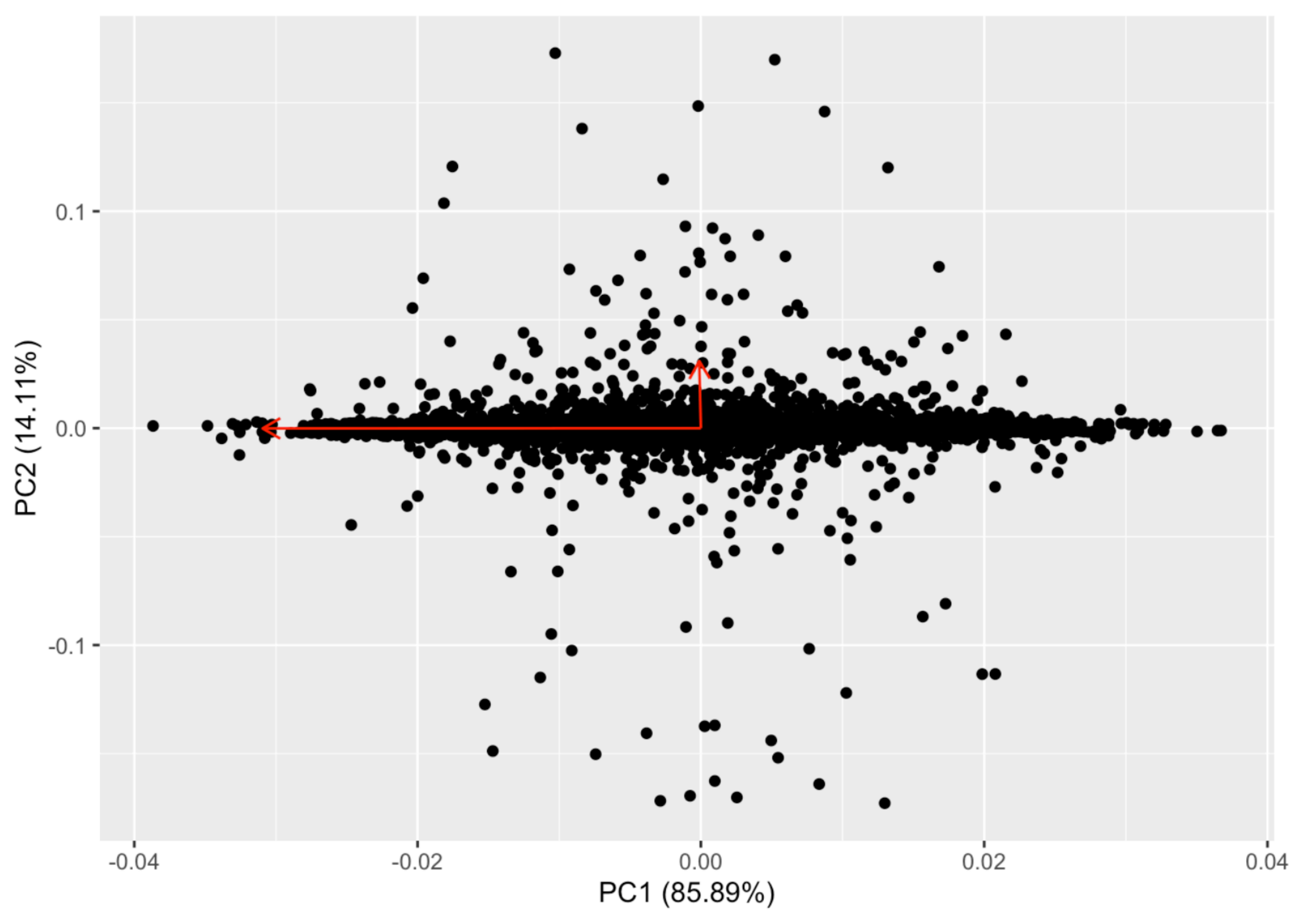
Improved MCMC with active subspaces



Richard Everitt, Warwick

and

Leonardo Ripoli, Reading

WARWICK
THE UNIVERSITY OF WARWICK

UKRI
Natural Environment Research Council

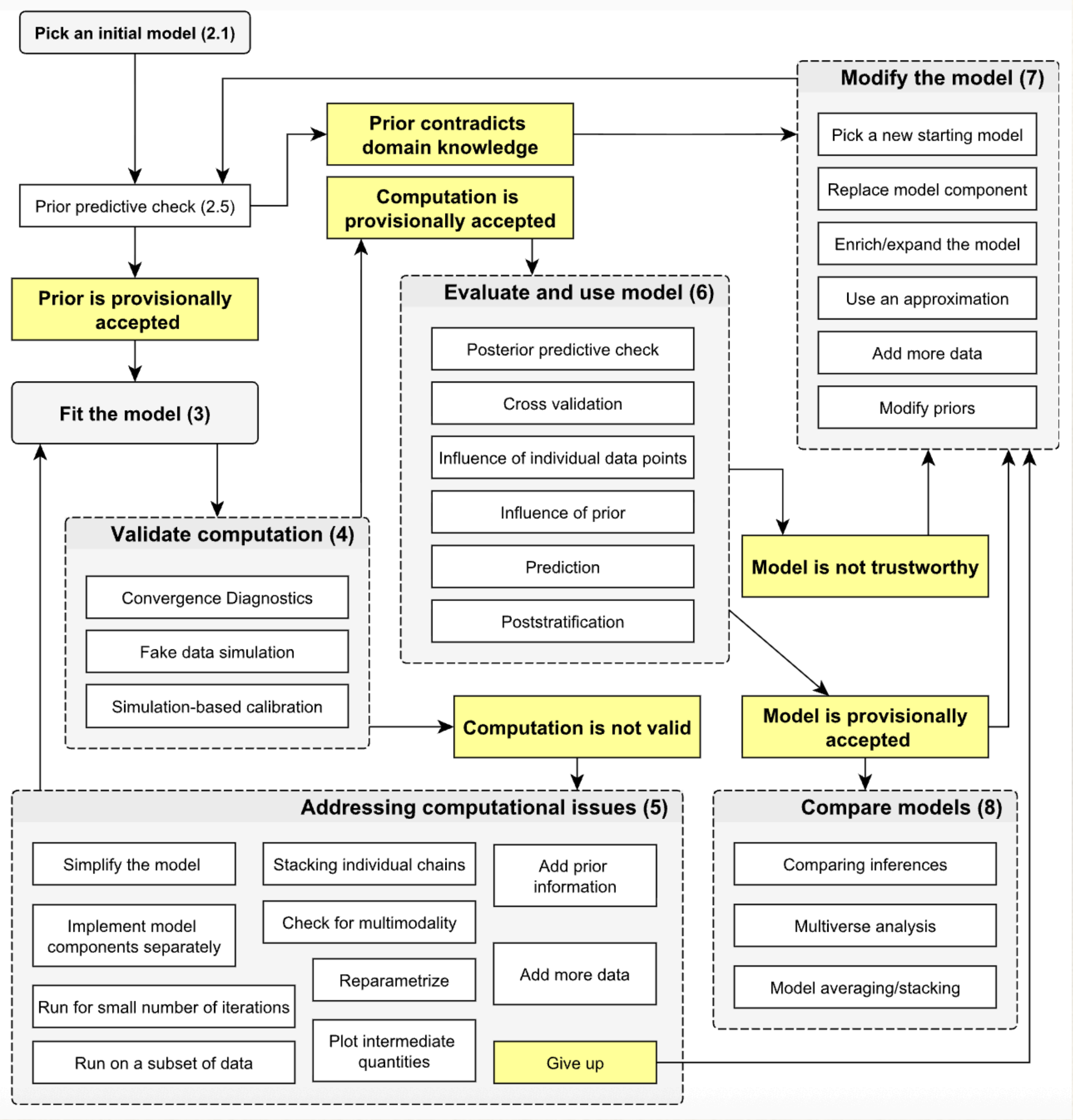
Warwick Statistics

The folk theorem of statistical computing

- **The Folk Theorem of Statistical Computing (Gelman 2008):**

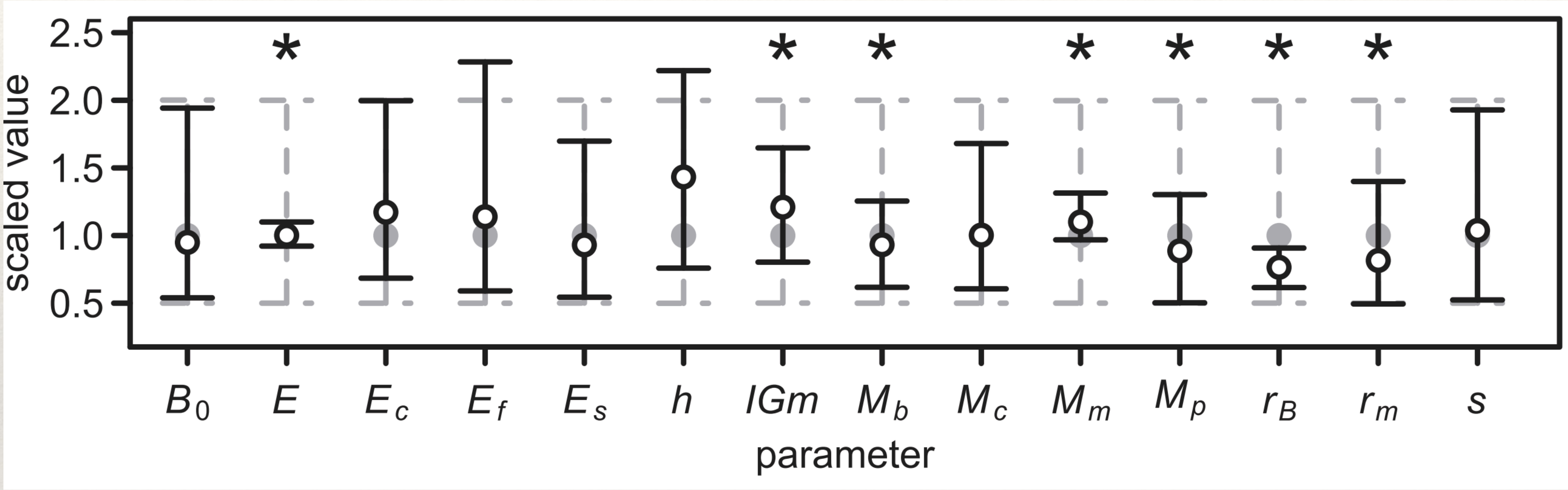
  When you have computational problems, often there's a problem with your model.

Gelman A, Vehtari A, Simpson D, et al. (2020) Bayesian workflow, arXiv.



Warwick Statistics

Unfortunate reality



van der Vaart, Beaumont, Johnston, Sibly (2015), Calibration and evaluation of individual-based models using Approximate Bayesian Computation, Ecological Modelling.

Plan of talk

- Introduce active subspaces.
- Active subspace Metropolis-Hastings (Constantine et al., 2016; Schuster et al., 2017).
- Active subspace PMMH.
- Alternatives and extensions.

Constantine, Kent and Bui-Thanh (2016), Accelerating Markov chain Monte Carlo with active subspaces, SIAM Journal on Statistical Computing.

Schuster, Constantine and Sullivan (2017), Exact active subspace Metropolis-Hastings, with applications to the Lorenz-96 system, arXiv.

Ripoli, L. and Everitt, R. G. (2025) Improved MCMC with active subspaces, arXiv.

Ripoli, L. and Everitt, R. G. (2024) Sequential Monte Carlo with active subspaces, arXiv.

Setup

- We wish to infer parameters $\theta \in \mathbb{R}^d$ from data y.
- Use Bayes:

$$\tilde{\pi}\left(\theta \mid y\right) = p\left(\theta\right) l\left(y \mid \theta\right) \qquad \pi\left(\theta \mid y\right) = \frac{\tilde{\pi}\left(\theta \mid y\right)}{p\left(y\right)}$$

- From hereon we denote
  - posterior as $\pi\left(\theta\right) := \pi\left(\theta \mid y\right)$
