Cost-aware simulation-based inference

Dr François-Xavier Briol
Department of Statistical Science
University College London
https://fxbriol.github.io/
https://fsml-ucl.github.io/

Ayush Bharti

Daolang Huang

Sami Kaski

Simulation-based inference (SBI)

Bayesian inference:

$$p(\theta \mid y_1, \ldots, y_n) \propto \prod_{i=1}^{n} p(y_i \mid \theta) p(\theta)$$

Simulation-based inference (SBI)

Bayesian inference:

$$p(\theta \mid y_1, \ldots, y_n) \propto \prod_{i=1}^{n} p(y_i \mid \theta) p(\theta)$$

Simulation-based inference (SBI)

Bayesian inference:

$$p(\theta \mid y_1, \ldots, y_n) \propto \prod_{i=1}^{n} p(y_i \mid \theta) p(\theta)$$

Two main approaches:

- Approximate Bayesian computation (ABC).

Beaumont, M. A. (2019). Approximate Bayesian computation. *Annual Review of Statistics and Its Application*, *6*, 379–403.

Simulation-based inference (SBI)

Bayesian inference:

$$p(\theta \mid y_1, \ldots, y_n) \propto \prod_{i=1}^{n} p(y_i \mid \theta) p(\theta)$$

Two main approaches:

- Approximate Bayesian computation (ABC).
- Neural-based simulation-based inference.

Beaumont, M. A. (2019). Approximate Bayesian computation. *Annual Review of Statistics and Its Application*, *6*, 379–403.

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(48).

Challenge for SBI

Simulators can be really computationally expensive!

Challenge for SBI

Simulators can be really computationally expensive!

- Most simulators used in SBI papers take only a few seconds (or less) to run.

Challenge for SBI

Simulators can be really computationally expensive!

- Most simulators used in SBI papers take only a few seconds (or less) to run.
- Even if a simulator takes only a few minutes, we typically need thousands of simulations!

Challenge for SBI

Simulators can be really computationally expensive!

- Most simulators used in SBI papers take only a few seconds (or less) to run.
- Even if a simulator takes only a few minutes, we typically need thousands of simulations!
- Simulators that take more time are currently out of reach of existing methods.

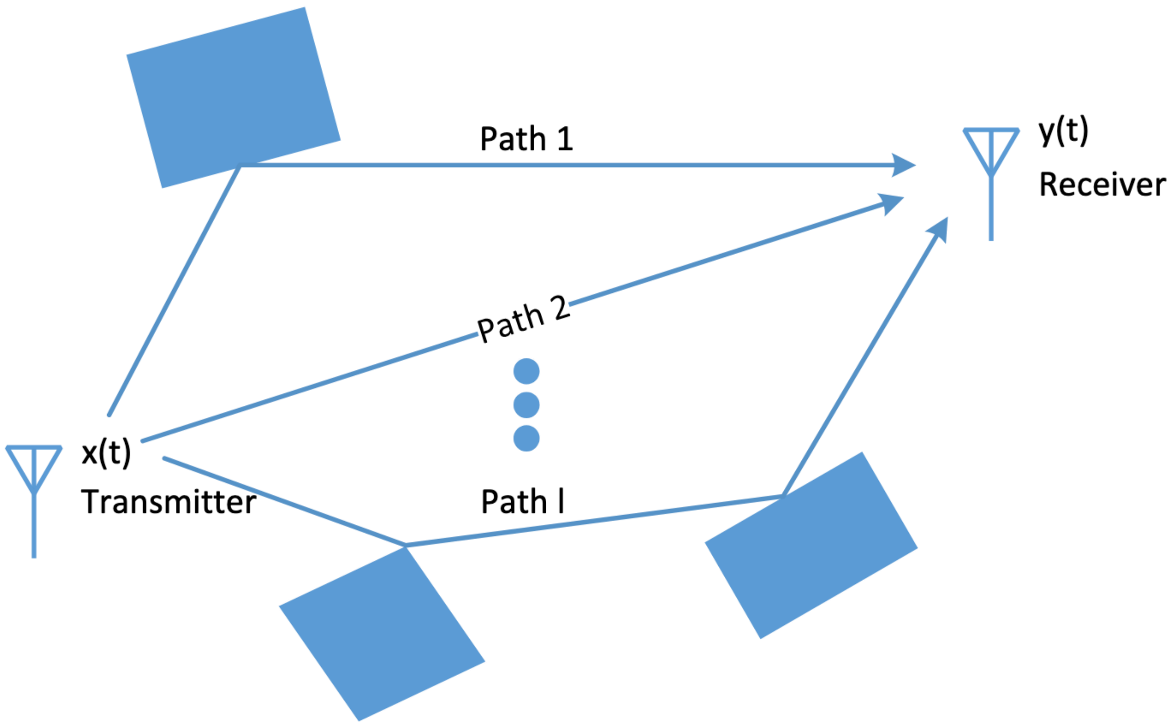
Challenge for SBI

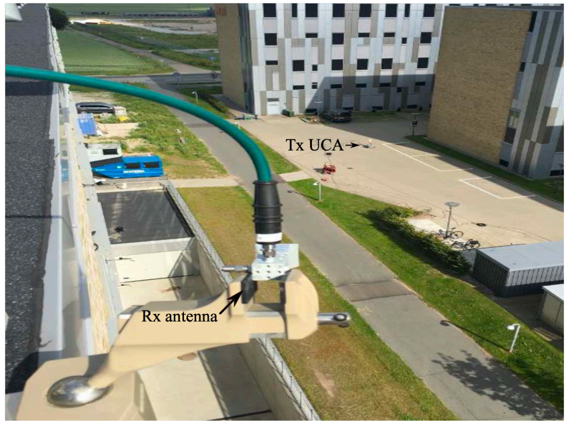
Simulators can be really computationally expensive!

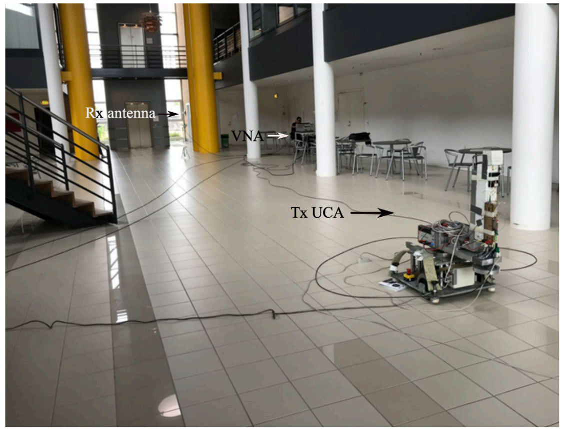
- Most simulators used in SBI papers take only a few seconds (or less) to run.
- Even if a simulator takes only a few minutes, we typically need thousands of simulations!
- Simulators that take more time are currently out of reach of existing methods.

→ A great playground for computational statisticians!

