# Simulation-based inference for implicitly defined models

11th World Congress in Probability and Statistics, Aug 14 2024

Joonha Park, University of Kansas, Department of Mathematics

# Outline

- Introduction: implicitly defined model
- Concept: Simulation-based inference using log-likelihood estimator
- Simulation metamodel
- Method for simulation-based parameter inference
- Numerical demonstration
- Summary & Discussion

# Introduction: Implicitly defined models

# Introduction: Implicitly defined model

- Stochastic system parametrized by $\theta \in \mathbb{R}^d$
- Realization $X \sim P_\theta$.
- A model is called *implicitly defined*[1], if $P_\theta$ can be simulated, but the density cannot be evaluated.
- Computer simulation of complicated stochastic systems
  - Increasing use due to widespread adoption of digital twins.
  - Example: Epidemiological stochastic dynamic model

$$d\mathbf{X}(t) = a\{\mathbf{X}(t)\} \cdot dt + B\{\mathbf{X}(t)\} \cdot dW(t)$$

- Physical simulation of a process

1. P. J. Diggle and R. J. Gratton. Monte Carlo methods of inference for implicit statistical models. Journal of the Royal Statistical Society. Series B (Methodological), pages 193–227, 1984.

# Inference for implicitly defined models & challenges

- Noisy or partial observations: $Y_i \,|\, X \overset{ind}{\sim} g_i(y_i \,|\, X; \theta), \quad i \in 1:n$.

- Likelihood: $L(\theta; y_{1:n}) = \int \boxed{g(y_{1:n} \,|\, x; \theta)}\, dP_\theta(x)$

$$\boxed{\prod_{i=1}^{n} g_i(y_i \,|\, x; \theta)}$$

- Unbiased estimator for $L(\theta; y_{1:n})$:

  Simulate $X^j(\theta) \sim P_\theta, \quad j \in 1:J$ and let $\hat{L}(\theta; y_{1:n}) = \dfrac{1}{J} \displaystyle\sum_{j=1}^{J} g(y_{1:n} \,|\, X^j(\theta); \theta)$.

# Inference for implicitly defined models & challenges

- Partial or noisy observations: $Y_i \,|\, X \overset{ind}{\sim} g_i(y_i \,|\, X; \theta), \quad i \in 1:n.$

- Likelihood: $L(\theta; y_{1:n}) = \int \boxed{g(y_{1:n} \,|\, x; \theta)} \, dP_\theta(x)$

  $$\boxed{\prod_{i=1}^{n} g_i(y_i \,|\, x; \theta)}$$

- Unbiased estimator for $L(\theta; y_{1:n})$:

  Simulate $X^j(\theta) \sim P_\theta, \quad j \in 1:J$ and let $\hat{L}(\theta; y_{1:n}) = \frac{1}{J} \sum_{j=1}^{J} g(y_{1:n} \,|\, X^j(\theta); \theta).$

- Challenge: Variance of $\hat{L}(\theta; y_{1:n})$ grows exponentially with $n$.

# Inference for implicitly defined models & challenges

- Partial or noisy observations: $Y_i \mid X \overset{ind}{\sim} g_i(y_i \mid X; \theta), \quad i \in 1:n.$
- Likelihood: $L(\theta; y_{1:n}) = \int \boxed{g(y_{1:n} \mid x; \theta)}\, dP_\theta(x)$

$$\boxed{\prod_{i=1}^{n} g_i(y_i \mid x; \theta)}$$

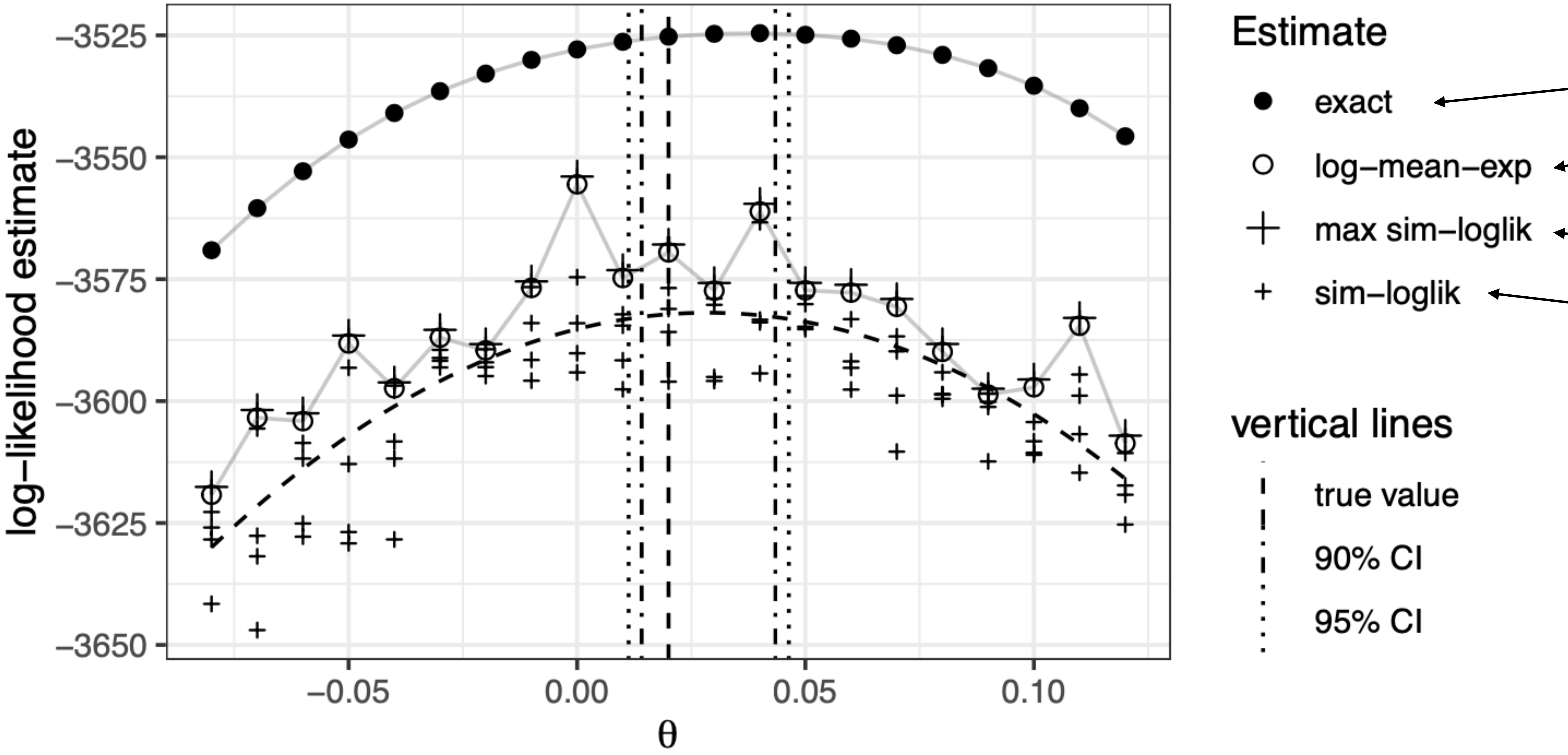
- Unbiased estimator for $L(\theta; y_{1:n})$:

  Simulate $X^j(\theta) \sim P_\theta, \quad j \in 1:J$ and let $\hat{L}(\theta; y_{1:n}) = \dfrac{1}{J} \displaystyle\sum_{j=1}^{J} g(y_{1:n} \mid X^j(\theta); \theta).$

- Challenge: Variance of $\hat{L}(\theta; y_{1:n})$ grows exponentially with $n$.
- New approach: Use the *log*-likelihood estimator, $\hat{\ell}(\theta; y_{1:n}) = \log \hat{L}(\theta; y_{1:n})$.

# Simulation-based inference using the *log*-likelihood estimator

- $J = 5$ simulations per $\theta$, $X^j(\theta) \sim P_\theta$, $j \in 1:J$,
- $\ell_j^S(\theta) = \log g(y_{1:n} \,|\, X^j(\theta))$ (Simulation log-likelihood)




$\ell(\theta)$

$\log\left[\frac{1}{J}\sum_{j=1}^{J} \exp\{\ell_j^S(\theta)\}\right]$

$\max_{j\in 1:5} \ell_j^S(\theta)$

$\ell_j^S(\theta)$

Jensen bias: $\ell(\theta) = \log \mathbb{E} g(y_{1:n} \,|\, X(\theta)) \geq \mathbb{E} \log g(y_{1:n} \,|\, X(\theta)) = \mu(\theta)$.

# Simulation log-likelihood for hidden Markov models

- Consider an implicitly defined, partially observed latent Markov process $X_{1:n}$ (hidden Markov model)
  - The sequence of filtering distributions $\{\mathscr{L}(X_t \,|\, y_{1:t}); t \in 1:n\}$, can be approximated by a recursive Monte Carlo algorithm, called the *bootstrap particle filter.*
  - An unbiased likelihood estimator $\hat{L}(\theta)$ can be obtained by running the bootstrap particle filter.
- $\ell^S(\theta) := \log \hat{L}(\theta)$ can be used as a *simulation log-likelihood*.

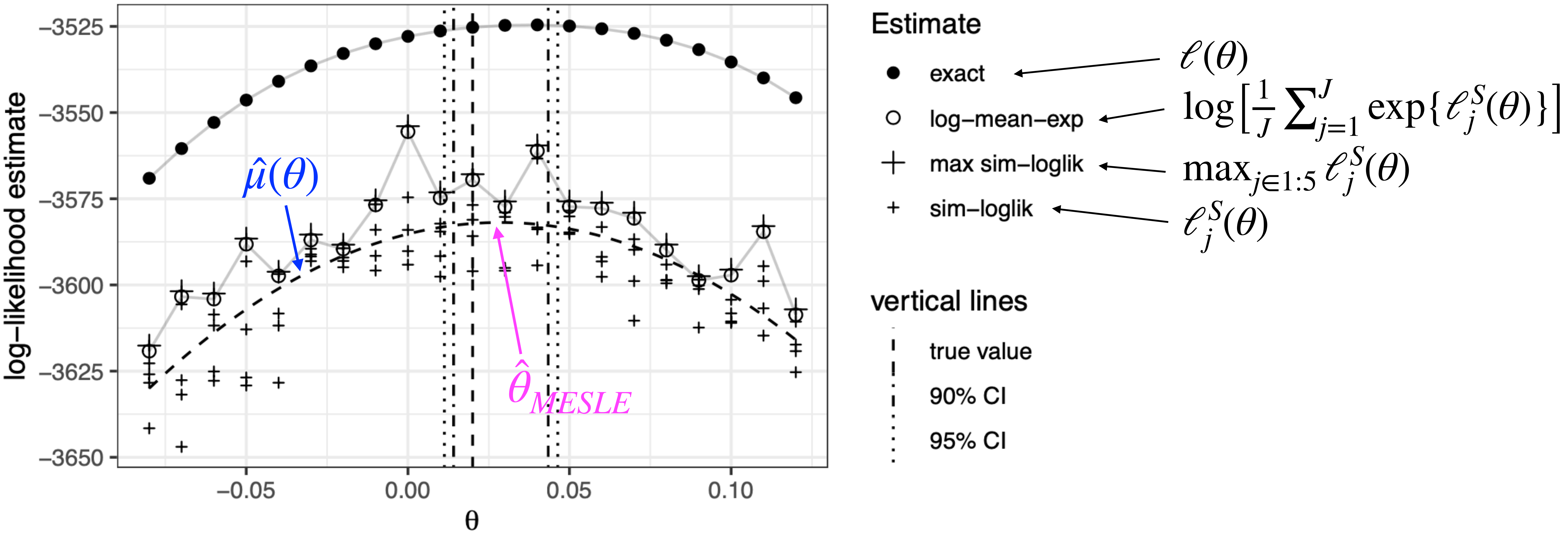
# Simulation metamodel

# Mean of the simulation log-likelihood

- Expected simulation log-likelihood: $\mu(\theta; y_{1:n}) = \mathbb{E}\,\ell^S(\theta; y_{1:n})$.