  - likelihood as $l\left(\theta\right) := l\left(y \mid \theta\right)$

Warwick Statistics

Active subspaces in Monte Carlo methods

- **Aim:** to find a reparameterisation of θ of the form
$$\theta = B_a a + B_i i,$$
such that the marginal distribution of a captures most of the variability of π, where $B_a \in \mathbb{R}^{d \times d_a}$ and $B_i \in \mathbb{R}^{d \times d_i}$ are such that $[B_a, B_i]$ is an orthonormal basis of $\mathbb{R}^d$.
- Ideally we would run a MCMC algorithm on a rather than θ.

Identifiability and sensitivity

- "A model parameter is said to be *identifiable* if it can be uniquely determined from the model output".
- The *sensitivity Fisher Information matrix* (sFIM) is defined as:
$$F(\theta) = \nabla \log l(\theta) \nabla \log l(\theta)^T$$
- Useful since $\nabla \log l(\theta)$ tells us about how a change in θ affects the likelihood (sensitivity).
- If F has full rank at θ, then we say that the model is (locally) *structurally identifiable* at θ.

Brouwer and Eisenberg (2018), The underlying connections between identifiability, active subspaces, and parameter space dimension reduction, arXiv.

Active subspaces

- The sFIM measures identifiability local to θ.
- *Active subspaces* are identified by finding the expectation of sFIM over some distribution ϕ:

$$\int_\theta \nabla \log l(\theta) \nabla \log l(\theta)^T \phi(\theta)\, d\theta\,.$$

- This measures global (across ϕ) identifiability.
- Take the eigendecomposition of this matrix and partition the the eigenvalues and eigenvectors as:

$$\begin{pmatrix} \Lambda_a & \\ & \Lambda_i \end{pmatrix} \qquad \begin{pmatrix} B_a & B_i \end{pmatrix},$$

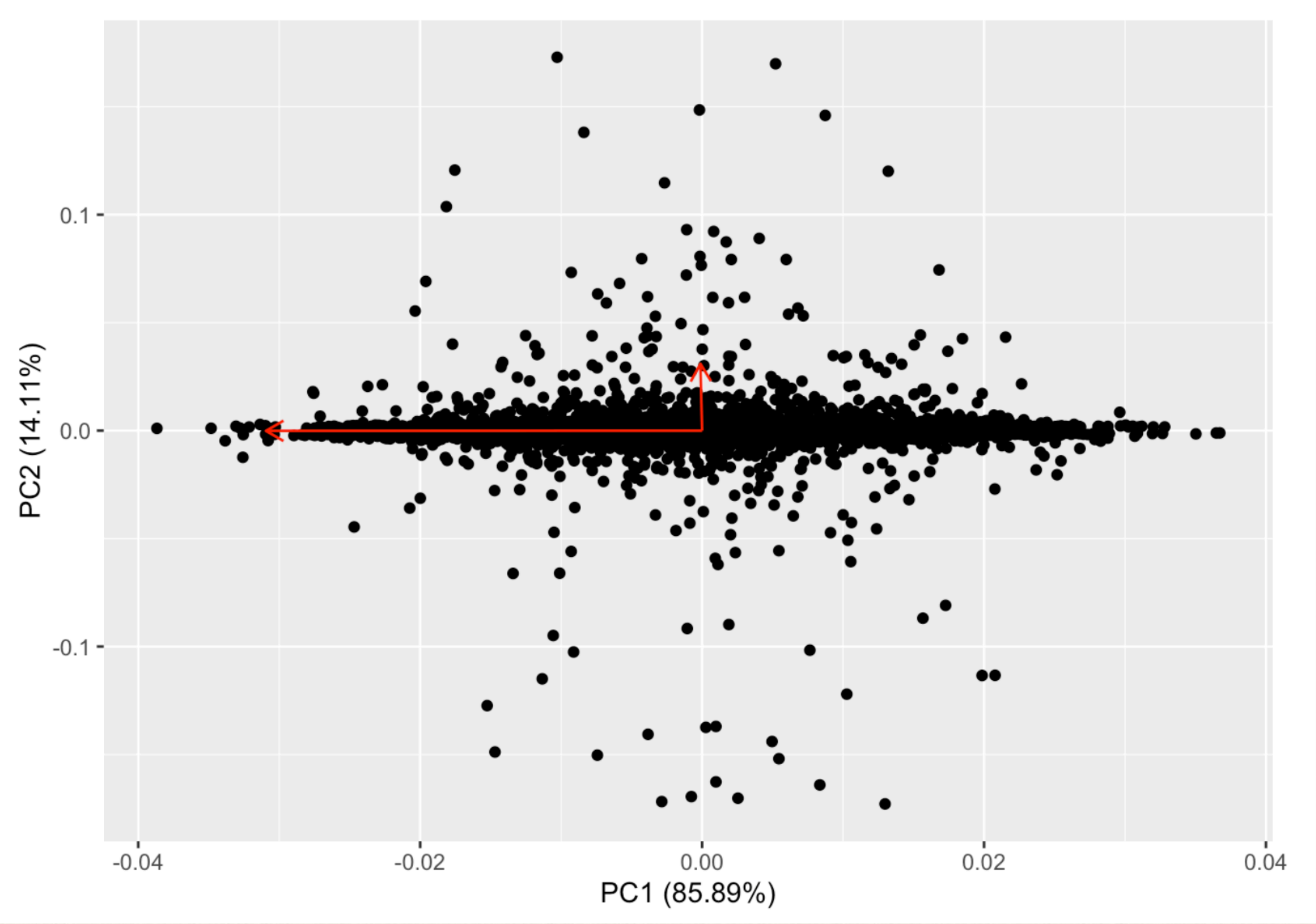
using a gap in the eigenvalues to determine this split.