SBI for radio-propagation

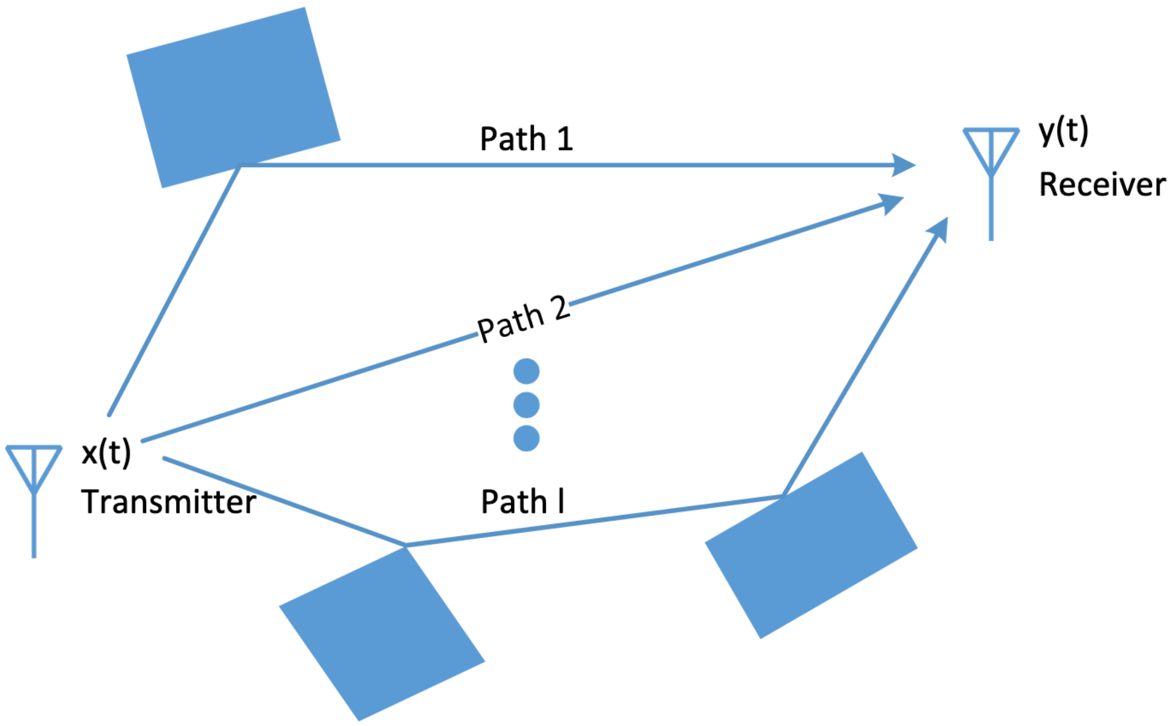


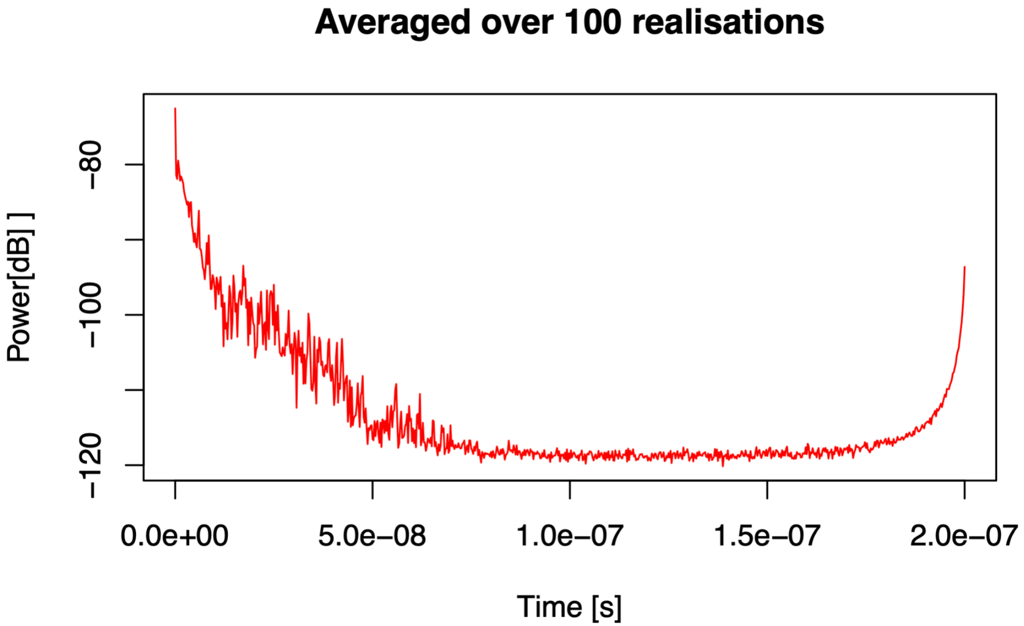




Bharti, A., **Briol, F-X.**, Pedersen, T. (2022). A general method for calibrating stochastic radio channel models with kernels. IEEE Transactions on Antennas and Propagation, vol. 70, no. 6, pp. 3986-4001, June 2022.

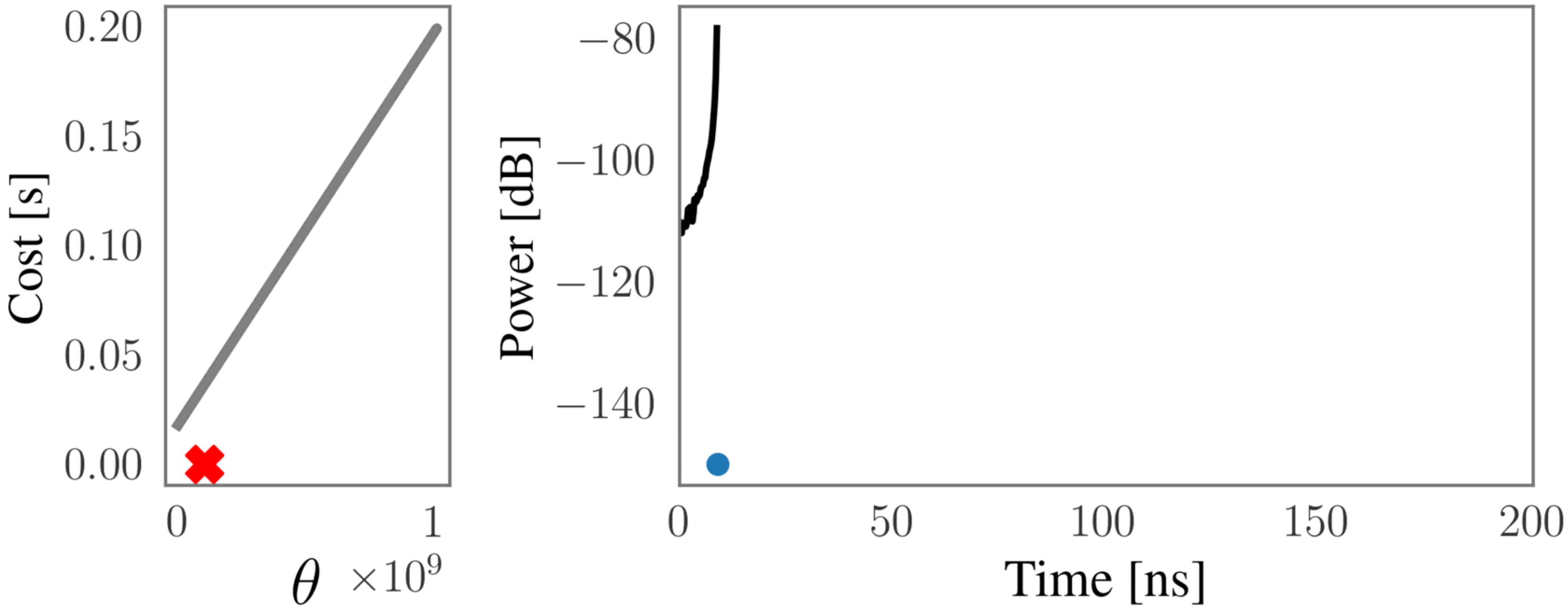
SBI for radio-propagation





Bharti, A., **Briol, F-X.**, Pedersen, T. (2022). A general method for calibrating stochastic radio channel models with kernels. IEEE Transactions on Antennas and Propagation, vol. 70, no. 6, pp. 3986-4001, June 2022.

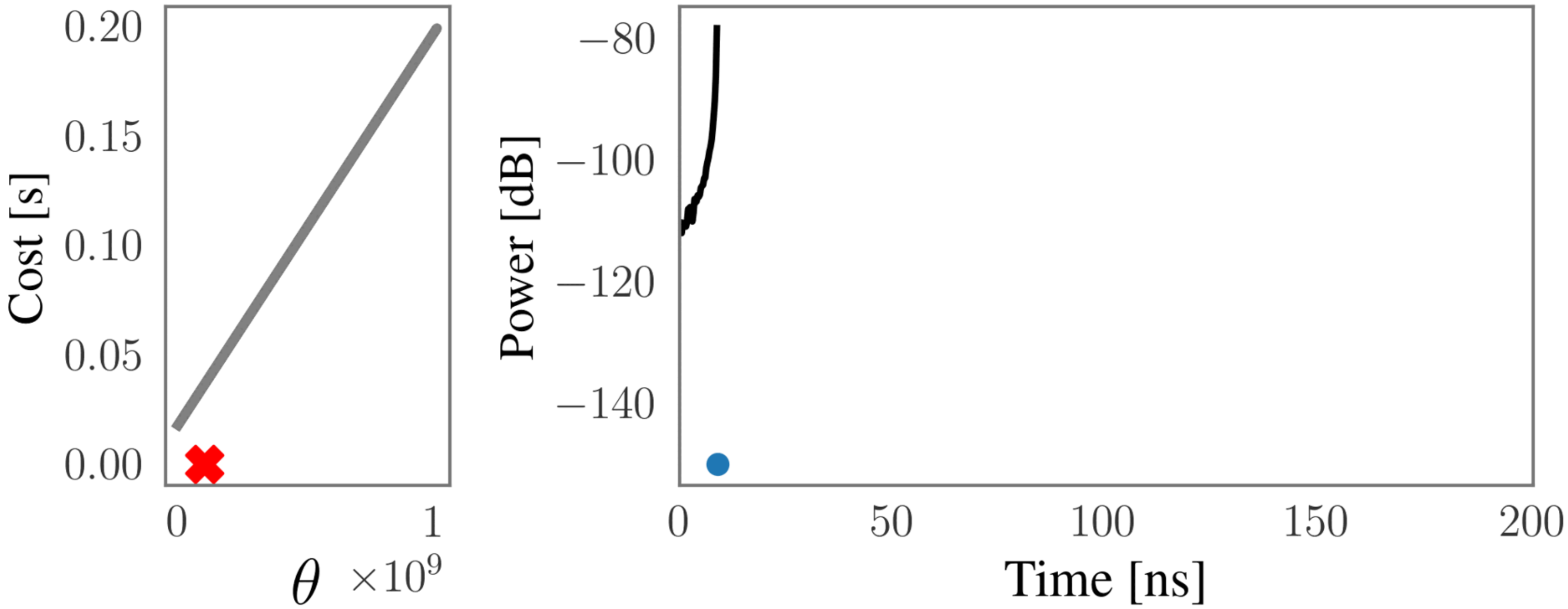
The cost of simulations is not constant…



Rate parameter of a Poisson process!

[https://fxbriol.github.io/images/ca-SBI.mp4]

The cost of simulations is not constant…



Rate parameter of a Poisson process!

[https://fxbriol.github.io/images/ca-SBI.mp4]

Neural likelihood estimation (NLE)

- **Step 1:** train a conditional density model $q_\phi(\cdot \mid \theta)$ to approximate the likelihood using samples from the prior ($\theta_1, \ldots, \theta_n \sim \pi$) and simulator ($\mathbf{x}_i \sim p(\cdot \mid \theta_i)$):

$$\arg\min_{\phi} \ell_{\mathrm{NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n} \log q_{\phi}(\mathbf{x}_i \mid \theta_i) \approx -\mathbb{E}_{\theta \sim p(\theta)}[\mathbb{E}_{\mathbf{x} \sim \mathbb{P}_\theta}[\log q_{\phi}(\mathbf{x} \mid \theta)]]$$

Neural likelihood estimation (NLE)

- **Step 1:** train a conditional density model $q_\phi(\cdot \mid \theta)$ to approximate the likelihood using samples from the prior ($\theta_1, \ldots, \theta_n \sim \pi$) and simulator ($\mathbf{x}_i \sim p(\cdot \mid \theta_i)$):

$$\arg\min_{\phi} \ell_{\mathrm{NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n} \log q_\phi(\mathbf{x}_i \mid \theta_i) \approx -\mathbb{E}_{\theta\sim p(\theta)}[\mathbb{E}_{\mathbf{x}\sim\mathbb{P}_\theta}[\log q_\phi(\mathbf{x} \mid \theta)]]$$

- **Step 2:** Do Bayes with approximate likelihood!

$$p(\theta \mid y_1, \ldots, y_n) \propto \prod_{i=1}^{n} q_{\phi^*}(y_i \mid \theta) p(\theta)$$

Neural likelihood estimation (NLE)

- **Step 1:** train a conditional density model $q_\phi(\cdot\,|\,\theta)$ to approximate the likelihood using samples from the prior ($\theta_1, \ldots, \theta_n \sim \pi$) and simulator ($\mathbf{x}_i \sim p(\cdot\,|\,\theta_i)$):

$$\arg\min_{\phi} \ell_{\text{NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n} \log q_\phi(\mathbf{x}_i\,|\,\theta_i) \approx -\mathbb{E}_{\theta\sim p(\theta)}[\mathbb{E}_{\mathbf{x}\sim\mathbb{P}_\theta}[\log q_\phi(\mathbf{x}\,|\,\theta)]]$$

Typically the most **computationally expensive** step!!