- Maximum Expected Simulation Log-likelihood Estimator (MESLE):

$$\theta_{MESLE}(y_{1:n}) = \arg\max_{\theta} \mu(\theta; y_{1:n})$$



# Simulation metamodel conditional on data

- Simulation metamodel conditional on data $y_{1:n}$:

$$\ell^S(\theta; y_{1:n}) \sim N\left(a(y_{1:n}) + b(y_{1:n})^\top \theta + \theta^\top c(y_{1:n})\theta,\ \frac{\sigma^2(y_{1:n})}{w(\theta)}\right).$$

$\Rightarrow$ Summarizes the distribution of $\ell^S(\theta)$ due to randomness in simulations.

- Asymptotically valid when $X = (X_1, \ldots, X_n),\ X_i$ independent, and $Y_i$ depends only on $X_i$.

- Also valid in some dependent cases where mixing occurs.

# Simulation metamodel

- Parameter inference should consider randomness in observations as well.
- Data-averaged expected simulation log-likelihood: $U(\theta_0; \theta) = \mathbb{E}_{Y_{1:n} \sim P^Y_{\theta_0}} \mu(\theta; Y_{1:n})$
- Simulation-based parameter surrogate : $\theta_* \overset{def.}{=} \arg\max_{\theta} U(\theta_0; \theta)$

# Simulation metamodel

- Parameter inference should consider randomness in observations as well.
- Data-averaged expected simulation log-likelihood: $U(\theta_0; \theta) = \mathbb{E}_{Y_{1:n} \sim P^Y_{\theta_0}} \mu(\theta; Y_{1:n})$
- Simulation-based parameter surrogate : $\theta_* \overset{def.}{=} \arg\max_\theta U(\theta_0; \theta)$
- $\theta_* = \theta_0$ in some models, but in general they are different.
- Bias $|\theta_* - \theta_0|$ may be bounded when the Jensen bias $|\ell(\theta) - \mu(\theta)|$ is approximately constant.

# Local asymptotic normality (LAN) for simulation log-likelihood

Simulation metamodel conditional on data $y_{1:n}$:

$$\ell^S(\theta; y_{1:n}) \sim N\left(a(y_{1:n}) + b(y_{1:n})^\top \theta + \theta^\top c(y_{1:n})\theta,\ \frac{\sigma^2(y_{1:n})}{w(\theta)}\right).$$

# Local asymptotic normality (LAN) for simulation log-likelihood

Simulation metamodel conditional on data $y_{1:n}$:

$$\ell^S(\theta; y_{1:n}) \sim N\left(a(y_{1:n}) + b(y_{1:n})^\top \theta + \theta^\top c(y_{1:n})\theta, \frac{\sigma^2(y_{1:n})}{w(\theta)}\right).$$

$$\mu(\theta; Y_{1:n})$$

$$= \mu(\theta_*; Y_{1:n}) + \frac{1}{\sqrt{n}} \frac{\partial \mu}{\partial \theta}(\theta_*; Y_{1:n}) \cdot \sqrt{n}(\theta - \theta_*) + \frac{1}{2n} \frac{\partial^2 \mu}{\partial \theta^2}(\theta_*; Y_{1:n}) \cdot n(\theta - \theta_*)^2 + o(n\|\theta - \theta_*\|^2)$$

$$= \mu(\theta_*; Y_{1:n}) + S_n \cdot \sqrt{n}(\theta - \theta_*) - \frac{1}{2} K_2 \{\sqrt{n}(\theta - \theta_*)^2\} + o(n\|\theta - \theta_*\|^2)$$

$$S_n = \frac{1}{\sqrt{n}} \frac{\partial \mu}{\partial \theta}(\theta_*; Y_{1:n}) \underset{n\to\infty}{\Rightarrow} N(0, K_1), \qquad K_2 = -\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \mu}{\partial \theta^2}(\theta_*; Y_{1:n}).$$

# Local asymptotic normality (LAN) for simulation log-likelihood

$$
\begin{aligned}
&\mu(\theta; Y_{1:n}) \\
&= \mu(\theta_*; Y_{1:n}) + \frac{1}{\sqrt{n}} \frac{\partial \mu}{\partial \theta}(\theta_*; Y_{1:n}) \cdot \sqrt{n}(\theta - \theta_*) + \frac{1}{2n} \frac{\partial^2 \mu}{\partial \theta^2}(\theta_*; Y_{1:n}) \cdot n(\theta - \theta_*)^2 + o(n\|\theta - \theta_*\|^2) \\
&= \mu(\theta_*; Y_{1:n}) + S_n \cdot \sqrt{n}(\theta - \theta_*) - \frac{1}{2} K_2 \{\sqrt{n}(\theta - \theta_*)^2\} + o(n\|\theta - \theta_*\|^2)
\end{aligned}
$$

$$
S_n = \frac{1}{\sqrt{n}} \frac{\partial \mu}{\partial \theta}(\theta_*; Y_{1:n}) \underset{n\to\infty}{\Rightarrow} N(0, K_1), \qquad K_2 = -\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \mu}{\partial \theta^2}(\theta_*; Y_{1:n}).
$$

Local asymptotic normality for log-likelihood $\ell(\theta; Y_{1:n})$ (Le Cam, 1986):

$$
\ell(\theta; Y_{1:n}) = \ell(\theta_0; Y_{1:n}) + S'_n \cdot \sqrt{n}(\theta - \theta_0) - \frac{1}{2} I(\theta_0) \cdot \{\sqrt{n}(\theta - \theta_0)\}^2 + o_p(n(\theta - \theta_0)^2).
$$

$S'_n \sim N(0, I(\theta_0))$

$I(\theta_0)$: Fisher information at $\theta_0$.

