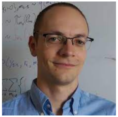
Robust Bayesian Inference for Simulator-based Models via the MMD Posterior Bootstrap

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



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

PHYSTAT-SBI 2024, Max Planck Institute for Physics

Notion of robustness in this talk

Robustness in SBI:

- Model Misspecification [1]
- Measurement Error [2]

[1] Dellaporta, C., Knoblauch, J., Damoulas, T. and Briol, F.X., 2022, May. Robust Bayesian inference for simulator-based models via the MMD posterior bootstrap. In International Conference on Artificial Intelligence and Statistics (pp. 943-970). PMLR.

[2] Dellaporta, C. and Damoulas, T., 2023. Robust Bayesian Inference for Measurement Error Models. arXiv preprint arXiv:2306.01468.

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

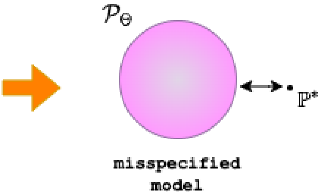
Simulator-based model family

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

$\mathbb{P}_\theta$ 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 $x_{1:n} \sim \mathbb{P}^*$



Main idea

Standard Bayesian Inference:

$$\underbrace{P(\theta|x_1,\ldots,x_n)}_{\text{Posterior}} \propto \underbrace{P(x_1,\ldots,x_n|\theta)}_{\text{Likelihood}} \cdot \underbrace{P(\theta)}_{\text{Prior}}$$

- Likelihood / model is assumed to be correct
- Uncertainty is set directly on θ via the prior!

Bayesian Nonparametric Learning (NPL):

$$\underbrace{\mathbb{P}^\star \sim P(\mathbb{P}^\star)}_{\text{Nonp. Prior on } \mathbb{P}^\star} \longrightarrow \underbrace{P(\mathbb{P}^\star \mid x_{1:n}) \propto P(x_{1:n} \mid \mathbb{P}^\star) \cdot P(\mathbb{P}^\star)}_{\text{Nonp. Posterior on } \mathbb{P}^\star} \longrightarrow \underbrace{\Pi_{\mathrm{NPL}}(\theta)}_{\text{NPL posterior on } \theta}$$

- Model not assumed to be correct
- Uncertainty set directly on $\mathbb{P}^\star$ and propagated to θ

Main idea

Standard Bayesian Inference:

$$\underbrace{P(\theta|x_1,\ldots,x_n)}_{\text{Posterior}} \propto \underbrace{P(x_1,\ldots,x_n|\theta)}_{\text{Likelihood}} \cdot \underbrace{P(\theta)}_{\text{Prior}}$$

- Likelihood / model is assumed to be correct
- Uncertainty is set directly on θ via the prior!

Bayesian Nonparametric Learning (NPL):

$$\underbrace{\mathbb{P}^\star \sim P(\mathbb{P}^\star)}_{\text{Nonp. Prior on } \mathbb{P}^\star} \longrightarrow \underbrace{P(\mathbb{P}^\star \mid x_{1:n}) \propto P(x_{1:n} \mid \mathbb{P}^\star) \cdot P(\mathbb{P}^\star)}_{\text{Nonp. Posterior on } \mathbb{P}^\star} \longrightarrow \underbrace{\Pi_{\mathrm{NPL}}(\theta)}_{\text{NPL posterior on } \theta}$$

- Model not assumed to be correct
- Uncertainty set directly on $\mathbb{P}^\star$ and propagated to θ

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

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

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

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}.

Proposal: Maximum Mean Discrepancy (MMD) based loss

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

- Map θ^* now corresponds to a minimum MMD estimator as in Briol et al. (2019); Chérief-Abdellatif and Alquier (2022)
- The MMD between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ in an RKHS $\mathcal{H}_k$ with associated kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ can be expressed as

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

Proposal: Maximum Mean Discrepancy (MMD) based loss

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

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

- Map θ^* now corresponds to a minimum MMD estimator as in Briol et al. (2019); Chérief-Abdellatif and Alquier (2022)
- The MMD between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ in an RKHS $\mathcal{H}_k$ with associated kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ can be expressed as

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

Proposal: Maximum Mean Discrepancy (MMD) based loss

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

- Map θ^* now corresponds to a minimum MMD estimator as in Briol et al. (2019); Chérief-Abdellatif and Alquier (2022)
- The MMD between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ in an RKHS $\mathcal{H}_k$ with associated kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ can be expressed as