- **Step 2:** Do Bayes with approximate likelihood!

$$p(\theta\,|\,y_1, \ldots, y_n) \propto \prod_{i=1}^{n} q_{\phi^*}(y_i\,|\,\theta)p(\theta)$$

Neural likelihood estimation (NLE)

- **Step 1:** train a conditional density model $q_\phi(\cdot\,|\,\theta)$ to approximate the likelihood using samples from the prior ($\theta_1, \ldots, \theta_n \sim \pi$) and simulator ($\mathbf{x}_i \sim p(\cdot\,|\,\theta_i)$):

$$\arg\min_{\phi} \ell_{\text{NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n} \log q_{\phi}(\mathbf{x}_i\,|\,\theta_i) \approx -\mathbb{E}_{\theta\sim p(\theta)}[\mathbb{E}_{\mathbf{x}\sim\mathbb{P}_\theta}[\log q_\phi(\mathbf{x}\,|\,\theta)]]$$

Typically the most **computationally expensive** step!!

- **Step 2:** Do Bayes with approximate likelihood!

$$p(\theta\,|\,y_1, \ldots, y_n) \propto \prod_{i=1}^{n} q_{\phi^*}(y_i\,|\,\theta)p(\theta)$$

- Can do similarly and approximate a posterior….. Neural posterior estimation (NPE).

A cheaper step 1?

$$\ell_{\mathrm{NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n} \log q_\phi(\mathbf{x}_i \,|\, \theta_i) \approx -\mathbb{E}_{\theta\sim p(\theta)}[\mathbb{E}_{\mathbf{x}\sim\mathbb{P}_\theta}[\log q_\phi(\mathbf{x} \,|\, \theta)]]$$

Can we do this better/cheaper?!

A cheaper step 1?

$$\ell_{\text{NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n}\log q_\phi(\mathbf{x}_i \,|\, \theta_i) \approx -\mathbb{E}_{\theta\sim p(\theta)}[\mathbb{E}_{\mathbf{x}\sim\mathbb{P}_\theta}[\log q_\phi(\mathbf{x}\,|\,\theta)]]$$

Can we do this better/cheaper?!

Idea:

- Let's make use of the cost function $c : \Theta \rightarrow \mathbb{R}$.

A cheaper step 1?

$$\ell_{\text{NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n} \log q_\phi(\mathbf{x}_i \mid \theta_i) \approx -\mathbb{E}_{\theta \sim p(\theta)}[\mathbb{E}_{\mathbf{x} \sim \mathbb{P}_\theta}[\log q_\phi(\mathbf{x} \mid \theta)]]$$

Can we do this better/cheaper?!

<u>Idea:</u>

- Let's make use of the cost function $c : \Theta \to \mathbb{R}$.
- We can try to sample less often in expensive regions ……
  but we still want to target the right objective.

Importance sampling

$$\mu = \int_{\Theta} f(\theta)\pi(\theta)d\theta$$

Importance sampling

$$\mu = \int_{\Theta} f(\theta)\pi(\theta)d\theta \quad = \int_{\Theta} f(\theta)\frac{\pi(\theta)}{\tilde{\pi}(\theta)}\tilde{\pi}(\theta)d\theta$$

Importance sampling

$$\mu = \int_\Theta f(\theta)\pi(\theta)d\theta \quad = \int_\Theta f(\theta)\frac{\pi(\theta)}{\tilde{\pi}(\theta)}\tilde{\pi}(\theta)d\theta$$

$$\approx \sum_{i=1}^{N} w(\theta_i)f(\theta_i) \qquad \theta_1, \ldots, \theta_N \sim \tilde{\pi}$$

Importance sampling

$$\mu = \int_\Theta f(\theta)\pi(\theta)d\theta \quad = \int_\Theta f(\theta)\frac{\pi(\theta)}{\tilde{\pi}(\theta)}\tilde{\pi}(\theta)d\theta$$

$$\approx \sum_{i=1}^{N} w(\theta_i)f(\theta_i) \qquad \theta_1, \ldots, \theta_N \sim \tilde{\pi}$$

$$w_{\text{IS}}(\theta_i) = \frac{1}{N}\frac{\pi(\theta_i)}{\tilde{\pi}(\theta_i)}$$

Importance sampling

$$\mu = \int_\Theta f(\theta)\pi(\theta)d\theta \quad = \int_\Theta f(\theta)\frac{\pi(\theta)}{\tilde{\pi}(\theta)}\tilde{\pi}(\theta)d\theta$$

$$\approx \sum_{i=1}^{N} w(\theta_i)f(\theta_i) \qquad \theta_1, \ldots, \theta_N \sim \tilde{\pi}$$

$$w_{\text{IS}}(\theta_i) = \frac{1}{N}\frac{\pi(\theta_i)}{\tilde{\pi}(\theta_i)} \qquad w_{\text{SNIS}}(\theta_i) = \frac{w_{\text{IS}}(\theta_i)}{\sum_{j=1}^{N} w_{\text{IS}}(\theta_j)}$$

Importance sampling

$$\mu = \int_\Theta f(\theta)\pi(\theta)d\theta \quad = \int_\Theta f(\theta)\frac{\pi(\theta)}{\tilde{\pi}(\theta)}\tilde{\pi}(\theta)d\theta$$

$$\approx \sum_{i=1}^{N} w(\theta_i)f(\theta_i) \qquad \theta_1, \ldots, \theta_N \sim \tilde{\pi}$$

$$w_{\text{IS}}(\theta_i) = \frac{1}{N}\frac{\pi(\theta_i)}{\tilde{\pi}(\theta_i)} \qquad w_{\text{SNIS}}(\theta_i) = \frac{w_{\text{IS}}(\theta_i)}{\sum_{j=1}^{N} w_{\text{IS}}(\theta_j)}$$

Question: How do you pick the importance distribution?

Cost-aware importance sampling

$$\tilde{\pi}_g(\theta) \propto \frac{\pi(\theta)}{g(c(\theta))},$$

Cost-aware importance sampling

$$\tilde{\pi}_g(\theta) \propto \frac{\boxed{\pi(\theta)}}{g(c(\theta))},$$

We want a distribution similar to our target π

Cost-aware importance sampling

$$\tilde{\pi}_g(\theta) \propto \frac{\pi(\theta)}{g(\boxed{c(\theta)})},$$

We do not want to sample often where the cost is large!

Cost-aware importance sampling

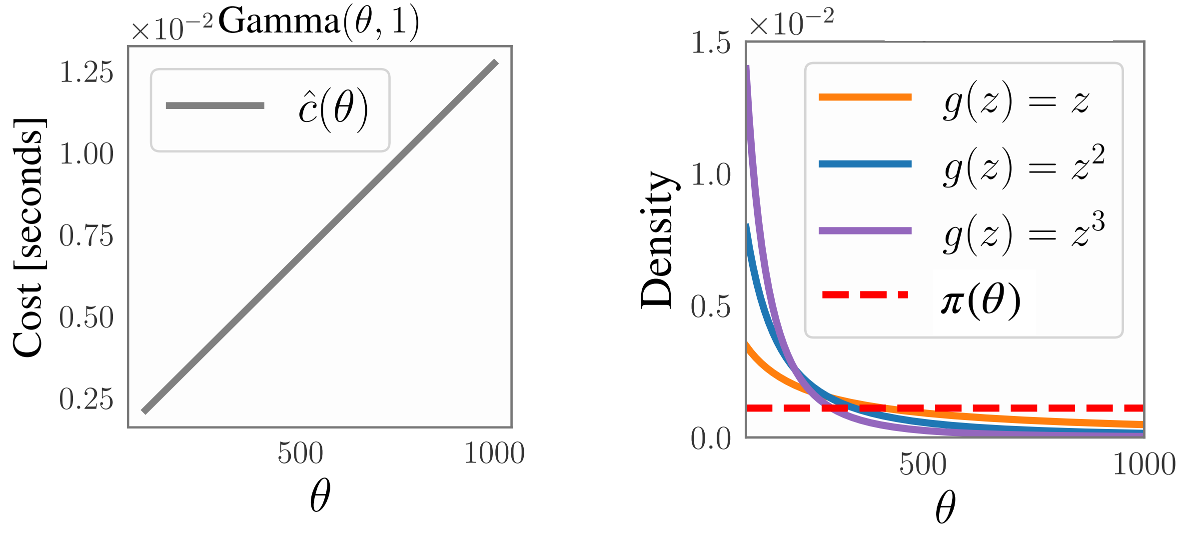
$$\tilde{\pi}_g(\theta) \propto \frac{\pi(\theta)}{\boxed{g}(c(\theta))},$$

$g : (0,\infty) \to (0,\infty)$ taken to be non-decreasing.

Represents how much we dislike ‘expensive’ parameters!