Estimated active subspaces

- Estimate
$$\int_{\theta} \nabla \log l(\theta) \nabla \log l(\theta)^T \phi(\theta)\, d\theta$$
by
$$\frac{1}{N}\sum_{m=1}^{N} \nabla \log l\left(\theta^m\right) \nabla \log l\left(\theta^m\right)^T$$
for $\theta^i \sim \phi$ for $n = 1 : M$. Then do the eigendecompostion.
- Constantine et al. choose ϕ to be the prior.
- Equivalent to performing a non-centered, non-scaled PCA to determine the directions that explain the most variation in the gradient.

Warwick Statistics

Estimated active subspaces



Warwick Statistics

Toy model

- Consider the family of models.

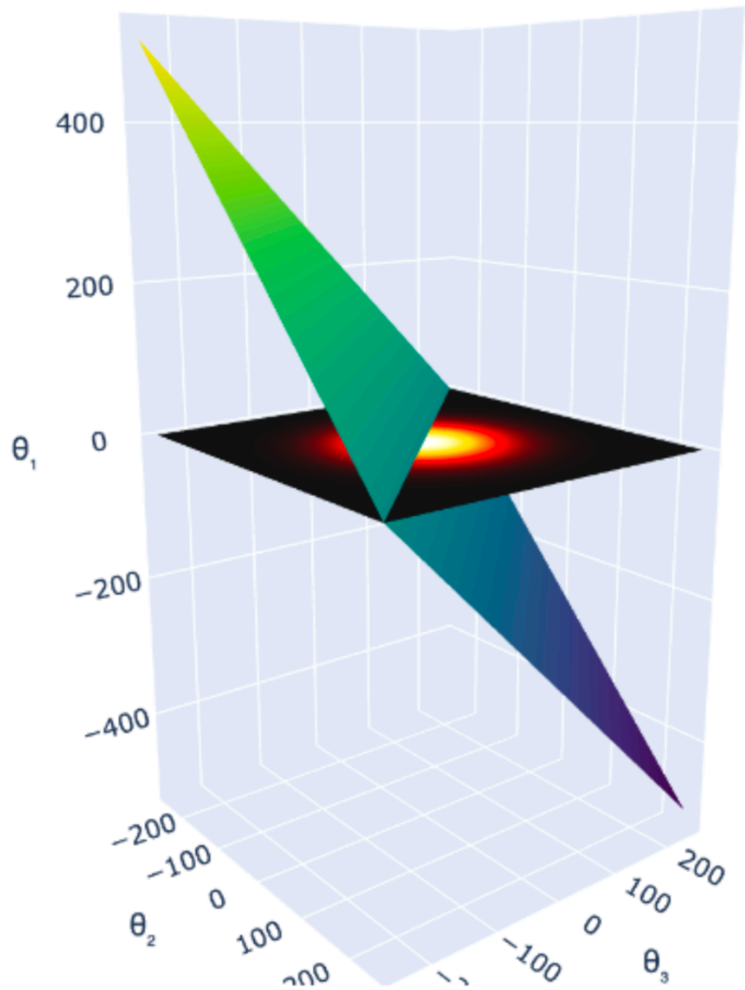
$$\theta \sim \mathscr{MVN}\left(0_d, 1000\mathbb{I}_d\right)$$

$$y \sim \mathscr{MVN}\left(\left(\sum_{i=1}^{d} \theta_i\right) + b\left(\sum_{i=1}^{k} \theta_i^2\right), \mathbb{I}\right)$$

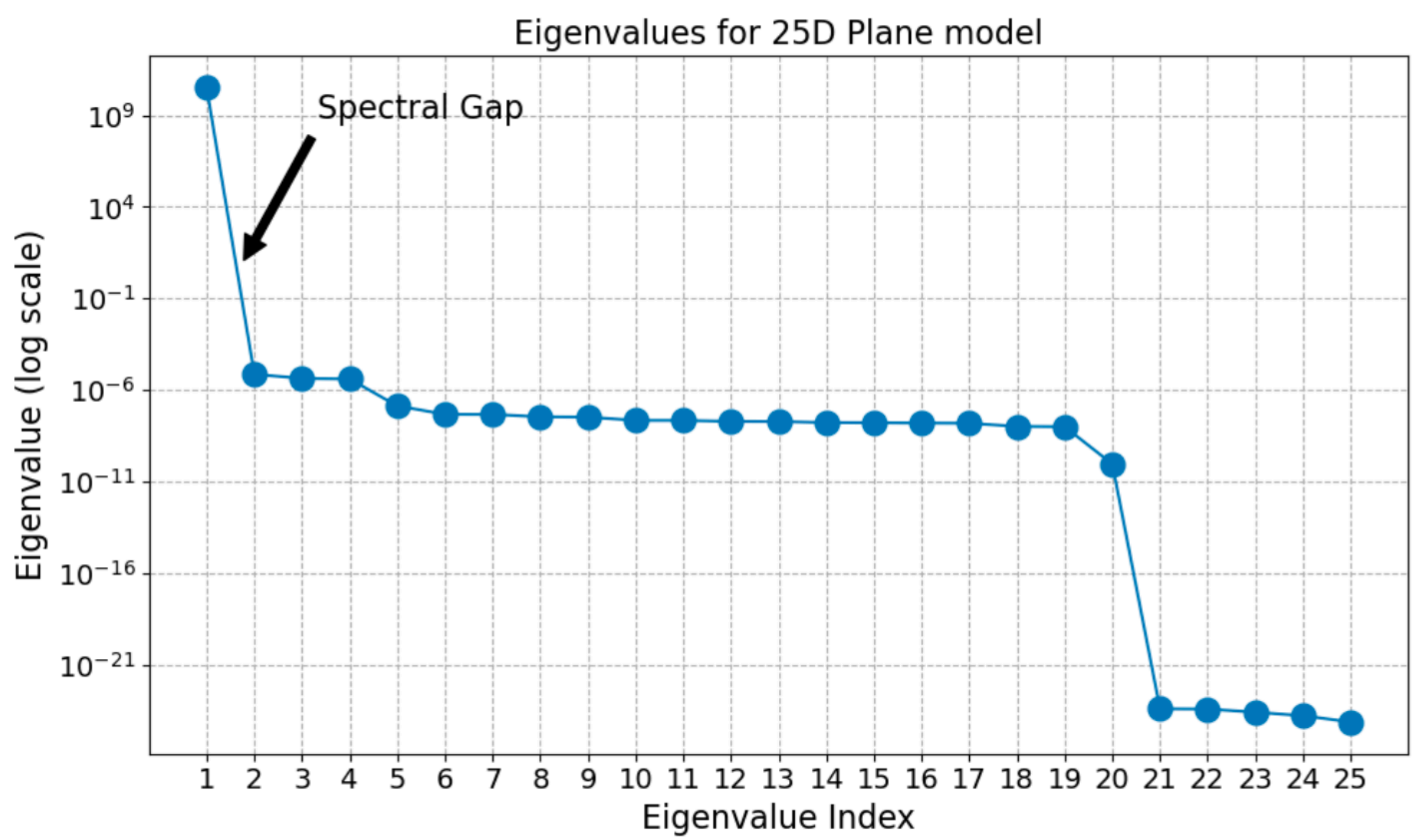
- The parameters are not identifiable.
- For $b = 0$, we are in the linear setting, the ideal case for the active subspace approach.
- For $b \neq 0$, we are in a non-linear setting.