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

Proposal: Maximum Mean Discrepancy (MMD) based loss

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

$$\downarrow$$

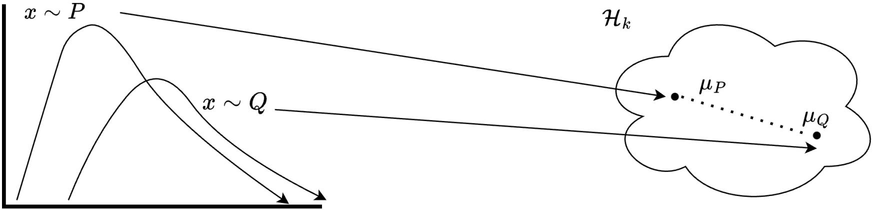
$$\mathrm{MMD}(\mathbb{P}_\theta, \mathbb{P}^*)$$

- Map θ^* now corresponds to a minimum MMD estimator as in Briol et al. (2019); Chérief-Abdellatif and Alquier (2022)
- The MMD between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ in an RKHS $\mathcal{H}_k$ with associated kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ can be expressed as

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

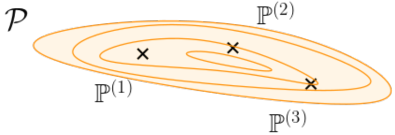
MMD

$$\mathrm{MMD}(P,Q) = \|\mu_P - \mu_Q\|_{\mathcal{H}_k} = \|\mathbb{E}_{x\sim P}[k(x,\cdot)] - \mathbb{E}_{x\sim Q}[k(x,\cdot)]\|_{\mathcal{H}_k}$$



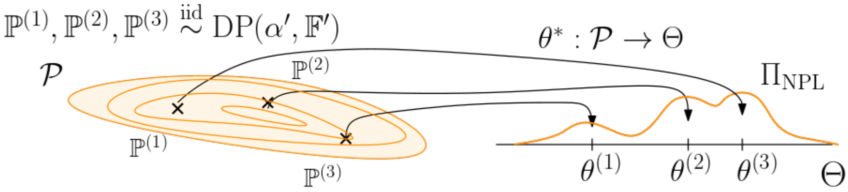
MMD Posterior Bootstrap

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



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

MMD Posterior Bootstrap



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

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

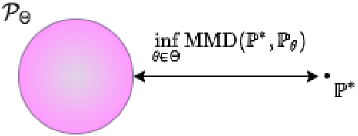
- **Assumption:** $\sup_{x,x'\in\mathcal{X}} |k(x,x')| < \infty$, W.L.O.G. $|k(x,x')| \leq 1 \quad \forall \quad x,x' \in \mathcal{X}$
- $\nu := DP(\alpha', \mathbb{F}')$

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

Generalisation error

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

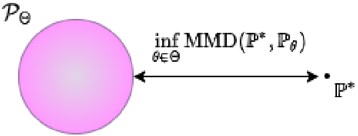
$$\inf_{\theta \in \Theta} \mathrm{MMD}(\mathbb{P}^*, \mathbb{P}_\theta) \quad \le \quad \mathbb{E}_{x_{1:n} \overset{\text{iid}}{\sim} \mathbb{P}^*} \left[\mathbb{E}_{\mathbb{P} \sim \nu} \left[\mathrm{MMD}(\mathbb{P}^*, \mathbb{P}_{\theta^*(\mathbb{P})}) \right] \right]$$



Generalisation error

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

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

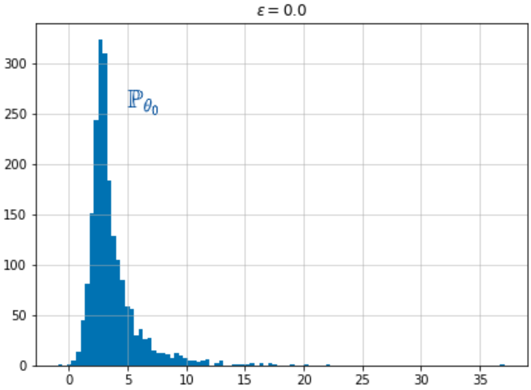


Extensions:

- Robustness to outliers for contaminated model $\mathbb{P}^* = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{Q}$
- Posterior consistency in the frequentist sense

