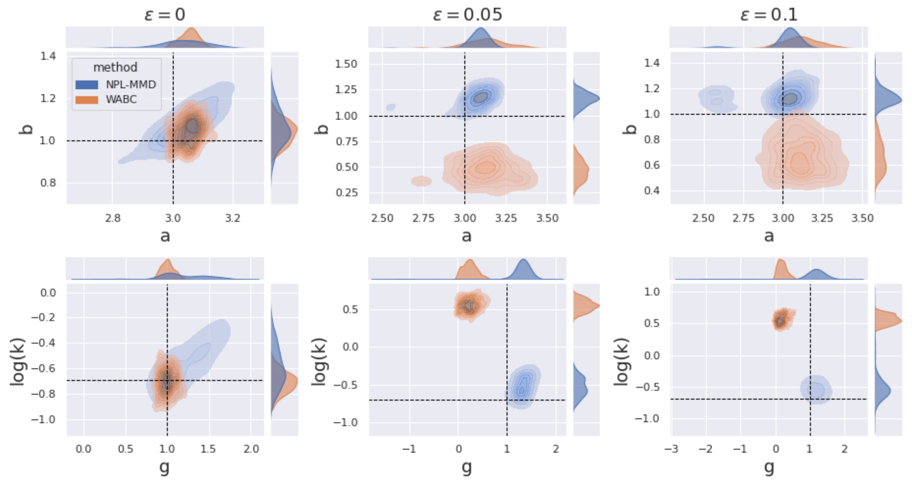


Figure 1: Comparison of posterior marginal distributions obtained using the MMD Posterior Bootstrap (NPL-MMD) and the Wasserstein-ABC (WABC) method in Bernton et al. (2019).

Example: Toggle switch model with Cauchy noise

- Arising in Systems Biology (see e.g. Bonassi et al., 2011)
- A dynamic model used to study cellular networks describing the interaction of two genes over time
- For cell i and unknown parameters $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \mu, \sigma, \gamma)^\top$, the simulator input is $u_i = (u_{i,1,1}, u_{i,1,2}, \ldots, u_{i,T,1}, u_{i,T,2}, u_{i,T+1,1})^\top \sim \text{Unif}([0,1]^{2T+1})$ and the simulator G_θ is defined through:

$$G_\theta(u_i) = \Phi^{-1}\left(\Phi\left(\frac{-(\mu+v_{i,T})v_{i,T}^\gamma}{\mu\sigma}\right) + u_{i,T+1,1}\left(1 - \Phi\left(\frac{-(\mu+v_{i,T})v_{i,T}^\gamma}{\mu\sigma}\right)\right)\right)\frac{\mu\sigma}{v_{i,T}^\gamma} + (\mu + v_{i,T})$$

where for $t = 1, \ldots, T-1$, we have

$$\begin{aligned}
\tilde{v}_{i,t+1} &= v_{i,t} + \frac{\alpha_1}{1+w_{i,t}^{\beta_1}} - (1 + 0.03v_{i,t}) \\
\tilde{w}_{i,t+1} &= w_{i,t} + \frac{\alpha_2}{1+v_{i,t}^{\beta_2}} - (1 + 0.03w_{i,t}) \\
v_{i,t+1} &= \tilde{v}_{i,t+1} + 0.5\Phi^{-1}(\Phi(-2\tilde{v}_{i,t+1}) + u_{i,t,1}(1 - \Phi(-2\tilde{v}_{i,t+1}))) \\
w_{i,t+1} &= \tilde{w}_{i,t+1} + 0.5\Phi^{-1}(\Phi(-2\tilde{w}_{i,t+1}) + u_{i,t,2}(1 - \Phi(-2\tilde{w}_{i,t+1})))
\end{aligned}$$

Example: Toggle switch model with Cauchy noise

- Arising in Systems Biology (see e.g. Bonassi et al., 2011)
- A dynamic model used to study cellular networks describing the interaction of two genes over time
- For cell i and unknown parameters $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \mu, \sigma, \gamma)^\top$, the simulator input is $u_i = (u_{i,1,1}, u_{i,1,2}, \ldots, u_{i,T,1}, u_{i,T,2}, u_{i,T+1,1})^\top \sim \mathrm{Unif}([0,1]^{2T+1})$ and the simulator G_θ is defined through:

$$G_\theta(u_i) = \Phi^{-1}\Big(\Phi\Big(\tfrac{-(\mu+v_{i,T})v_{i,T}^{\gamma}}{\mu\sigma}\Big) + u_{i,T+1,1}\Big(1-\Phi\Big(\tfrac{-(\mu+v_{i,T})v_{i,T}^{\gamma}}{\mu\sigma}\Big)\Big)\Big)\tfrac{\mu\sigma}{v_{i,T}^{\gamma}} + (\mu + v_{i,T})$$

where for $t = 1, \ldots, T-1$, we have

$$\tilde{v}_{i,t+1} = v_{i,t} + \tfrac{\alpha_1}{1+w_{i,t}^{\beta_1}} - (1 + 0.03 v_{i,t})$$

$$\tilde{w}_{i,t+1} = w_{i,t} + \tfrac{\alpha_2}{1+v_{i,t}^{\beta_2}} - (1 + 0.03 w_{i,t})$$

$$v_{i,t+1} = \tilde{v}_{i,t+1} + 0.5\Phi^{-1}\big(\Phi(-2\tilde{v}_{i,t+1}) + u_{i,t,1}(1-\Phi(-2\tilde{v}_{i,t+1}))\big)$$

$$w_{i,t+1} = \tilde{w}_{i,t+1} + 0.5\Phi^{-1}\big(\Phi(-2\tilde{w}_{i,t+1}) + u_{i,t,2}(1-\Phi(-2\tilde{w}_{i,t+1}))\big)$$

Example: Toggle switch model with Cauchy noise

- Arising in Systems Biology (see e.g. Bonassi et al., 2011)
- A dynamic model used to study cellular networks describing the interaction of two genes over time
- For cell i and unknown parameters $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \mu, \sigma, \gamma)^\top$, the simulator input is $u_i = (u_{i,1,1}, u_{i,1,2}, \ldots, u_{i,T,1}, u_{i,T,2}, u_{i,T+1,1})^\top \sim \text{Unif}([0,1]^{2T+1})$ and the simulator G_θ is defined through:

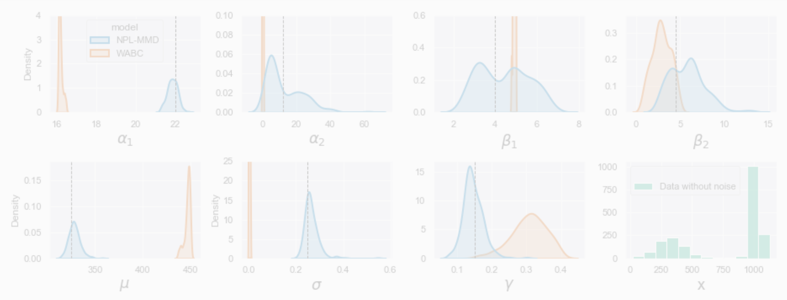
$$G_\theta(u_i) = \Phi^{-1}\Big(\Phi\Big(\tfrac{-(\mu+v_{i,T})v_{i,T}^{\gamma}}{\mu\sigma}\Big) + u_{i,T+1,1}\Big(1 - \Phi\Big(\tfrac{-(\mu+v_{i,T})v_{i,T}^{\gamma}}{\mu\sigma}\Big)\Big)\Big)\tfrac{\mu\sigma}{v_{i,T}^{\gamma}} + (\mu + v_{i,T})$$

where for $t = 1, \ldots, T-1$, we have

$$\begin{aligned}
\tilde{v}_{i,t+1} &= v_{i,t} + \tfrac{\alpha_1}{1+w_{i,t}^{\beta_1}} - (1 + 0.03v_{i,t}) \\
\tilde{w}_{i,t+1} &= w_{i,t} + \tfrac{\alpha_2}{1+v_{i,t}^{\beta_2}} - (1 + 0.03w_{i,t}) \\
v_{i,t+1} &= \tilde{v}_{i,t+1} + 0.5\Phi^{-1}\big(\Phi(-2\tilde{v}_{i,t+1}) + u_{i,t,1}(1 - \Phi(-2\tilde{v}_{i,t+1}))\big) \\
w_{i,t+1} &= \tilde{w}_{i,t+1} + 0.5\Phi^{-1}\big(\Phi(-2\tilde{w}_{i,t+1}) + u_{i,t,2}(1 - \Phi(-2\tilde{w}_{i,t+1}))\big)
\end{aligned}$$