Warwick Statistics

“Plane” model ($b = 0$)

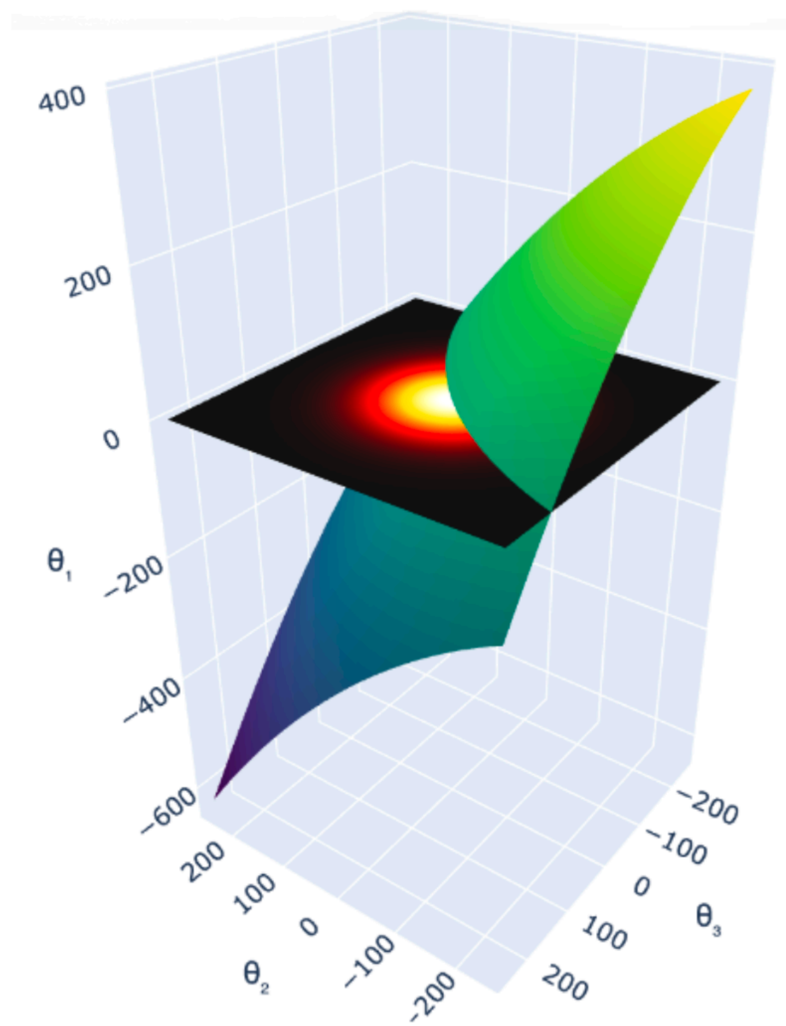


(a) A 2d slice of the Gaussian prior on the horizontal plane $\theta_1 = 0$ together with the level surface of the likelihood in the particular case $\sum_{k=1}^{3} \theta_k = 0$ (green plane).

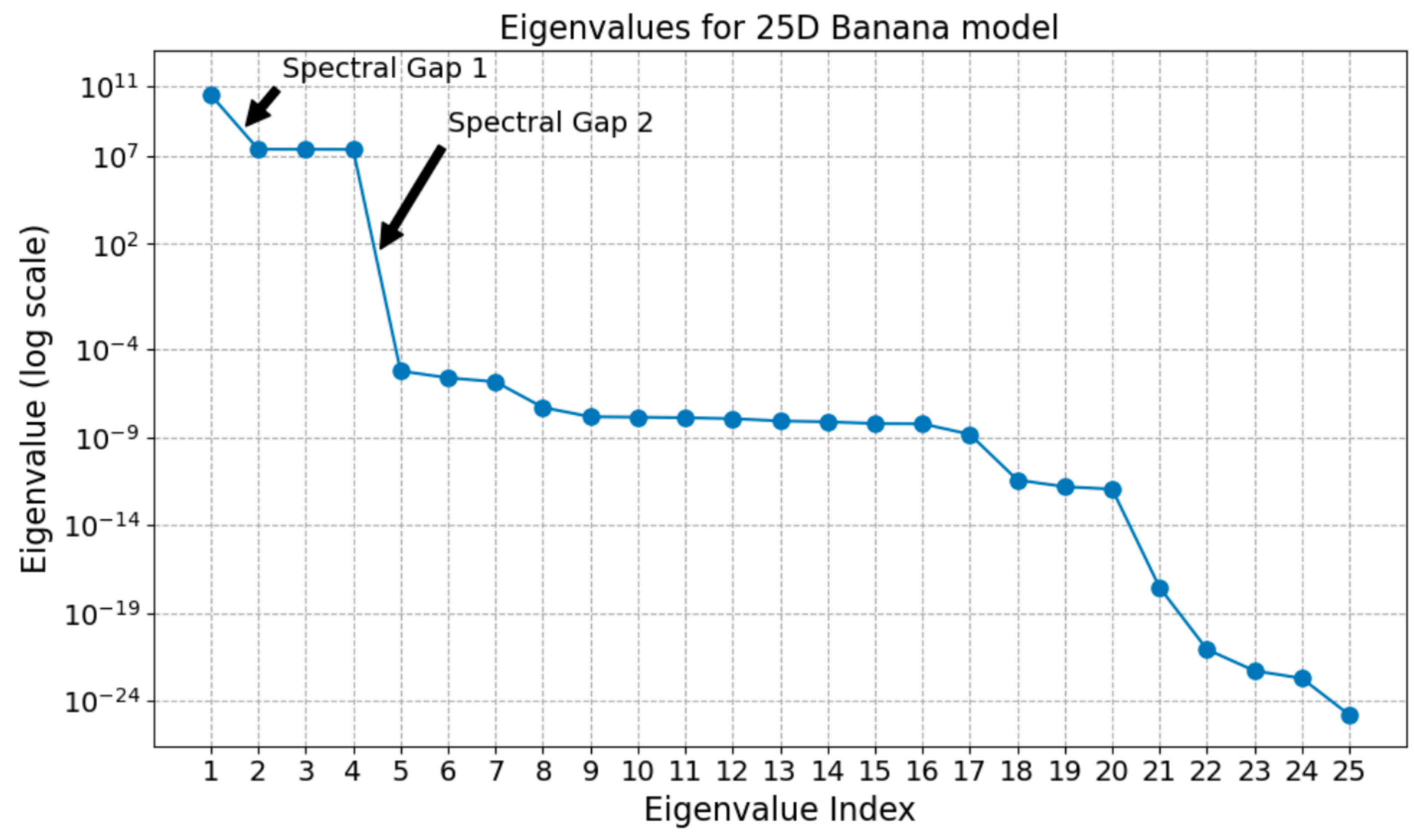


(b) The eigenvalues used in determining the active subspace.