# Simulation metamodel

- Simulation metamodel conditional on data:

$$\ell^S(\theta; y_{1:n}) \sim N\left( a(y_{1:n}) + b(y_{1:n})^\top \theta + \theta^\top c(y_{1:n})\theta,\ \frac{\sigma^2(y_{1:n})}{w(\theta)} \right).$$

$\rightarrow$ models randomness in simulations

- Marginal simulation metamodel:

$$a(Y_{1:n}) + b(Y_{1:n})^\top \theta + \theta^\top c(Y_{1:n})\theta = \mu(\theta_*; Y_{1:n}) + S_n \cdot \sqrt{n}(\theta - \theta_*) - \frac{1}{2} K_2 \{\sqrt{n}(\theta - \theta_*)^2\}$$

$$b = \sqrt{n} S_n + n K_2 \theta_* \sim N(n K_2 \theta_*,\ n K_1), \qquad c = -\frac{n}{2} K_2$$

$\rightarrow$ models randomness in observations

# Simulation-based parameter inference

# Point estimation of MESLE

- Simulate $X$ at $\theta_1, \ldots, \theta_M$ and obtain $\ell^S(\theta_m)$, $m \in 1:M$.

- Point estimate of metamodel parameters:

$$(\hat{a}, \hat{b}, \hat{c}) = \{(1, \theta_{1:M}, \theta^2_{1:M})^\top W (1, \theta_{1:M}, \theta^2_{1:M})\}^{-1} (1, \theta_{1:M}, \theta^2_{1:M})^\top W \ell^S_{1:M}$$

where $\theta_{1:M} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_M \end{pmatrix}$, $\theta^2_{1:M} = \begin{pmatrix} \theta_1^2 \\ \vdots \\ \theta_M^2 \end{pmatrix}$, $\ell^S_{1:M} = \begin{pmatrix} \ell^S(\theta_1) \\ \vdots \\ \ell^S(\theta_M) \end{pmatrix}$, and $W = \text{diag}(w(\theta_1), \ldots, w(\theta_M))$.

- Point estimator of the MESLE: $\hat{\theta}_{MESLE} = \arg\max_\theta \hat{a} + \hat{b}^\top\theta + \theta^\top \hat{c}\theta = -\frac{1}{2}\hat{c}^{-1}\hat{b}$.

# Hypothesis test about $\theta_{MESLE}$

- Consider a hypothesis test

$$H_0 : \theta_{MESLE} = \theta_0, \qquad H_1 : \theta_{MESLE} \neq \theta_0.$$

- Use the simulation metamodel conditional on data:

$$\ell^S(\theta; y_{1:n}) \sim N\left( a(y_{1:n}) + b(y_{1:n})^\top \theta + \theta^\top c(y_{1:n})\theta, \frac{\sigma^2(y_{1:n})}{w(\theta)} \right).$$

- Test statistic: $T_{MESLE} \sim F_{d, M - \frac{d^2+3d+2}{2}}$ under $H_0$.

- Confidence interval for $\theta_{MESLE}$ can be constructed.

# Hypothesis test about $\theta_*$

- Consider a hypothesis test about the simulation surrogate:

$$H_0 : \theta_* = \theta_{*,0}, \qquad\qquad H_1 : \theta_* \neq \theta_{*,0}$$

- Use the conditional metamodel and the marginal metamodel.

$$b = \sqrt{n}S_n + nK_2\theta_* \sim N(nK_2\theta_*,\ nK_1), \qquad\qquad c = -\frac{n}{2}K_2$$

- Test statistic: $T_{surrogate} \sim F_{d,M-\frac{d^2+3d+2}{2}}$ under $H_0$.

- A confidence interval for $\theta_*$ can be constructed.

- We use relative differences $(\ell^S(\theta_2) - \ell^S(\theta_1), \ldots, \ell^S(\theta_M) - \ell^S(\theta_1))$, which does not depend on $a$ (undefined under the simulation metamodel).

# Numerical demonstration

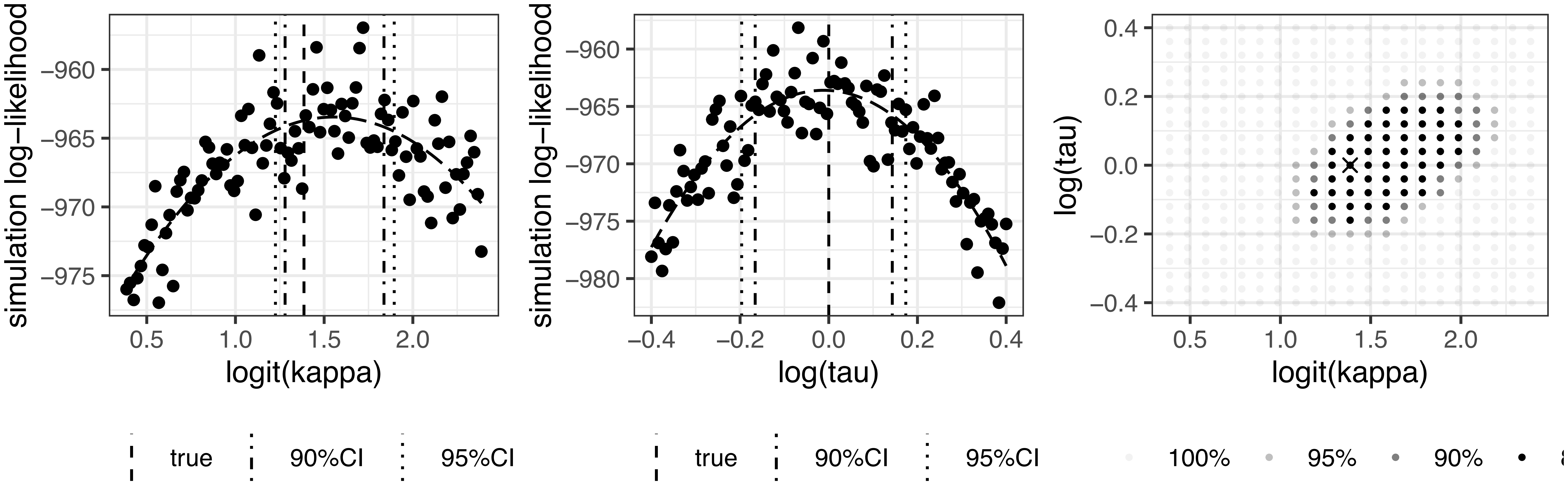
# Numerical results

- Stochastic volatility model

$$r_i = e^{s_i} W_i, \quad W_i \overset{iid}{\sim} t_5,$$

$$s_i = \kappa s_{i-1} + \tau\sqrt{1-\kappa^2}V_i \;\text{ for }\; i > 1, \quad s_1 = \tau V_1, \quad V_i \overset{iid}{\sim} N(0,1)\,.$$

- Simulation log-likelihood $\ell^S(\theta)$ was obtained by running the bootstrap particle filter.