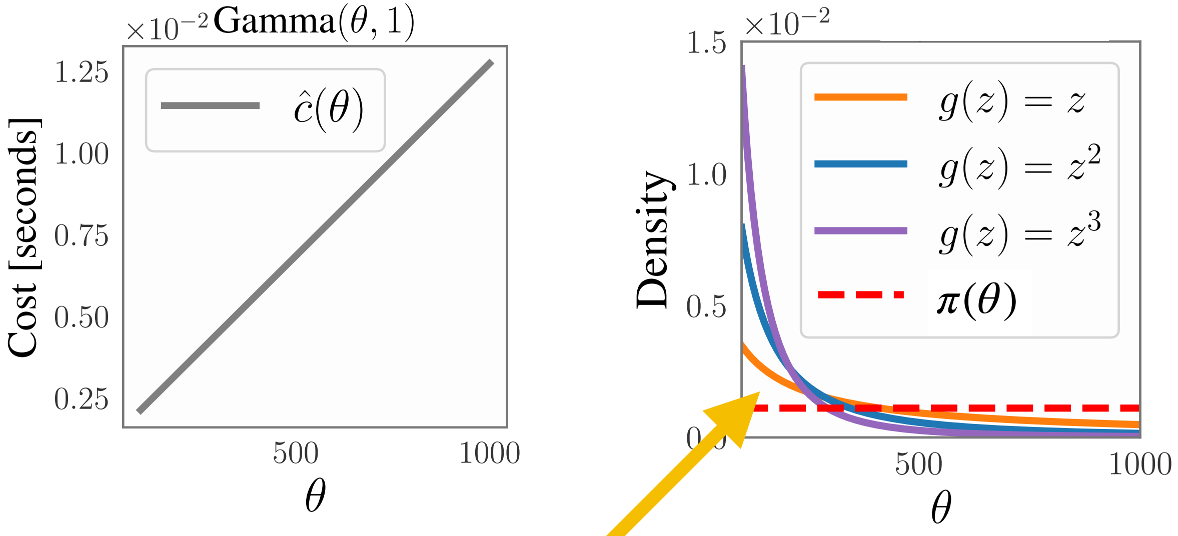
Cost-aware importance sampling

$$\tilde{\pi}_g(\theta) \propto \frac{\pi(\theta)}{g(c(\theta))},$$



Cost-aware importance sampling

$$\tilde{\pi}_g(\theta) \propto \frac{\pi(\theta)}{g(c(\theta))},$$



Upweight cheap region!

Cost-aware importance sampling

$$w(\theta) = \frac{1}{N}\frac{\pi(\theta)}{\tilde{\pi}_g(\theta)} = \frac{B\pi(\theta)g(c(\theta))}{N\pi(\theta)} \propto g(c(\theta))$$

Through $\tilde{\pi}_g$, we sample less often from expensive regions, so we need to up-weight expensive samples.

Cost-aware importance sampling

$$w(\theta) = \frac{1}{N}\frac{\pi(\theta)}{\tilde{\pi}_g(\theta)} = \frac{B\pi(\theta)g(c(\theta))}{N\pi(\theta)} \propto g(c(\theta))$$

$$w_{\mathrm{Ca}}(\theta_i) = \frac{w(\theta_i)}{\sum_{j=1}^{n} w(\theta_j)} = \frac{g(c(\theta_i))}{\sum_{j=1}^{n} g(c(\theta_j))}$$

We use SNIS weights

$$\mu = \int_\Theta f(\theta)\pi(\theta)d\theta \approx \sum_{i=1}^{n} w_{\mathrm{Ca}}(\theta_i)f(\theta_i) = \hat{\mu}_n^{\mathrm{Ca}}$$

Sampling from the cost-aware proposal

- We can use rejection sampling!

Sampling from the cost-aware proposal

- We can use rejection sampling!

Repeat until n samples are accepted:

1. Sample $\theta^\star \sim \pi(\theta)$.
2. Accept $\theta^\star$ as a sample from $\tilde{\pi}_g$ with probability $A(\theta)$.

Sampling from the cost-aware proposal

- We can use rejection sampling!

Repeat until n samples are accepted:

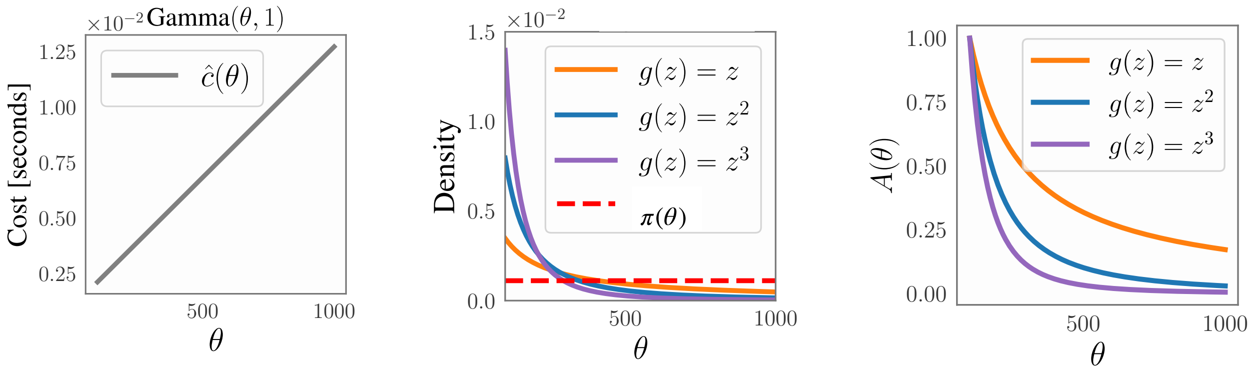
1. Sample $\theta^\star \sim \pi(\theta)$.
2. Accept $\theta^\star$ as a sample from $\tilde{\pi}_g$ with probability $A(\theta)$.

Proposition: Assume $g_{\min} := \inf_{\theta \in \Theta} g(c(\theta)) > 0$. Then

- $\tilde{\pi}_g$ is a density.
- The correct acceptance probability is $A(\theta) = \dfrac{g_{\min}}{g(c(\theta))}$

Sampling from the cost-aware proposal

- We can use rejection sampling!



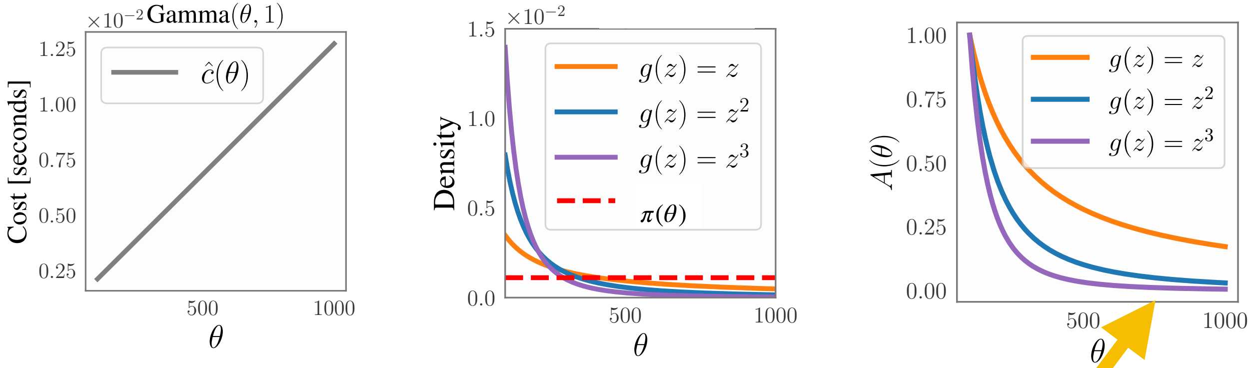
Repeat until n samples are accepted:

1. Sample $\theta^\star \sim \pi(\theta)$.
2. Accept $\theta^\star$ as a sample from $\tilde{\pi}_g$ with probability $A(\theta)$.

Proposition: Assume $g_{\min} := \inf_{\theta \in \Theta} g(c(\theta)) > 0$. Then

- $\tilde{\pi}_g$ is a density.
- The correct acceptance probability is $A(\theta) = \dfrac{g_{\min}}{g(c(\theta))}$

Sampling from the cost-aware proposal



- We can use rejection sampling!

Repeat until n samples are accepted:

1. Sample $\theta^\star \sim \pi(\theta)$.
2. Accept $\theta^\star$ as a sample from $\tilde{\pi}_g$ with probability $A(\theta)$.

Being cost-averse decreases acceptance prob!

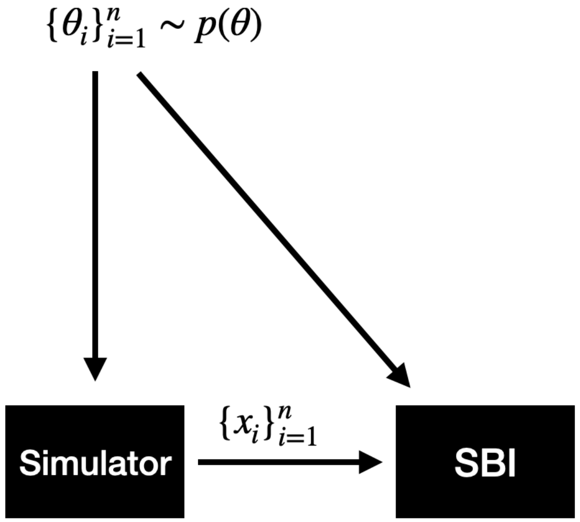
Proposition: Assume $g_{\min} := \inf_{\theta \in \Theta} g(c(\theta)) > 0$. Then

- $\tilde{\pi}_g$ is a density.
- The correct acceptance probability is $A(\theta) = \dfrac{g_{\min}}{g(c(\theta))}$

Putting it all together!