“Banana” model ($b = 0.001$)



(d) A 2d slice of the Gaussian prior on the horizontal plane $\theta_1 = 0$ together with the level surface of the likelihood in the particular case $\sum_{k=1}^{3} \theta_k + 0.001 \times \sum_{k=1}^{3} \theta_k^2 = 0$ (green).



(e) The eigenvalues used in determining the active subspace.

Accelerating MCMC with active subspaces

- **Idea:** run Metropolis-Hastings on a, rather than θ.
- Use the marginal model:

$$\tilde{\pi}_a(a) = p_a(a)\, l_a(a), \text{ where}$$

$$l_a(a) := \int_i p_i\left(i \mid a\right) l\left(B_a a + B_i i\right) di\,.$$

- The marginal likelihood $l_a\left(\theta_a\right)$ is approximated at each iteration using Monte Carlo integration:

$$\hat{l}_a(a) = \frac{1}{N_i}\sum_{n=1}^{N_i} \frac{p_i\left(i^n \mid a\right) l\left(B_a a + B_i i^n\right)}{q_i\left(i^n \mid a\right)},$$

where $i^n \sim q_i\left(\,\cdot \mid a\right)$ for $n = 1 : N_i$.

Active subspace Metropolis-Hastings

- At each iteration:

$$a^{*m} \sim q_a\left(\cdot \mid a^{m-1}\right)$$

- Accept with probability

$$\alpha_a^m = 1 \wedge \frac{p_a\left(a^{*m}\right)\bar{l}_a\left(a^{*m}\right)}{p_a\left(a^{m-1}\right)\bar{l}_a^{m-1}} \frac{q_a\left(a^{m-1} \mid a^{*m}\right)}{q_a\left(a^{*m} \mid a^{m-1}\right)},$$

where

$$\bar{l}_a\left(a^{*m}\right) = \frac{1}{N_i}\sum_{n=1}^{N_i} \frac{p_i\left(i^n \mid a^{*m}\right) l\left(B_a a^{*m} + B_i i^n\right)}{q_i\left(i^n \mid as\right)},$$

where $i^n \sim q_i\left(\cdot \mid \theta_a\right)$ for $n = 1 : N_i$, and $\bar{l}_a^{m-1}$ is the likelihood estimate for a^{m-1}.

Questions

1. Does this MCMC converge to the correct distribution?
2. Estimating the marginal likelihood introduces variance and additional per-iteration cost. Is this worth it?
3. Identifies only global linear identifiable subspaces: is this useful?
4. Is integrating with respect to the prior the best approach to finding the active subspace?
5. Do we really need to artificially split the active and inactive variables?

Warwick Statistics

Active subspace Metropolis-Hastings

1. Does this MCMC converge to the correct distribution?

No!

- Schuster et al. (2017) alter the algorithm so that it becomes a pseudo-marginal Metropolis-Hastings, which does have the desired invariant distribution.

Active subspace Metropolis-Hastings

2. *Estimating the marginal likelihood introduces variance and additional per-iteration cost. Is this worth it?*

Reframing as AS-MH sheds some light on this.

- Recall the marginal likelihood estimate:

$$\hat{l}_a(a) = \frac{1}{N_i} \sum_{n=1}^{N_i} \frac{p_i\left(i^n \mid a\right) l\left(B_a a + B_i i^n\right)}{q_i\left(i^n \mid a\right)}.$$

Warwick Statistics

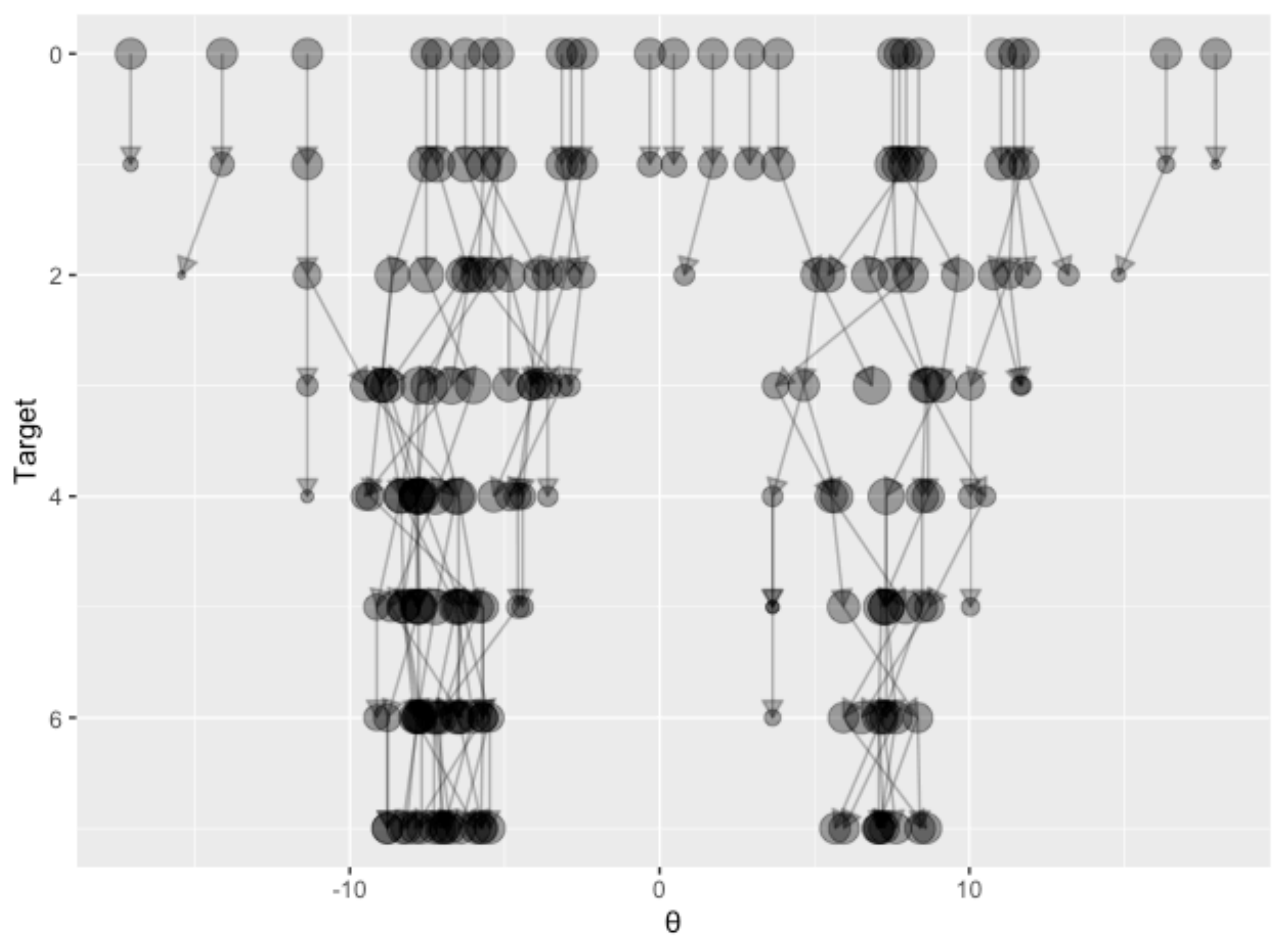
Active subspace Metropolis-Hastings

3. Identifies only global linear identifiable subspaces: is this useful?

Pseudo-marginal framing implies AS-MH will only work well in this setting.