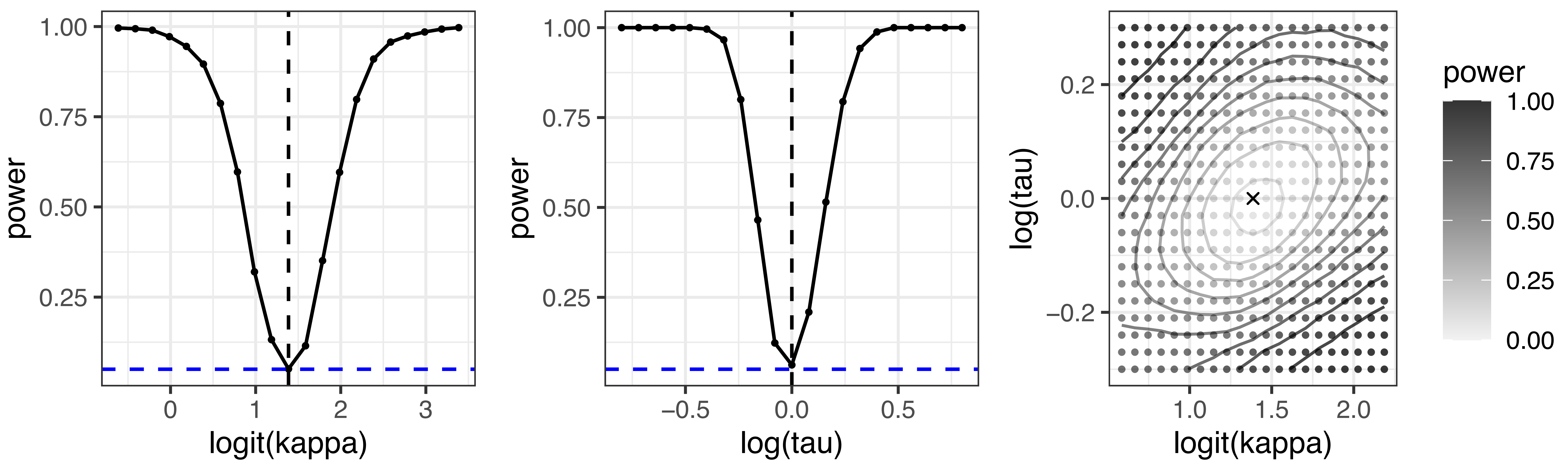
# Numerical results

- Stochastic volatility model

$$r_i = e^{s_i} W_i, \quad W_i \overset{iid}{\sim} t_5,$$

$$s_i = \kappa s_{i-1} + \tau\sqrt{1-\kappa^2} V_i \text{ for } i > 1,, \quad s_1 = \tau V_1, \quad V_i \overset{iid}{\sim} N(0,1)\,.$$

- Power curve at a 5% significance level



# Numerical example: Stochastic compartment model

- Stochastic compartment model for infectious disease transmission:

Susceptible → Exposed → Infectious → Recovered

$$dS(t) = -\left\{\left(\frac{R_0 s(t)(I(t)+\iota)^\alpha}{N(t)} + \mu\right) S(t)dt + dW_{SE}(t) + dW_{SD}(t)\right\} + db(t)$$

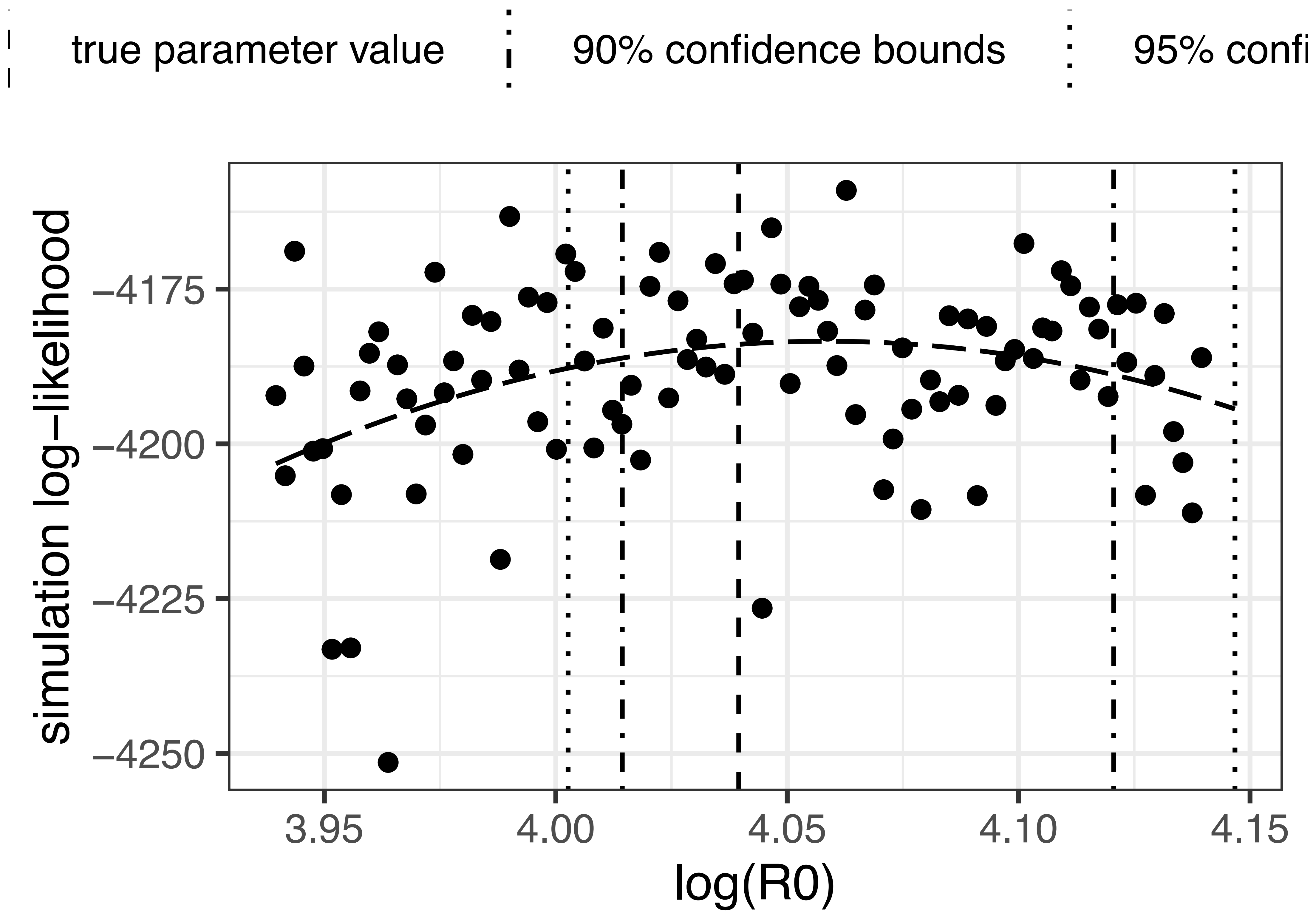
$$dE(t) = \left\{\frac{R_0 s(t)(I+\iota)^\alpha}{N(t)} S(t)dt + dW_{SE}(t)\right\} - \{(\gamma_{EI}+\mu)E(t)dt + dW_{EI}(t) + dW_{ED}(t)\}$$