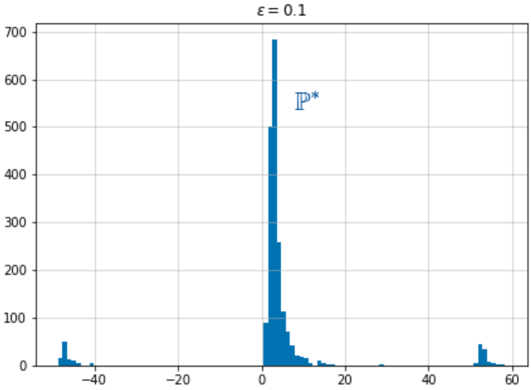
Example: Contaminated G-and-k distribution model

$\mathbb{P}^* = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{Q}$ 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 $\mathbb{Q} = \mathbb{P}_{\theta_0} \pm 50$



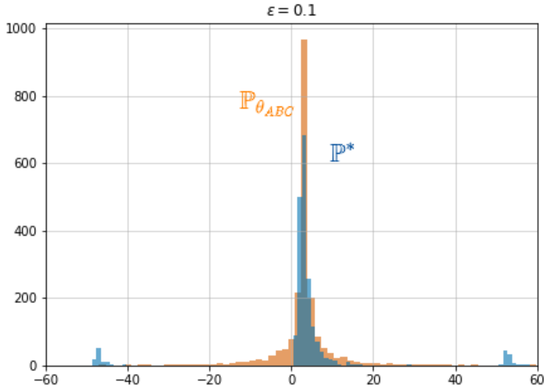
Example: Contaminated G-and-k distribution model

$\mathbb{P}^* = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{Q}$ 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 $\mathbb{Q} = \mathbb{P}_{\theta_0} \pm 50$



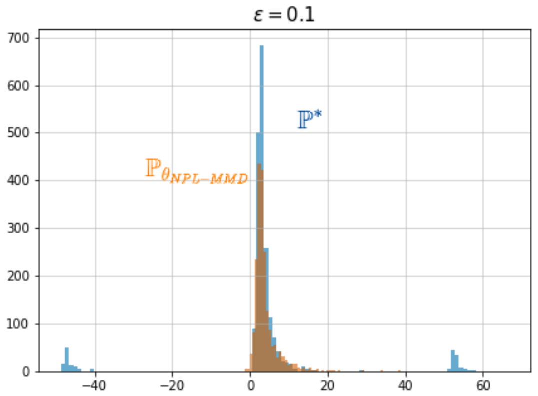
Example: Contaminated G-and-k distribution model

$\mathbb{P}^* = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{Q}$ 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

$\mathbb{P}^* = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{Q}$ 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

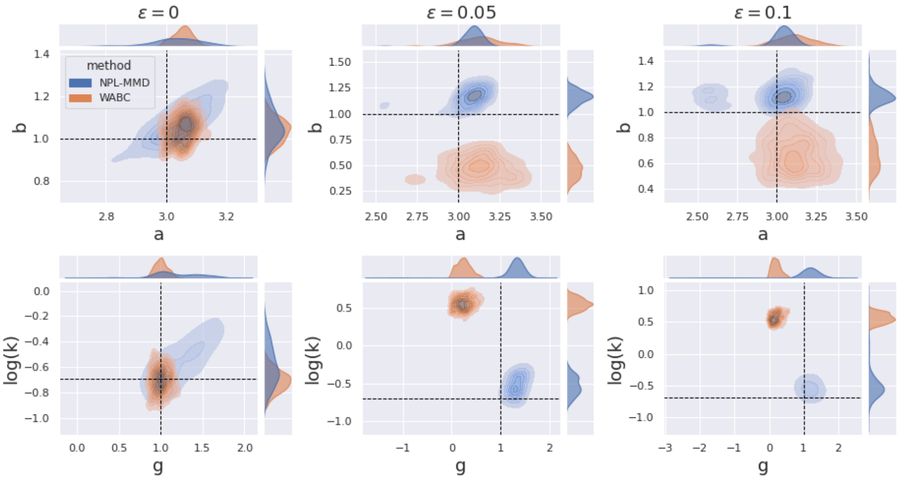


Figure: Comparison of posterior marginal distributions obtained using the MMD Posterior Bootstrap (NPL-MMD) and the Wasserstein-ABC (WABC) method in Bernton et al. (2019).