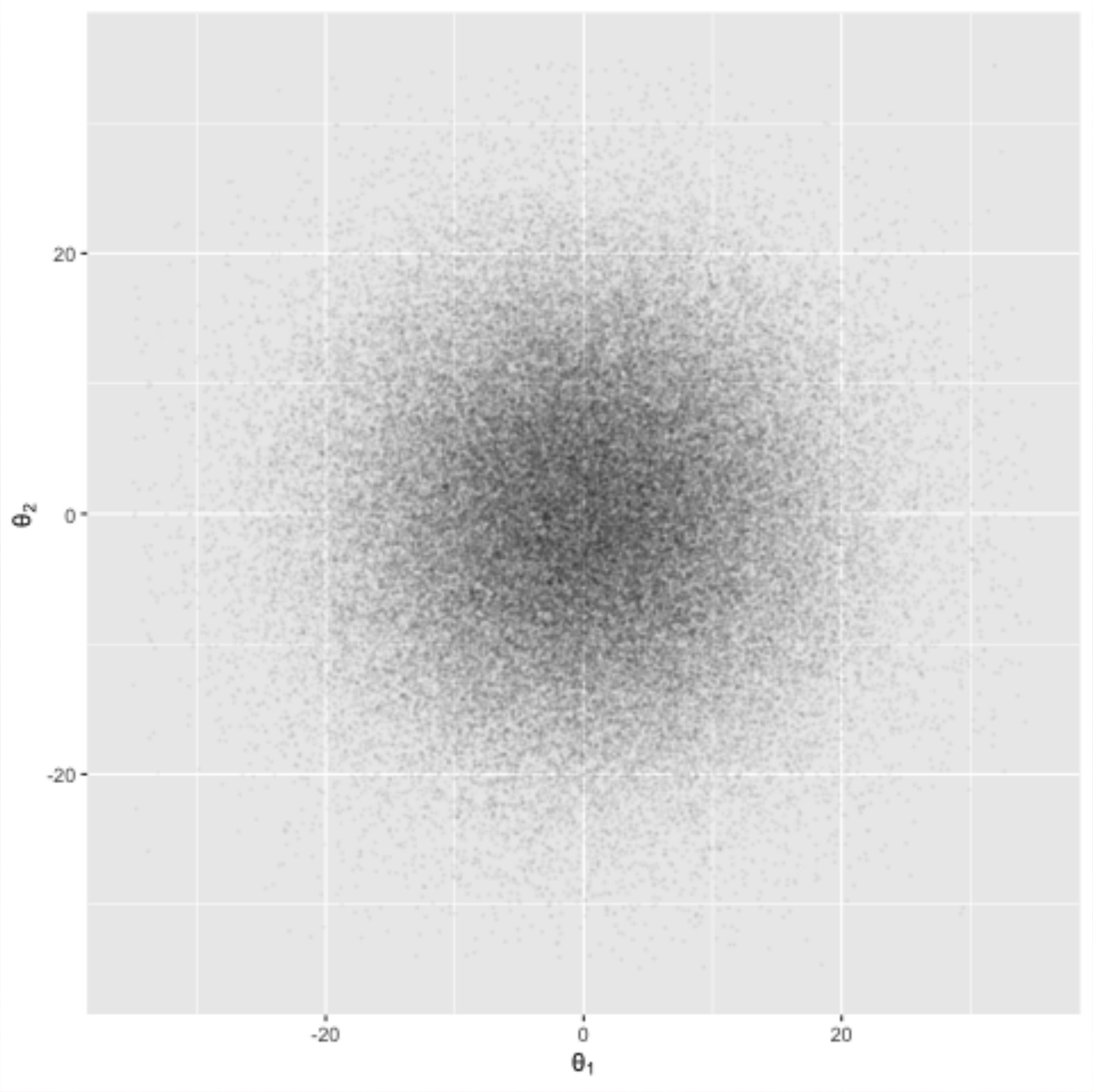
- Inefficient as the dimension of the inactive variables increases
- Could also try "AS-PMMH":
  - use SMC to estimate the marginal likelihood.

Sequential Monte Carlo



Warwick Statistics

Sequential Monte Carlo



Active subspace particle marginal MH (AS-PMMH)

- At each iteration:

$$a^{*m} \sim q_a\left(\cdot \mid a^{m-1}\right)$$

- Accept with probability

$$\alpha_a^m = 1 \wedge \frac{p_a\left(a^{*m}\right) \bar{l}_a\left(a^{*m}\right)}{p_a\left(a^{m-1}\right) \bar{l}_a^{m-1}} \frac{q_a\left(a^{m-1} \mid a^{*m}\right)}{q_a\left(a^{*m} \mid a^{m-1}\right)},$$

where $\bar{l}_a\left(\theta_a^{*m}\right)$ is an estimate from an SMC algorithm with sequence of targets

$$\pi_{t,i}\left(i \mid a\right) = \frac{p_i\left(i \mid a\right) l_{1:t}\left(B_a a + B_i i\right)}{l_{t,a}\left(a\right)},$$

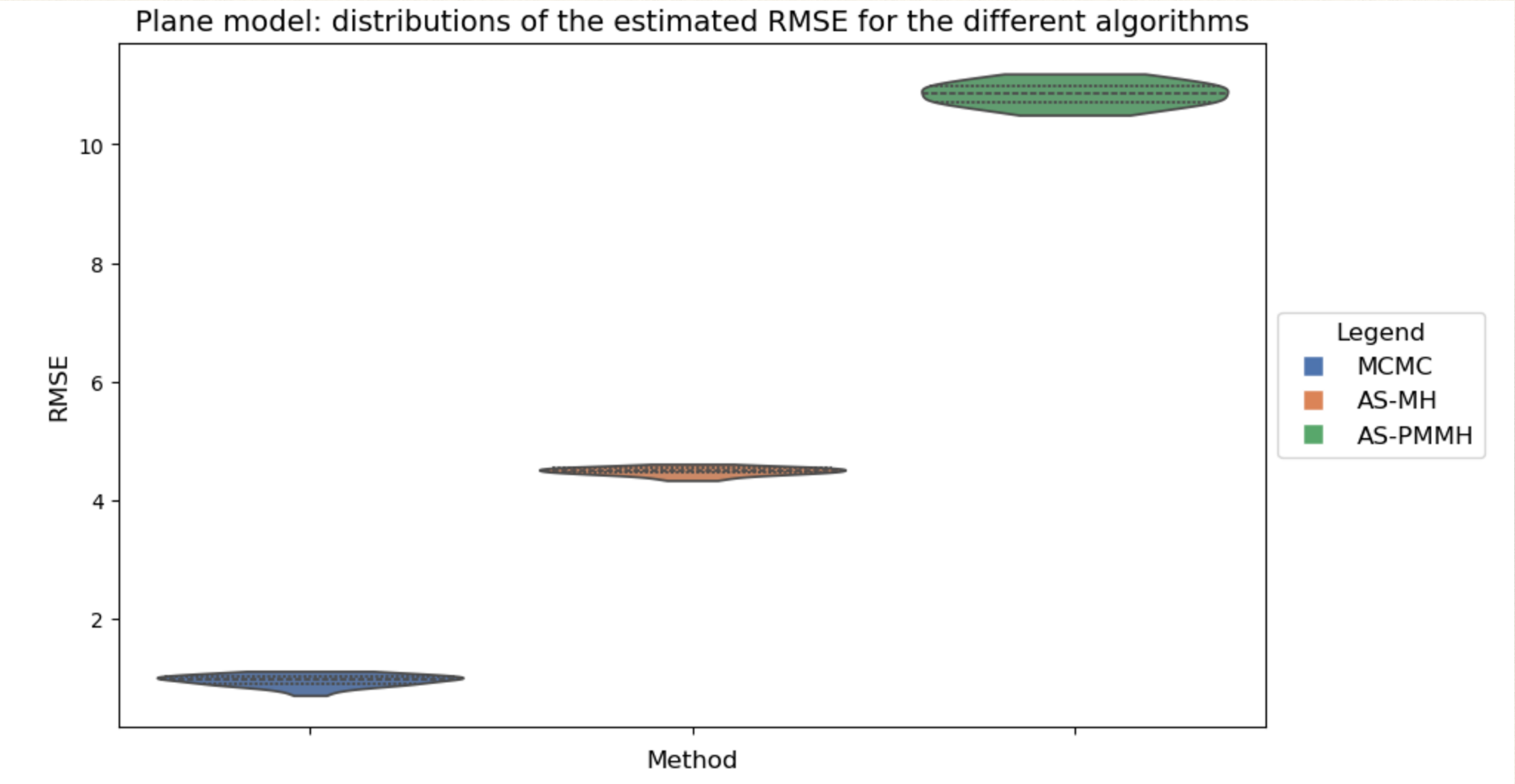
for $t = 1 : T$, and $\bar{l}_a^{m-1}$ is the likelihood estimate for a^{m-1}.