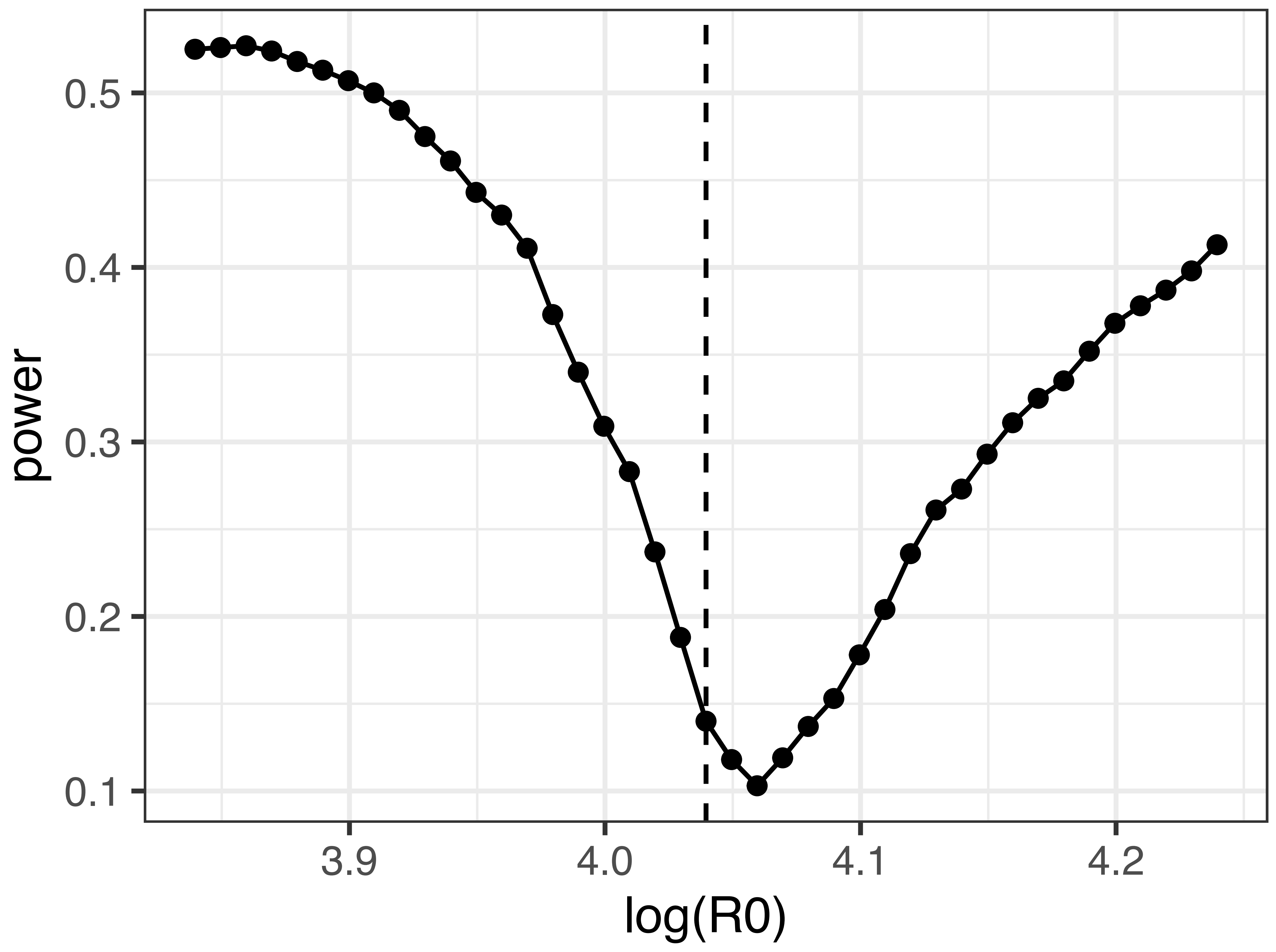
$$dI(t) = \{\gamma_{EI}dt + dW_{EI}(t)\} - \{(\gamma_{IR}+\mu)I(t)dt + dW_{IR}(t) + dW_{ID}(t)\}\,.$$

$R_0$: Basic reproduction number

# Numerical results: Stochastic compartment model

- Inference for $R_0$ (basic reproduction number)



# Comparison with particle Markov chain Monte Carlo

- Bayesian inference:

posterior distribution $\quad \pi(\theta \mid y_{1:n}) \propto \underbrace{h(\theta)}_{\text{prior}} \cdot \underbrace{L(\theta; y_{1:n})}_{\text{likelihood}}.$

- Particle Markov chain Monte Carlo (PMCMC) is a Bayesian parameter inference method using an unbiased estimator of the likelihood (Andrieu, Doucet, & Holenstein, 2010).