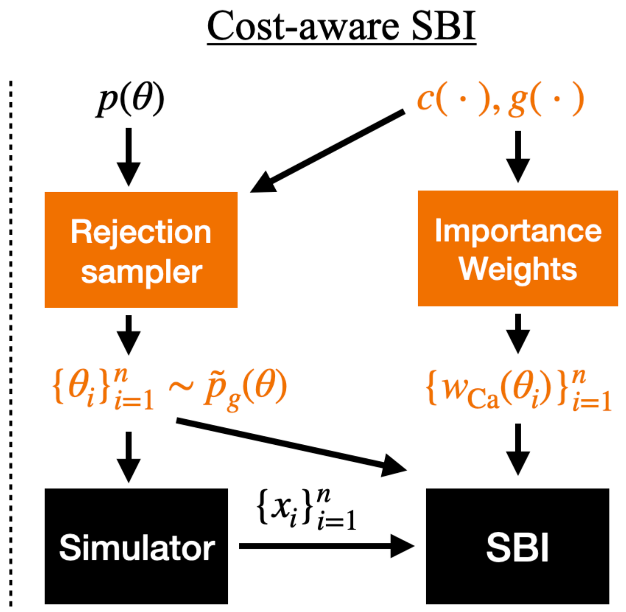
$$\ell_{\text{NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n} \log q_{\phi}(\mathbf{x}_i \,|\, \theta_i), \qquad \theta_i \sim p(\theta), \mathbf{x}_i \sim p(\,\cdot\,|\,\theta)$$

Standard SBI



Putting it all together!

$$\ell_{\text{Ca-NLE}}(\phi) = -\frac{1}{n}\sum_{i=1}^{n} w_{\text{Ca}}(\theta_i)\log q_\phi(\mathbf{x}_i \mid \theta_i), \qquad \theta_i \sim \tilde{p}_g(\theta), \mathbf{x}_i \sim p(\,\cdot \mid \theta)$$



Some reassuring results

Importance sampling can have infinite variance!!!

Some reassuring results

- Suppose that $g_{\max} = \sup_{\theta \in \Theta} g(c(\theta)) < \infty$. Then:

Some reassuring results

- Suppose that $g_{\max} = \sup_{\theta \in \Theta} g(c(\theta)) < \infty$. Then:
  1. The weights are bounded: $\dfrac{g_{\text{min}}}{n g_{\text{max}}} \leq w_{\text{Ca}}(\theta_i) \leq \dfrac{g_{\text{max}}}{n g_{\text{min}}} \qquad \forall i \in \{1, \ldots, n\},$

Some reassuring results

- Suppose that $g_{\max} = \sup_{\theta \in \Theta} g(c(\theta)) < \infty$. Then:

2. If f is square-integrable; i.e. $\int_\Theta f(\theta)^2 \pi(\theta) d\theta < \infty$, then $\mathrm{Var}(\hat{\mu}_{\mathrm{Ca}}) = \sigma^2_{\mathrm{Ca}}$ where:

$$\frac{g_{\min}}{g_{\max}} \left(\sigma^2_{\mathrm{MC}} - \frac{\mu^2}{n} \right) \leq \sigma^2_{\mathrm{Ca}} \leq \frac{g_{\max}}{g_{\min}} \left(\sigma^2_{\mathrm{MC}} - \frac{\mu^2}{n} \right).$$

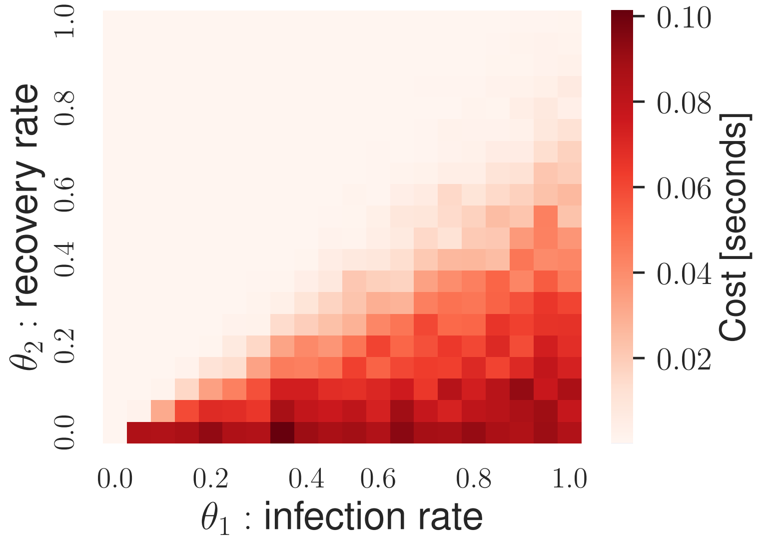
Some reassuring results

- Suppose that $g_{\max} = \sup_{\theta \in \Theta} g(c(\theta)) < \infty$. Then:

3. The ESS is bounded: $\left(\frac{g_{\text{min}}}{g_{\text{max}}}\right)^2 \leq \text{ESS} \leq \left(\frac{g_{\text{max}}}{g_{\text{min}}}\right)^2$.

Some epidemiological models

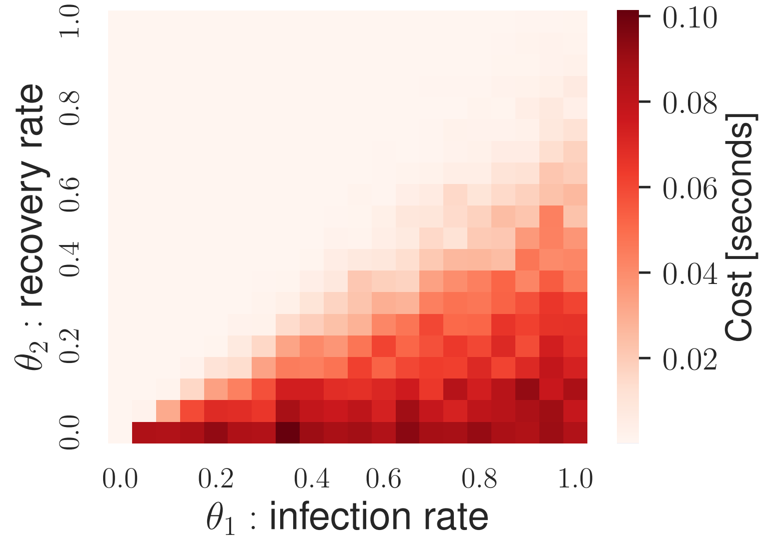
- We consider three different models with 1,2 and 3 parameters respectively, and use NPE.



Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. Mathematical Biosciences, 287:42–53.

Some epidemiological models

- We consider three different models with 1,2 and 3 parameters respectively, and use NPE.

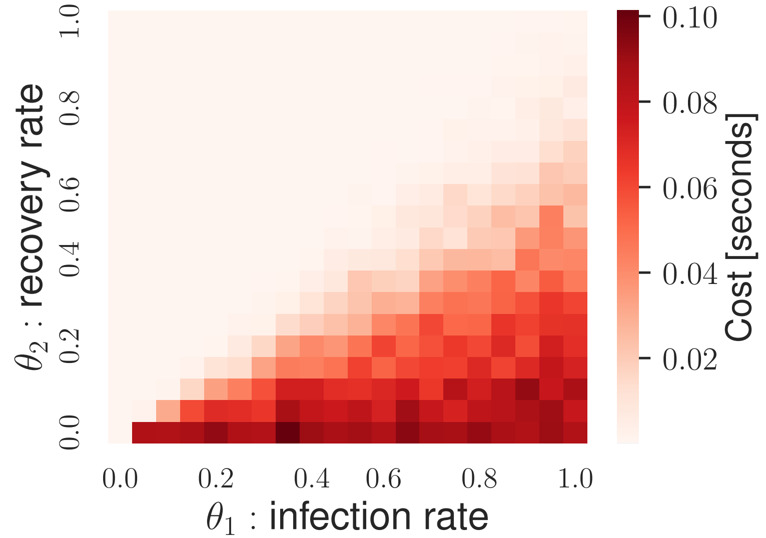


	$\widehat{\mathrm{MMD}}^2$ (↓)					Time saved (↑)			
	NPE	Ca-NPE $g(z)=z^{0.5}$	Ca-NPE $g(z)=z$	Ca-NPE $g(z)=z^2$	Ca-NPE multiple	Ca-NPE $g(z)=z^{0.5}$	Ca-NPE $g(z)=z$	Ca-NPE $g(z)=z^2$	Ca-NPE multiple
Homogen.	0.02(0.02)	0.02(0.01)	0.02(0.02)	0.23(0.08)	0.05(0.04)	16%(2)	38%(2)	70%(2)	30%(5)
Temporal	0.03(0.03)	0.06(0.03)	0.07(0.03)	0.07(0.03)	0.05(0.04)	36%(4)	65%(2)	85%(1)	24%(5)
Bernoulli	0.02(0.00)	0.02(0.00)	0.02(0.01)	0.04(0.01)	0.02(0.00)	23%(4)	37%(4)	47%(3)	25%(6)

Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. Mathematical Biosciences, 287:42–53.

Some epidemiological models

- We consider three different models with 1,2 and 3 parameters respectively, and use NPE.



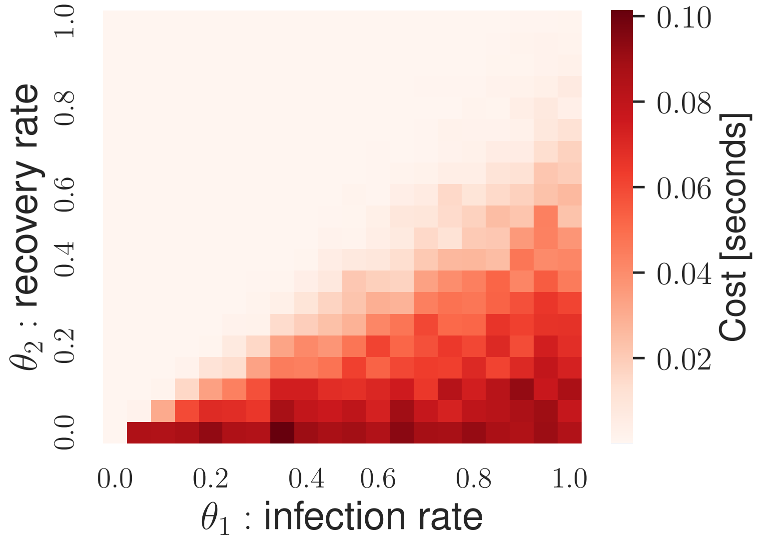
$g(z) = z^{0.5}$: Same accuracy but modest improvement!