- Here

$$l_{1:t}(\theta) = \prod_{s=1}^{t} l_s(\theta), \text{ for example } l(\theta) = \prod_{s=1}^{T} l^{\eta_t - \eta_{t-1}}(\theta).$$

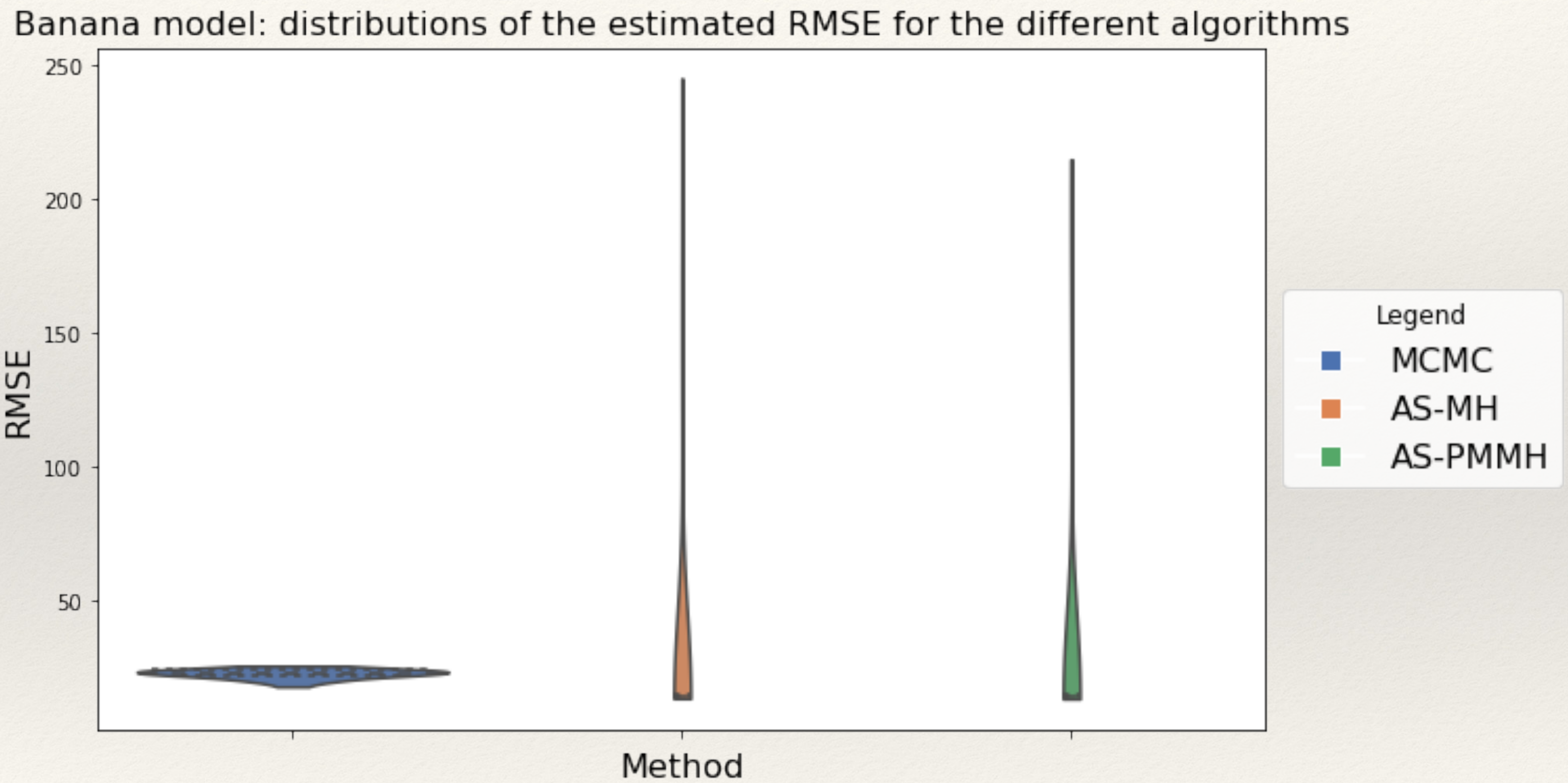
Warwick Statistics

Results: plane



Warwick Statistics

Results: banana



Warwick Statistics

An alternative: Gibbs sampler

- Factorise $p(a, i) = p_a(a) p_{i|a}(i \mid a)$.
- Gibbs sampler:
  - $i \sim p_{i|a}(i \mid a) l \left(B_a a + B_i i \right)$
  - $a \sim p_a(a) l \left(B_a a + B_i i \right)$

Warwick Statistics

An alternative: Gibbs sampler

- Factorise $p(a, i) = p_a(a)p_{i|a}(i \mid a)$.
- Gibbs sampler:
  - $i \sim p_{i|a}(i \mid a)$
  - $a \sim p_a(a)l\left(B_a a + B_i i\right)$