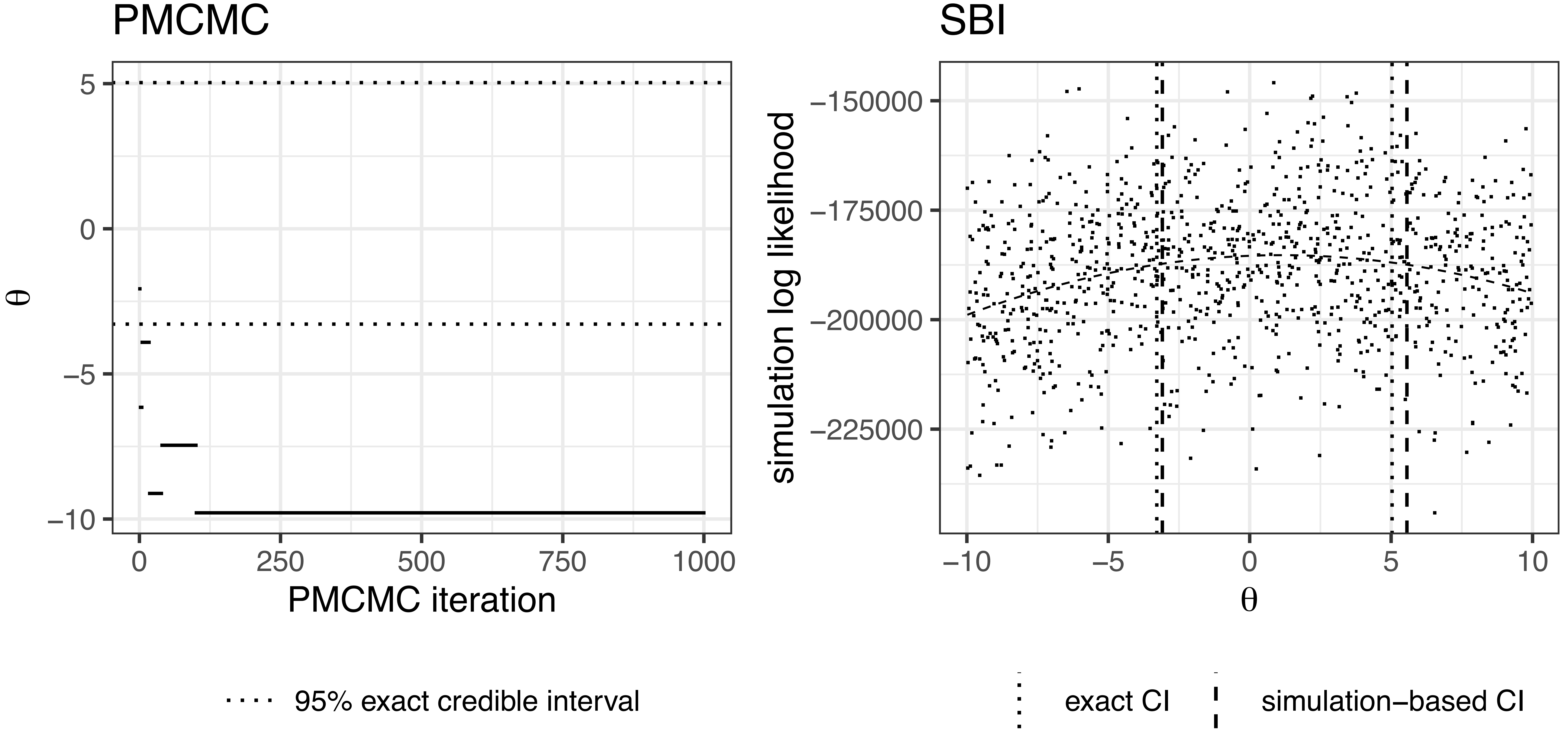
    - $\theta$: current state of a constructed Markov chain
    - $\theta' \sim q(\theta' \mid \theta)$ a proposed candidate
    - $\hat{L}(\theta')$ is obtained via simulation
    - $\theta'$ is accepted with probability $\min\left(1, \dfrac{h(\theta')q(\theta \mid \theta')\hat{L}(\theta')}{h(\theta)q(\theta' \mid \theta)\hat{L}(\theta)}\right).$