	$\widehat{\text{MMD}}^2$ (↓)					Time saved (↑)			
	NPE	Ca-NPE $g(z)=z^{0.5}$	Ca-NPE $g(z)=z$	Ca-NPE $g(z)=z^2$	Ca-NPE multiple	Ca-NPE $g(z)=z^{0.5}$	Ca-NPE $g(z)=z$	Ca-NPE $g(z)=z^2$	Ca-NPE multiple
Homogen.	0.02(0.02)	0.02(0.01)	0.02(0.02)	0.23(0.08)	0.05(0.04)	16%(2)	38%(2)	70%(2)	30%(5)
Temporal	0.03(0.03)	0.06(0.03)	0.07(0.03)	0.07(0.03)	0.05(0.04)	36%(4)	65%(2)	85%(1)	24%(5)
Bernoulli	0.02(0.00)	0.02(0.00)	0.02(0.01)	0.04(0.01)	0.02(0.00)	23%(4)	37%(4)	47%(3)	25%(6)

Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. Mathematical Biosciences, 287:42–53.

Some epidemiological models

- We consider three different models with 1,2 and 3 parameters respectively, and use NPE.



$g(z) = z$: Still same accuracy but slightly better improvement!

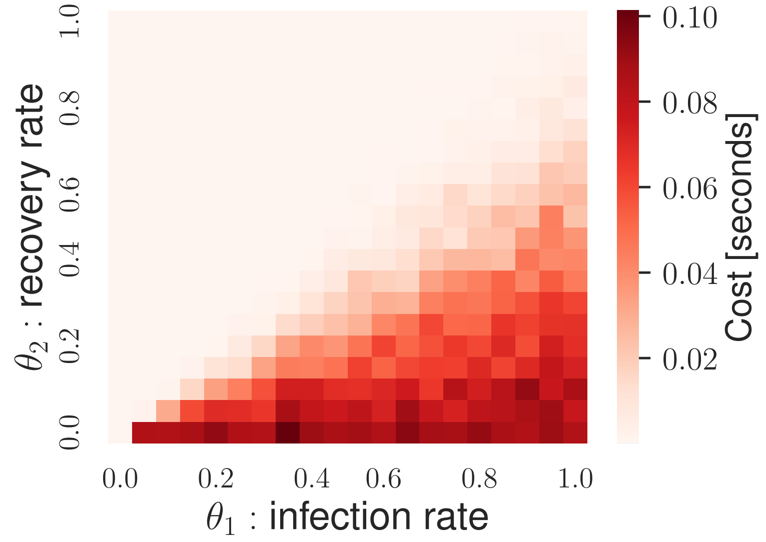
	$\widehat{\text{MMD}}^2$ (↓)					Time saved (↑)			
	NPE	Ca-NPE $g(z) = z^{0.5}$	Ca-NPE $g(z) = z$	Ca-NPE $g(z) = z^2$	Ca-NPE multiple	Ca-NPE $g(z) = z^{0.5}$	Ca-NPE $g(z) = z$	Ca-NPE $g(z) = z^2$	Ca-NPE multiple
Homogen.	0.02(0.02)	0.02(0.01)	$\boxed{0.02(0.02)}$	0.23(0.08)	0.05(0.04)	16%(2)	$\boxed{38\%(2)}$	70%(2)	30%(5)
Temporal	0.03(0.03)	0.06(0.03)	0.07(0.03)	0.07(0.03)	0.05(0.04)	36%(4)	65%(2)	85%(1)	24%(5)
Bernoulli	0.02(0.00)	0.02(0.00)	0.02(0.01)	0.04(0.01)	0.02(0.00)	23%(4)	37%(4)	47%(3)	25%(6)

Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. Mathematical Biosciences, 287:42–53.

Some epidemiological models

- We consider three different models with 1,2 and 3 parameters respectively, and use NPE.

$g(z) = z^2$: Worse accuracy but much cheaper

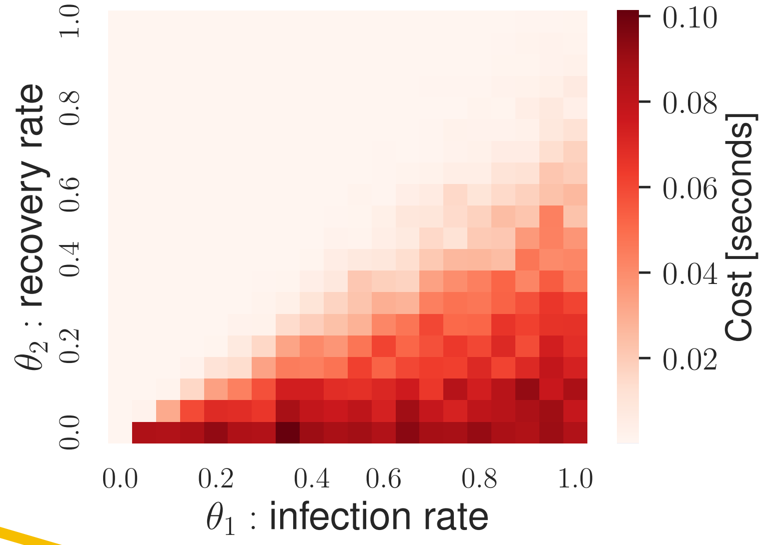


	$\widehat{\text{MMD}}^2$ (↓)					Time saved (↑)			
	NPE	Ca-NPE $g(z) = z^{0.5}$	Ca-NPE $g(z) = z$	Ca-NPE $g(z) = z^2$	Ca-NPE multiple	Ca-NPE $g(z) = z^{0.5}$	Ca-NPE $g(z) = z$	Ca-NPE $g(z) = z^2$	Ca-NPE multiple
Homogen.	0.02(0.02)	0.02(0.01)	0.02(0.02)	$\boxed{0.23(0.08)}$	0.05(0.04)	16%(2)	38%(2)	$\boxed{70\%(2)}$	30%(5)
Temporal	0.03(0.03)	0.06(0.03)	0.07(0.03)	0.07(0.03)	0.05(0.04)	36%(4)	65%(2)	85%(1)	24%(5)
Bernoulli	0.02(0.00)	0.02(0.00)	0.02(0.01)	0.04(0.01)	0.02(0.00)	23%(4)	37%(4)	47%(3)	25%(6)

Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. Mathematical Biosciences, 287:42–53.

Some epidemiological models

- We consider three different models with 1,2 and 3 parameters respectively, and use NPE.



Typically slight loss of accuracy but decent reduction in cost!

	$\widehat{\text{MMD}}^2$ (↓)					Time saved (↑)			
	NPE	Ca-NPE $g(z) = z^{0.5}$	Ca-NPE $g(z) = z$	Ca-NPE $g(z) = z^2$	Ca-NPE multiple	Ca-NPE $g(z) = z^{0.5}$	Ca-NPE $g(z) = z$	Ca-NPE $g(z) = z^2$	Ca-NPE multiple
Homogen.	0.02(0.02)	0.02(0.01)	0.02(0.02)	0.23(0.08)	0.05(0.04)	16%(2)	38%(2)	70%(2)	30%(5)
Temporal	0.03(0.03)	0.06(0.03)	0.07(0.03)	0.07(0.03)	0.05(0.04)	36%(4)	65%(2)	85%(1)	24%(5)
Bernoulli	0.02(0.00)	0.02(0.00)	0.02(0.01)	0.04(0.01)	0.02(0.00)	23%(4)	37%(4)	47%(3)	25%(6)

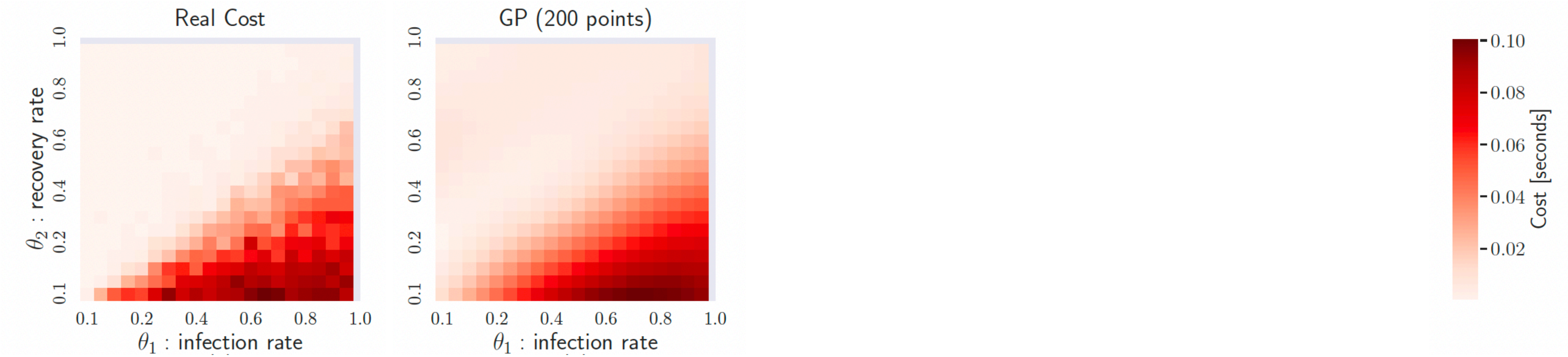
Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. Mathematical Biosciences, 287:42–53.

Estimating the cost function

When the cost function is unknown, it can be estimated through simulations+regression.
This is typically very cheap, and simulations can be re-used for inference!