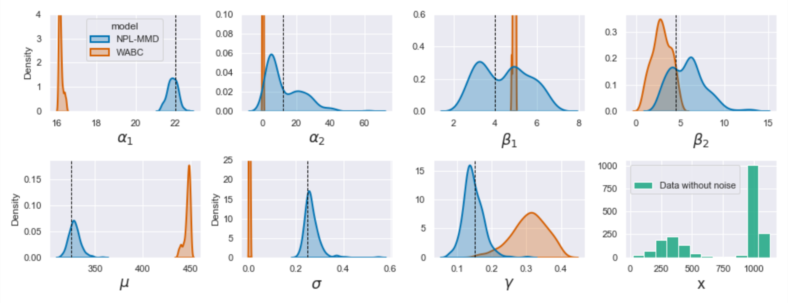
Example: Toggle switch model with Cauchy noise

- Inference on $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \mu, \sigma, \gamma)$ for $n = 2000$ data points simulated from the toggle-switch model in which 10% of the data have some added Cauchy noise of location parameter 0 and scale parameter 10.



Example: Toggle switch model with Cauchy noise

- Inference on $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \mu, \sigma, \gamma)$ for $n = 2000$ data points simulated from the toggle-switch model in which 10% of the data have some added Cauchy noise of location parameter 0 and scale parameter 10.



Discussion & current work

Strengths & Limitations

Strengths:

- ✓ Posterior Bootstrap algorithm suitable for simulator-based models inducing robustness through:
    1. Bayesian Nonparametric Learning (NPL) framework
    2. Minimum Maximum Mean Discrepancy (MMD) estimators
- ✓ Highly parallelisable algorithm
- ✓ Generalisation error, posterior consistency and robustness to outliers guarantees

Limitations:

- ✓ Optimisation; objective usually not convex
- ✓ Kernel choice and associated hyperparameters
- ✓ MMD approximation; estimators with improved sample complexity e.g. Bharti et al. (2023)

Strengths & Limitations

Strengths:

- ✓ Posterior Bootstrap algorithm suitable for simulator-based models inducing robustness through:
  1. Bayesian Nonparametric Learning (NPL) framework
  2. Minimum Maximum Mean Discrepancy (MMD) estimators
- ✓ Highly parallelisable algorithm
- ✓ Generalisation error, posterior consistency and robustness to outliers guarantees

Limitations:

- ✓ Optimisation; objective usually not convex
- ✓ Kernel choice and associated hyperparameters
- ✓ MMD approximation; estimators with improved sample complexity e.g. Bharti et al. (2023)

Strengths & Limitations

Strengths:

- ✓ Posterior Bootstrap algorithm suitable for simulator-based models inducing robustness through:
  1. Bayesian Nonparametric Learning (NPL) framework
  2. Minimum Maximum Mean Discrepancy (MMD) estimators
- ✓ Highly parallelisable algorithm
- ✓ Generalisation error, posterior consistency and robustness to outliers guarantees

Limitations:

- ✓ Optimisation; objective usually not convex
- ✓ Kernel choice and associated hyperparameters
- ✓ MMD approximation; estimators with improved sample complexity e.g. Bharti et al. (2023)

Strengths & Limitations

Strengths:

- ✓ Posterior Bootstrap algorithm suitable for simulator-based models inducing robustness through:
  1. Bayesian Nonparametric Learning (NPL) framework
  2. Minimum Maximum Mean Discrepancy (MMD) estimators
- ✓ Highly parallelisable algorithm
- ✓ Generalisation error, posterior consistency and robustness to outliers guarantees

Limitations:

- ✓ Optimisation; objective usually not convex
- ✓ Kernel choice and associated hyperparameters
- ✓ MMD approximation; estimators with improved sample complexity e.g. Bharti et al. (2023)

Strengths & Limitations

Strengths:

- ✓ Posterior Bootstrap algorithm suitable for simulator-based models inducing robustness through:
  1. Bayesian Nonparametric Learning (NPL) framework
  2. Minimum Maximum Mean Discrepancy (MMD) estimators
- ✓ Highly parallelisable algorithm
- ✓ Generalisation error, posterior consistency and robustness to outliers guarantees

Limitations:

- ✓ Optimisation; objective usually not convex
- ✓ Kernel choice and associated hyperparameters
- ✓ MMD approximation; estimators with improved sample complexity e.g. Bharti et al. (2023)

Strengths & Limitations

Strengths:

- ✓ Posterior Bootstrap algorithm suitable for simulator-based models inducing robustness through:
  1. Bayesian Nonparametric Learning (NPL) framework
  2. Minimum Maximum Mean Discrepancy (MMD) estimators
- ✓ Highly parallelisable algorithm
- ✓ Generalisation error, posterior consistency and robustness to outliers guarantees

Limitations:

- ✓ Optimisation; objective usually not convex
- ✓ Kernel choice and associated hyperparameters
- ✓ MMD approximation; estimators with improved sample complexity e.g. Bharti et al. (2023)