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 $u_i = (u_{i,1,1}, u_{i,1,2}, \ldots, u_{i,T,1}, u_{i,T,2}, u_{i,T+1,1})^\top \sim \text{Unif}([0,1]^{2T+1})$ 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

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

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

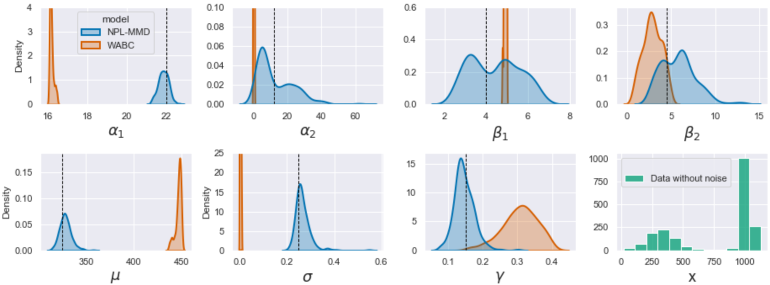
$$\begin{aligned}
\tilde{v}_{i,t+1} &= v_{i,t} + \frac{\alpha_1}{1+w_{i,t}^{\beta_1}} - (1 + 0.03v_{i,t}) \\
\tilde{w}_{i,t+1} &= w_{i,t} + \frac{\alpha_2}{1+v_{i,t}^{\beta_2}} - (1 + 0.03w_{i,t}) \\
v_{i,t+1} &= \tilde{v}_{i,t+1} + 0.5\Phi^{-1}(\Phi(-2\tilde{v}_{i,t+1}) + u_{i,t,1}(1 - \Phi(-2\tilde{v}_{i,t+1})))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

Example: Toggle switch model with Cauchy noise

- 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.



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
- ✓ Parallelisable algorithm
- ✓ Generalisation error, posterior consistency and robustness to outliers guarantees

Limitations:

- ✓ Optimisation; objective usually not convex, kernel choice and associated hyperparameters
- ✓ MMD scales well with dimensions but optimisation needs to be repeated a lot of times
- ✓ MMD approximation; estimators with improved sample complexity e.g. Bharti et al. (2023)
- ✓ Larger scale applications to follow

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
- ✓ Parallelisable algorithm
- ✓ Generalisation error, posterior consistency and robustness to outliers guarantees

Limitations:

- ✓ Optimisation; objective usually not convex, kernel choice and associated hyperparameters
- ✓ MMD scales well with dimensions but optimisation needs to be repeated a lot of times
- ✓ MMD approximation; estimators with improved sample complexity e.g. Bharti et al. (2023)
- ✓ Larger scale applications to follow

Extension to Measurement Error models

- *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: Estimate θ_0 when realisations of (W, Y) are available and true distributions of N and E are unknown (prior beliefs might be available)

Extension to Measurement Error models

- *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: Estimate θ_0 when realisations of (W, Y) are available and true distributions of N and E are unknown (prior beliefs might be available)

Extension to Measurement Error models

- *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: Estimate θ_0 when realisations of (W, Y) are available and true distributions of N and E are unknown (prior beliefs might be available)

Extension to Measurement Error models

- *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:

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

$$\Downarrow$$

Goal: Estimate θ_0 when realisations of (W, Y) are available and true distributions of N and E are unknown (prior beliefs might be available)

Extension to Measurement Error models

- *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: Estimate θ_0 when realisations of (W, Y) are available and true distributions of N and E are unknown (prior beliefs might be available)

Measurement Error

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

$$\theta_l^*(\mathbb{P}^\star_{X,Y}) = \arg\min_{\theta \in \Theta} \mathbb{E}_{(X,Y)\sim\mathbb{P}^\star_{X,Y}}[l(x,y;\theta)]$$

- Set uncertainty on the true distribution of $X \mid W = w_i$, denoted by $\mathbb{P}^*_{X \mid w_i}$
- Dirichlet Process (DP) prior for each $i = 1, \dots, n$:

$$\mathbb{P}_i \sim \mathrm{DP}(c, \mathbb{F}_{w_i})$$

- DP posterior:

$$\mathbb{P}_i \mid w_i \sim \mathrm{DP}(c+1, \mathbb{F}'_{w_i}), \quad \mathbb{F}'_{w_i} = \frac{1}{c+1}\delta_{w_i} + \frac{c}{c+1}\mathbb{F}_{w_i}$$

- Preprint under review [3]

[3] Dellaporta, C. and Damoulas, T., 2023. Robust Bayesian Inference for Measurement Error Models. arXiv preprint arXiv:2306.01468.

References I

Thank you! contact: c.dellaporta@warwick.ac.uk

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.

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