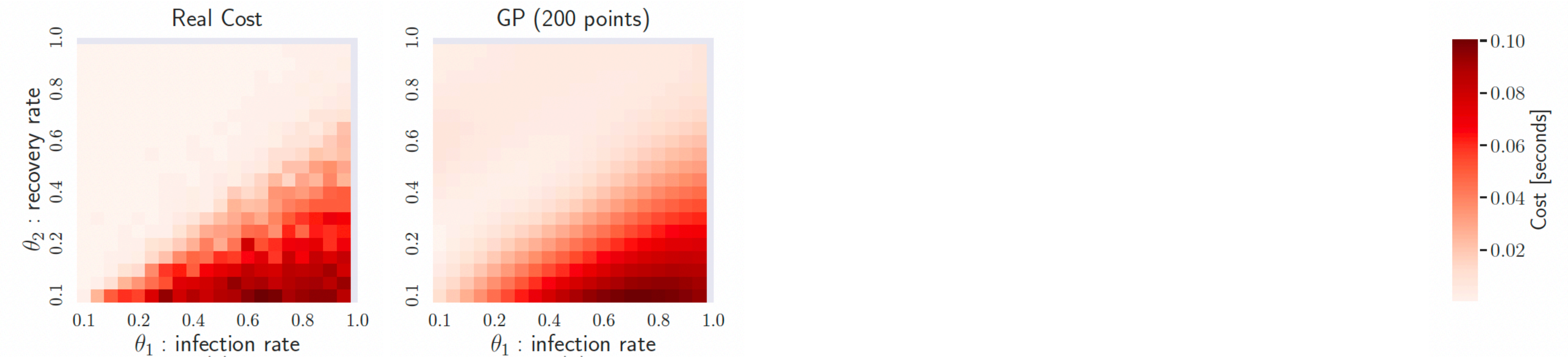
Estimating the cost function

When the cost function is unknown, it can be estimated through simulations+regression.
This is typically very cheap, and simulations can be re-used for inference!



Estimating the cost function

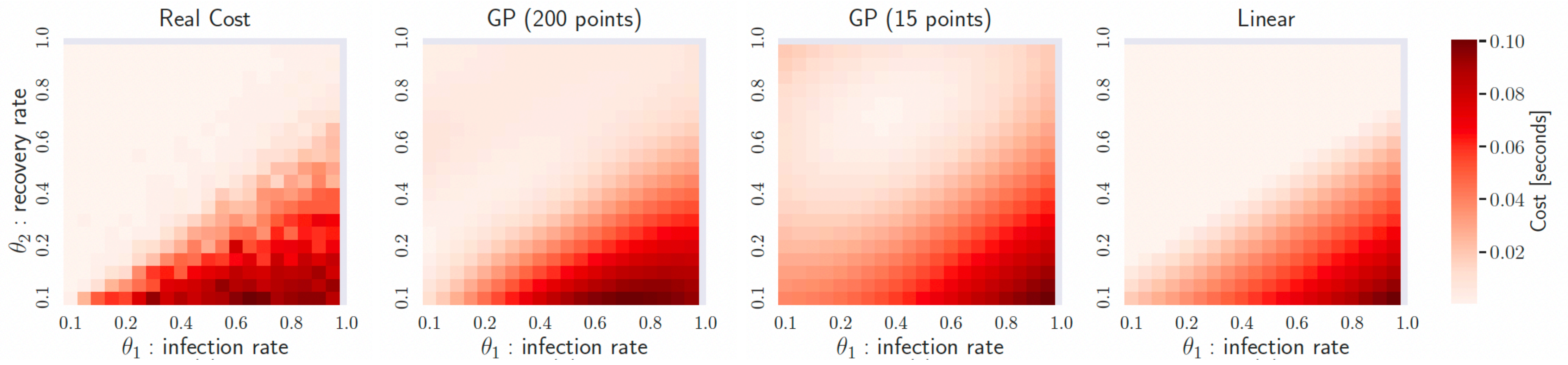
When the cost function is unknown, it can be estimated through simulations+regression.
This is typically very cheap, and simulations can be re-used for inference!



Very accurate!

Estimating the cost function

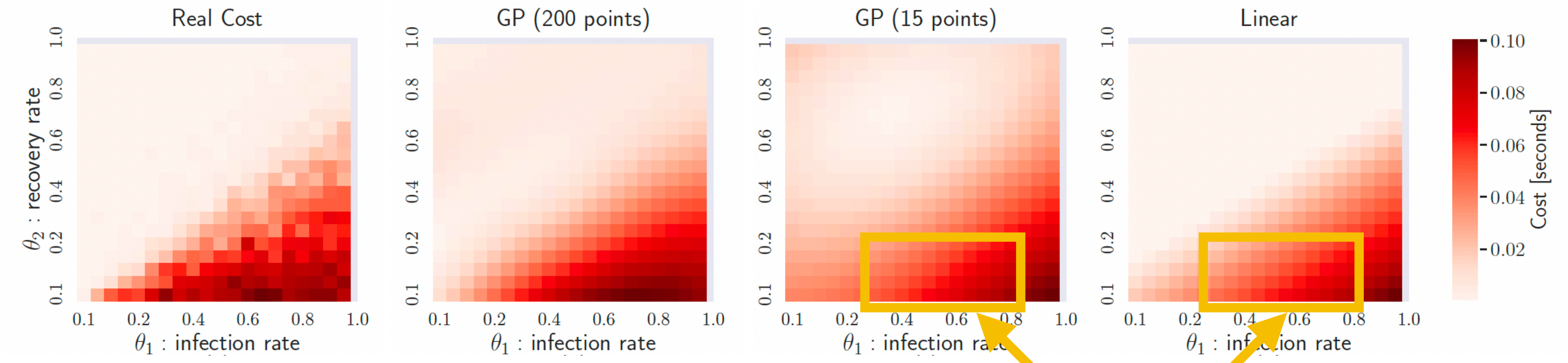
When the cost function is unknown, it can be estimated through simulations+regression.
This is typically very cheap, and simulations can be re-used for inference!



Very accurate!

Estimating the cost function

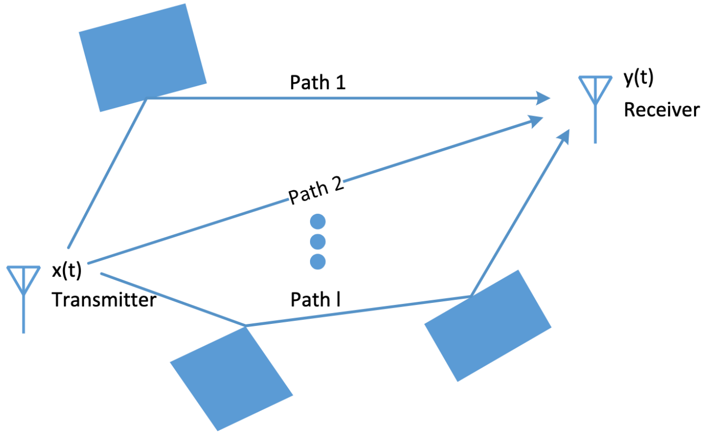
When the cost function is unknown, it can be estimated through simulations+regression.
This is typically very cheap, and simulations can be re-used for inference!



Very accurate!

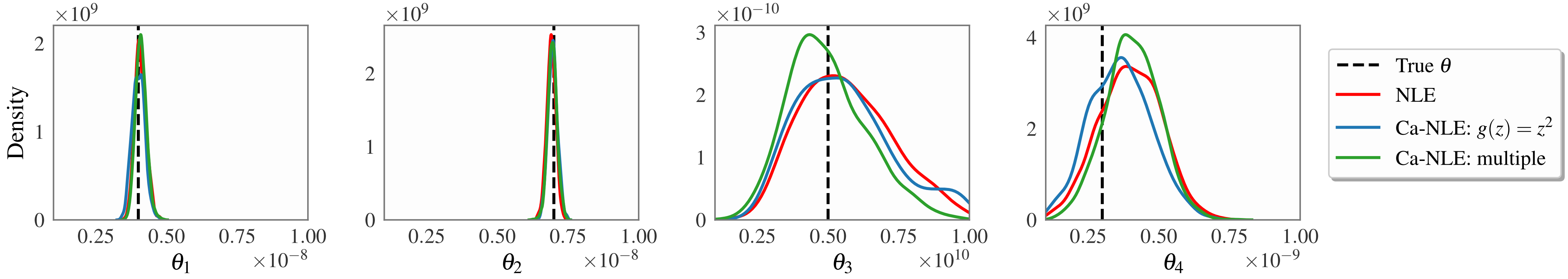
Clearly not perfect, but still pretty good…

Back to radio-propagation



<u>Computational Cost</u>

- **Standard NLE:** 15.6h,
- **Cost-aware NLE:** 8.8h!!



Bharti, A., **Briol, F-X.**, Pedersen, T. (2022). A general method for calibrating stochastic radio channel models with kernels. IEEE Transactions on Antennas and Propagation, vol. 70, no. 6, pp. 3986-4001, June 2022.

Conclusion

Conclusion

- We proposed a novel importance sampling algorithm which focuses on **down weighting sampling** in regions with a **large downstream cost.**

Conclusion

- We proposed a novel importance sampling algorithm which focuses on **down weighting sampling** in regions with a **large downstream cost.**
- Although I presented this for NLE/NPE, we also have experiments for ABC and it could be applied to any other sampling-based SBI method.