Current work

- *Measurement Error* in the covariates
- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E, \quad \mathbb{E}[E] = 0.$$

Goal: to estimate θ_0 when data from (W, Y) is available

- We do **not** have observations from $\mathbb{P}^*$ anymore!
- If our model assumes that there is no ME, how can we be robust?
- Poster later today :)

Current work

- *Measurement Error* in the covariates
- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E, \quad \mathbb{E}[E] = 0.$$

⇓

Goal: to estimate θ_0 when data from (W, Y) is available

- We do **not** have observations from $\mathbb{P}^*$ anymore!
- If our model assumes that there is no ME, how can we be robust?
- Poster later today :)

Current work

- *Measurement Error* in the covariates
- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E, \quad \mathbb{E}[E] = 0.$$

⇓

Goal: to estimate θ_0 when data from (W, Y) is available

- We do **not** have observations from $\mathbb{P}^*$ anymore!
- If our model assumes that there is no ME, how can we be robust?
- Poster later today :)

Current work

- *Measurement Error* in the covariates
- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E, \quad \mathbb{E}[E] = 0.$$

⬇

Goal: to estimate θ_0 when data from (W, Y) is available

- We do **not** have observations from $\mathbb{P}^*$ anymore!
- If our model assumes that there is no ME, how can we be robust?
- Poster later today :)

Current work

- *Measurement Error* in the covariates
- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E, \quad \mathbb{E}[E] = 0.$$

⇓

Goal: to estimate θ_0 when data from (W, Y) is available

- We do **not** have observations from $\mathbb{P}^*$ anymore!
- If our model assumes that there is no ME, how can we be robust?
- Poster later today :)

Current work

- *Measurement Error* in the covariates
- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E, \quad \mathbb{E}[E] = 0.$$

Goal: to estimate θ_0 when data from (W, Y) is available

- We do **not** have observations from $\mathbb{P}^*$ anymore!
- If our model assumes that there is no ME, how can we be robust?
- Poster later today :)

Current work

- *Measurement Error* in the covariates
- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E, \quad \mathbb{E}[E] = 0.$$

Goal: to estimate θ_0 when data from (W, Y) is available

- We do **not** have observations from $\mathbb{P}^*$ anymore!
- If our model assumes that there is no ME, how can we be robust?
- Poster later today :)

References i

Bernton, E., Jacob, P. E., Gerber, M., and Robert, C. P. (2019). Approximate bayesian computation with the wasserstein distance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(2):235–269.

Bharti, A., Naslidnyk, M., Key, O., Kaski, S., and Briol, F.-X. (2023). Optimally-weighted estimators of the maximum mean discrepancy for likelihood-free inference. *arXiv preprint arXiv:2301.11674*.

Bonassi, F. V., You, L., and West, M. (2011). Bayesian learning from marginal data in bionetwork models. *Statistical applications in genetics and molecular biology*, 10(1).

Briol, F.-X., Barp, A., Duncan, A. B., and Girolami, M. (2019). Statistical inference for generative models with maximum mean discrepancy. *arXiv preprint arXiv:1906.05944*.

Chérief-Abdellatif, B.-E. and Alquier, P. (2022). Finite sample properties of parametric mmd estimation: robustness to misspecification and dependence. *Bernoulli*, 28(1):181–213.

Fong, E., Lyddon, S., and Holmes, C. (2019). Scalable nonparametric sampling from multimodal posteriors with the posterior bootstrap. In *International Conference on Machine Learning*, pages 1952–1962. PMLR.

References ii

Gretton, A., Borgwardt, K., Rasch, M. J., Scholkopf, B., and Smola, A. J. (2008). A kernel method for the two-sample problem. *arXiv preprint arXiv:0805.2368*.

Lyddon, S., Walker, S., and Holmes, C. C. (2018). Nonparametric learning from Bayesian models with randomized objective functions. *arXiv preprint arXiv:1806.11544*.

Müller, A. (1997). Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, pages 429–443.

Pacchiardi, L. and Dutta, R. (2021). Generalized bayesian likelihood-free inference using scoring rules estimators. *arXiv preprint arXiv:2104.03889*.

Parr, W. C. and Schucany, W. R. (1980). Minimum distance and robust estimation. *Journal of the American Statistical Association*, 75(371):616–624.

Schmon, S. M., Cannon, P. W., and Knoblauch, J. (2020). Generalized posteriors in approximate bayesian computation. *arXiv preprint arXiv:2011.08644*.