# Comparison: PMCMC vs. our metamodel-based method

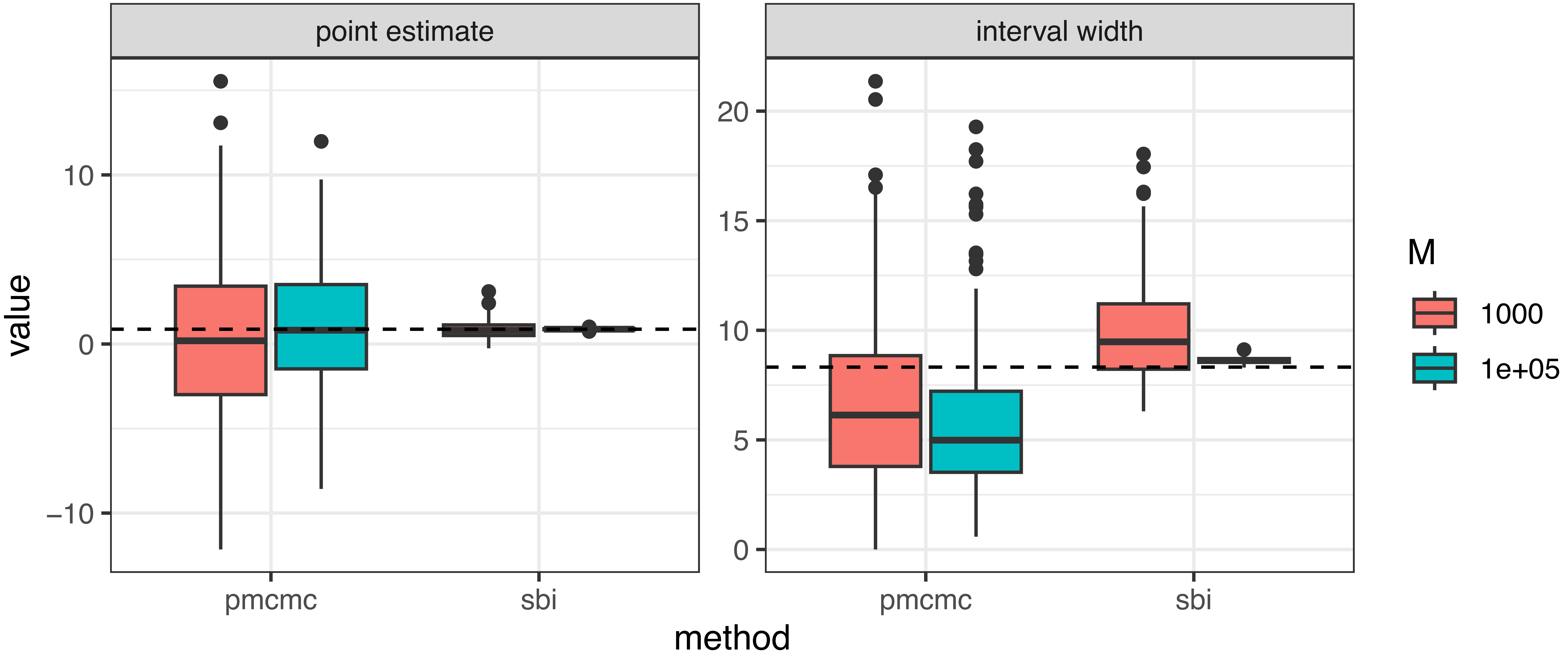
<u>Example</u> $X_{1:n} \overset{iid}{\sim} N(\theta, \tau^2), \quad Y_i \,|\, X_i \overset{ind}{\sim} N(X_i, 1) \qquad (\tau = 30,\ n = 200).$

- Exact 95% confidence interval for $\theta$

  = Exact 95% Bayesian credible interval for $\theta$ $= \bar{y} \pm \sqrt{(\tau^2 + 1)/n} \cdot z_{0.025}$



# Comparison: PMCMC vs. our metamodel-based method




95% Monte Carlo credible interval constructed using PMCMC

= (average of sample draws) $\pm z_{0.025} \cdot$ (standard deviation of sample draws)

# R package: sbi (simulation-based inference)

## How to use package sbi: simulation-based inference

### Introduction

This tutorial introduces you to the package `sbi` and explains how to use it using examples. This package implements parameter inference methods for stochastic models defined implicitly by a simulation algorithm, developed by Park, J. (2023) “On simulation-based inference for implicilty defined models” https://doi.org/10.48550/arXiv.2311.09446. First, the methodological and theoretical framework for inference is explained. Then how to create an `R` object that contains simulation-based log-likelihood estimates will be explained. How to carry out a hypothesis test will be explained first for independent and identically distributed (iid) data using a toy example. Conducting hypothesis tests for a certain class of models generating dependent observations will be explained next using an example of stochastic volatility model.

### Mathematical framework

This section provides a mathematical basis for the methods implemented in the package `sbi`. Further details can be found in the article Park (2023). If you want to learn only about how to use the package, you may skip to the next section.

We consider a collection of latent random variables $X$ distributed according to $P_\theta$, and partial observations $Y$ whose conditional distributions have density $g(y|x;\theta)$. The underlying process $P_\theta$ is not assumed to have a density that can be evaluated analytically pointwise. The parameter $\theta$ may affect both the latent process $X$ and the conditional measurement process $Y$ given $X$; however, $\theta$ may comprise two components each governing the latent process or the measurement process only.

Installable devtools::install_github(“joonhap/sbi")    url: https://github.com/joonhap/sbi

# Summary

- We developed simulation-based inference framework for implicitly defined models.
- Applicable to:
  - complex computational simulation models (digital twin), physical experiments.
- Used a **simulation metamodel** to enable *parameter estimation* and *uncertainty quantification.*
- Our method **scales** favorably with **increasing data size.**
- R package available at https://github.com/joonhap/sbi
- Paper available at **arXiv:2311.09446**.
- Future developments:
  - Bayesian inference employing a more general simulation metamodel
  - Adaptive selection of parameter values for simulation

# Inference bias

# Bias in simulation log-likelihood

- Suppose that $\hat{L}(\theta) = \exp\{\ell^S(\theta)\}$ is unbiased for $L(\theta)$.

- Jensen bias:$\ell(\theta) = \log \mathbb{E} \exp\{\ell^S(\theta)\} \geq \mathbb{E}\ell^S(\theta) = \mu(\theta)$.

- Jensen bias is upper bounded if

  - the Jensen bias for each observation piece is upper bounded, or

  - $\ell^S(\theta)$ has a sub-Gaussian upper tail.

# Bias in parameter inference

- $\mu(\theta; Y_{1:n}) \le \ell(\theta; Y_{1:n}) \le \mu(\theta; Y_{1:n}) + B(\theta; Y_{1:n})$

- $U(\theta_0, \theta) \le -H(\theta) \le U(\theta_0, \theta) + \bar{B}$

- If the quadratic approximations to $-H(\theta)$ and $U(\theta_0, \theta)$ are $\epsilon$-accurate, then

$$\|\theta_* - \theta_0\| \le \frac{2(\bar{B} + 2\epsilon)}{\delta\lambda}$$

where $\lambda$ is the smallest eigenvalue of $-\frac{\partial^2}{\partial\theta^2}U(\theta_0, \theta_*)$.