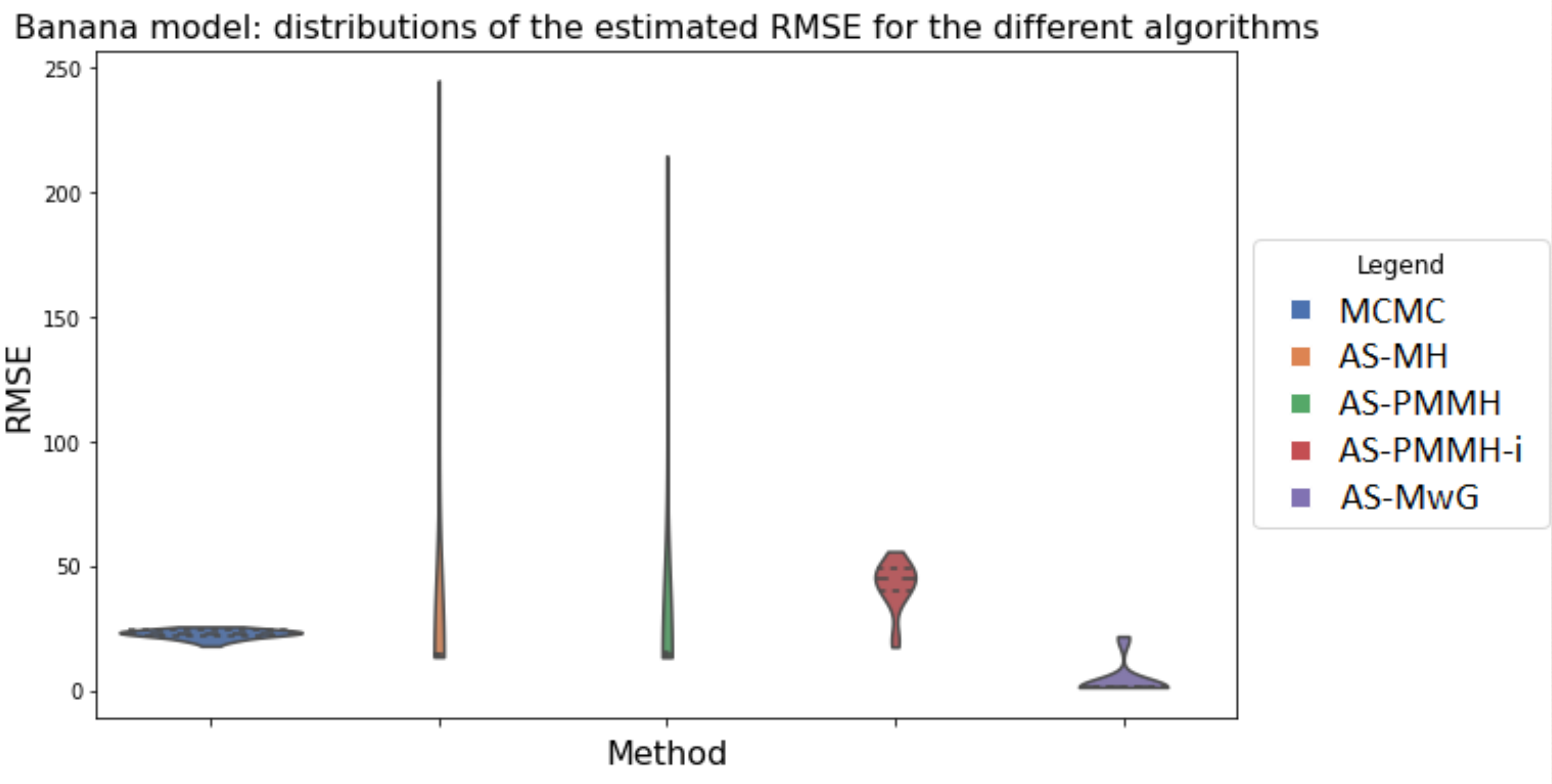
AS - Metropolis-within-Gibbs

- Factorise $p(a, i) = p_a(a)p_{i|a}(i \mid a)$.
- Gibbs sampler:
  - $i \sim p_{i|a}(i \mid a)l\left(B_a a + B_i i\right)$ using MH move with proposal $p_{i|a}$
  - $a \sim p_a(a)l\left(B_a a + B_i i\right)$ using MH move

1. Avoids estimating marginal likelihood.
2. Avoids computational focus on inactive variables.

Warwick Statistics

Results: banana



Warwick Statistics

Discussion

- Ripoli, L. and Everitt, R. G. (2025) Improved MCMC with active subspaces, arXiv.
  - also a conditional SMC algorithm for exploring the active variables
  - paper in revision.

- More generally:
  - determining a useful reparameterisation automatically.

Active subspace Metropolis-Hastings

4. *Is integrating with respect to the prior the best approach to finding the active subspace?*

Ripoli, L. and Everitt, R. G. (2024) Sequential Monte Carlo with active subspaces, arXiv.

Active subspace SMC

- Let us instead consider running an SMC on the active variables.
- Concept: consider a sequence of targets with likelihood $l_{1:t}\left(B_a a + B_i i\right)$.
- To begin, consider a sequence of targets proportional to

$$p_a(a)\prod_{j=1}^{N_i} q_{t,i}\left(i^n \mid a\right)\frac{1}{N_i}\sum_{n=1}^{N_i}\frac{p_i\left(i^n \mid a\right) l_{1:t}\left(B_a a + B_i i^n\right)}{q_{t,i}\left(i^n \mid a\right)}.$$

- We call this algorithm active subspace SMC (AS-SMC).

Warwick Statistics

Active subspace SMC²

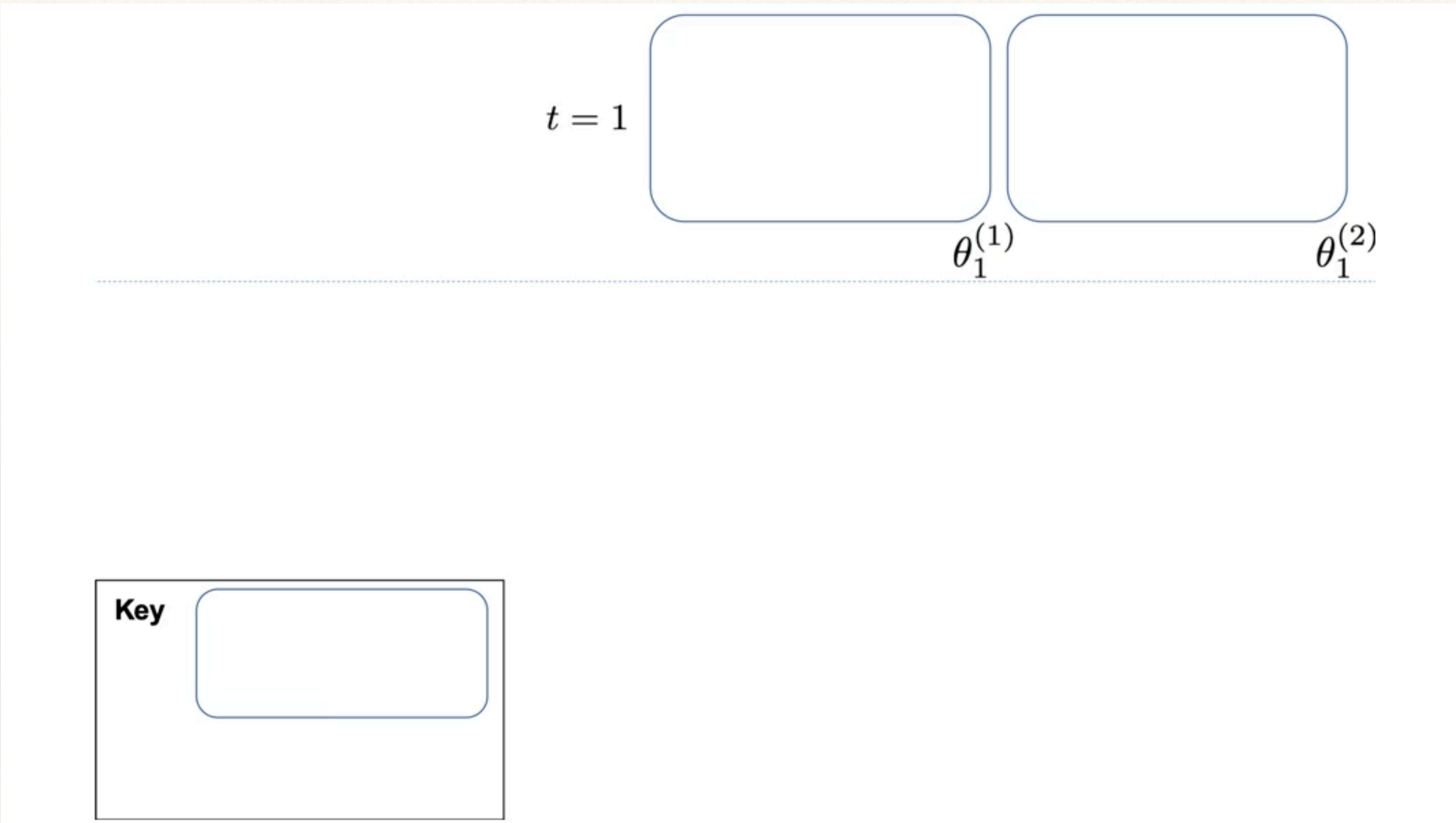
- AS-SMC inherits the weaknesses of AS-MH
  - it uses the an IS estimator for the marginal likelihood.
- We can improve it by using an SMC estimator of the marginal likelihood for each SMC particle on a
  - nested SMC approach that follows Chopin et al. (2013).
- Target distribution at each external SMC iteration is:

$$\pi_t\left(a, \left\{i^n_{0:t-1}, v^n_{0:t-1}\right\}_{n=1}^{N_i}\right) = p_a(a)\,\psi_{t-1}\left(\left\{i^n_{0:t-1}, v^n_{0:t-1}\right\}_{n=1}^{N_i} \mid a\right)\frac{\hat{l}_{t,a}(a)}{Z_t},$$

where ψ_{t-1} is the distribution of all of the random variables generated by the internal SMC up to time t.

Chopin, Jacob, and Papaspiliopoulos (2013). SMC² an efficient algorithm for sequential analysis of state space models. Journal of the Royal Statistical Society: Series B.

Structure of SMC²



Kerama, Thorne and Everitt (2022). Rare event ABC-SMC², arXiv.

AS-SMC and AS-SMC² results