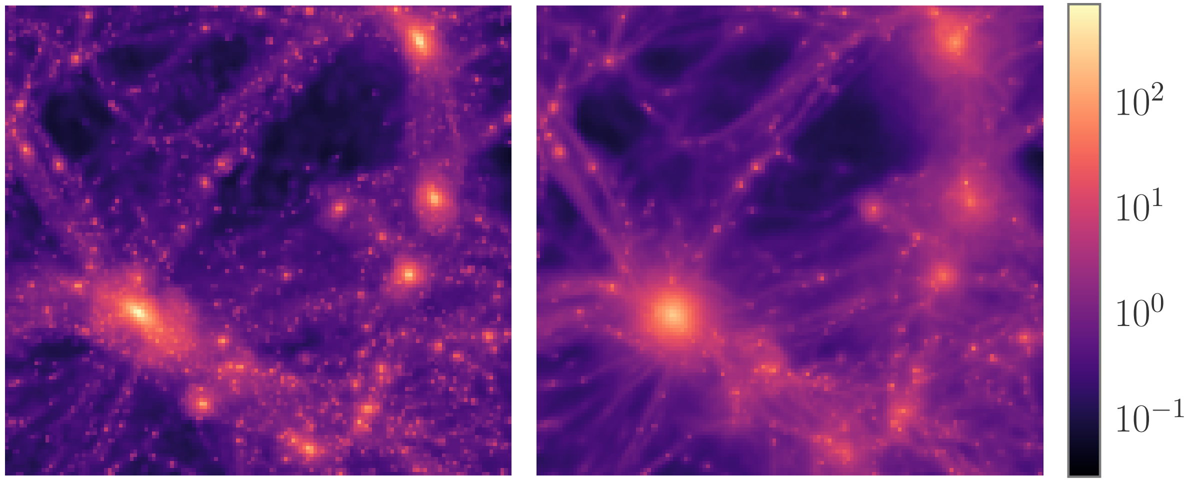
Conclusion

- We proposed a novel importance sampling algorithm which focuses on **down weighting sampling** in regions with a **large downstream cost.**

- Although I presented this for NLE/NPE, we also have experiments for ABC and it could be applied to any other sampling-based SBI method.

- Need more computational statisticians engaging with neural-based simulation inference!

From importance sampling to MLMC….

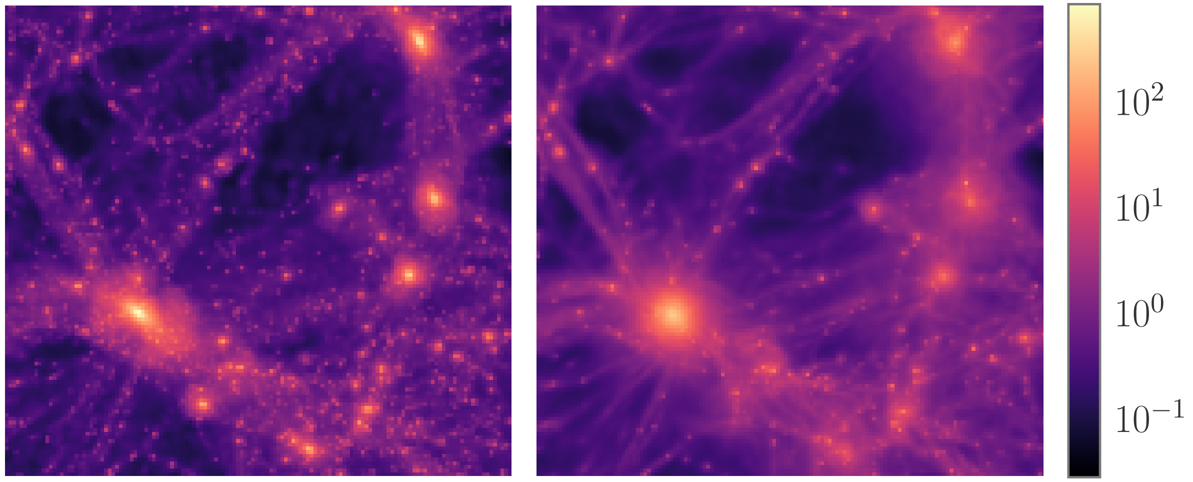


Low-fidelity High-fidelity

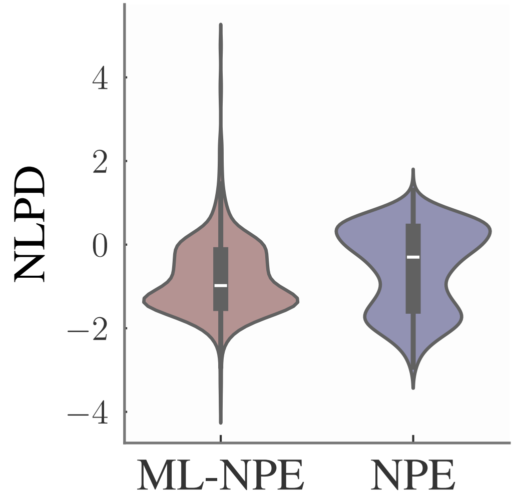
Complement very few expensive but (physically) accurate simulations to combine cheap but inaccurate simulations (with missing physics).

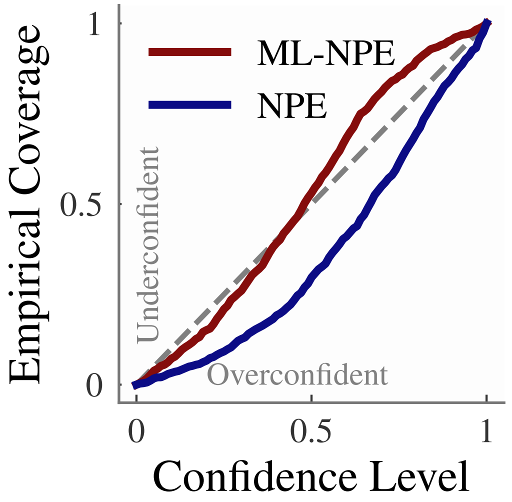
Hikida, Y., Bharti, A., Jeffrey, N. & **Briol, F-X** (2025). Multilevel neural simulation-based inference. Under review.

From importance sampling to MLMC….



Low-fidelity High-fidelity

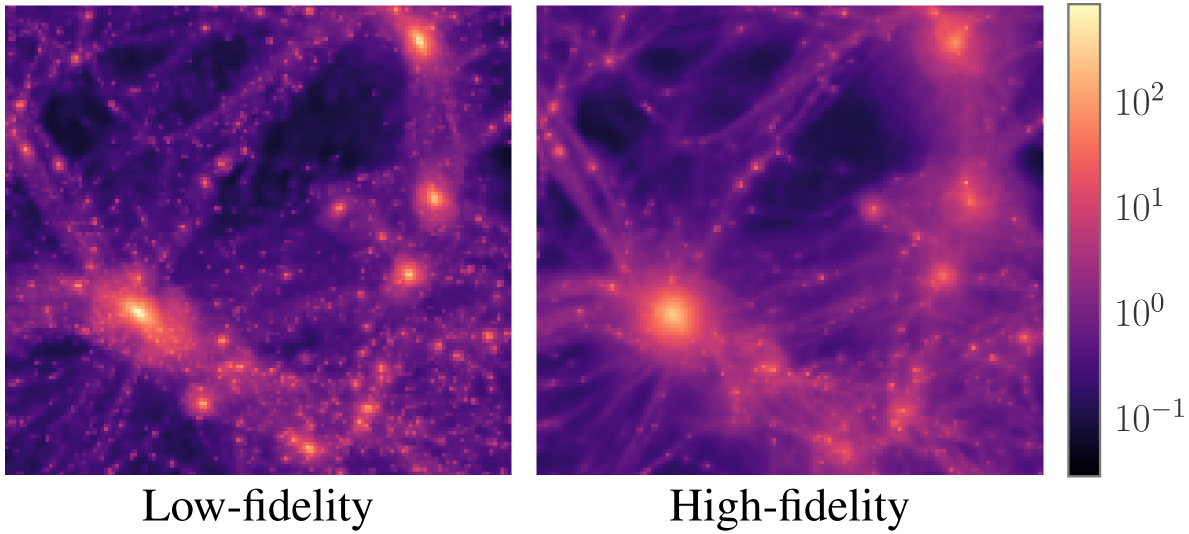


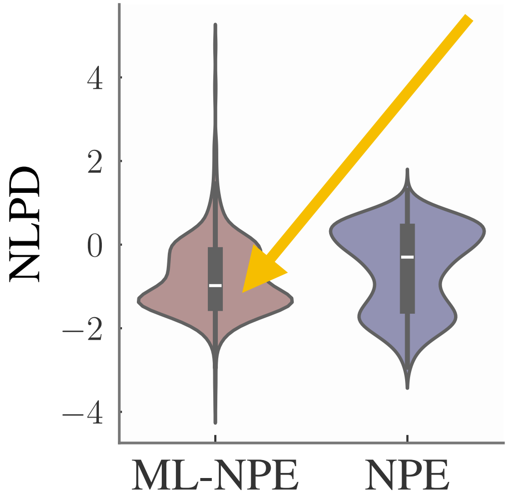


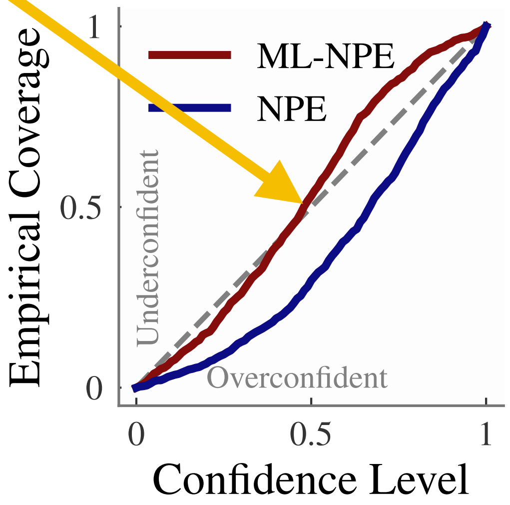
Complement very few expensive but (physically) accurate simulations to combine cheap but inaccurate simulations (with missing physics).

Hikida, Y., Bharti, A., Jeffrey, N. & **Briol, F-X** (2025). Multilevel neural simulation-based inference. Under review.

From importance sampling to MLMC….







Complement very few expensive but (physically) accurate simulations to combine cheap but inaccurate simulations (with missing physics).

Hikida, Y., Bharti, A., Jeffrey, N. & **Briol, F-X** (2025). Multilevel neural simulation-based inference. Under review.

Any Questions?

Cost-aware simulation-based inference

Ayush Bharti, Daolang Huang, Samuel Kaski, Francois-Xavier Briol *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, PMLR 258:28-36, 2025.

Code: https://github.com/huangdaolang/cost-aware-sbi