Stratified Markov Chain Monte Carlo

Brian Van Koten
University of Massachusetts, Amherst
Department of Mathematics and Statistics

with A. Dinner, J. Tempkin, E. Thiede, B. Vani, and J. Weare

June 28, 2019

Sampling Problems

What is the probability of finding a protein in a given conformation?

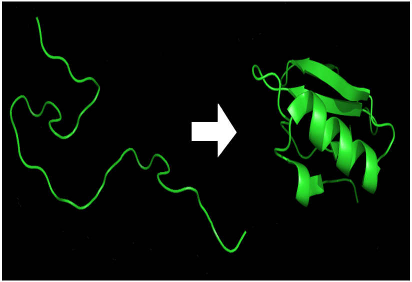


Figure from Folding@home

Compute sample from *Boltzmann distribution*.

Bayesian inference for ODE model of circadian rhythms.

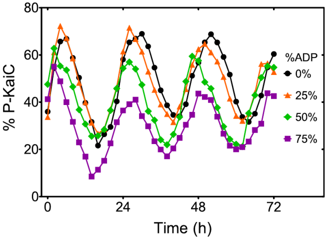


Figure from Phong, et al, PNAS, 2012

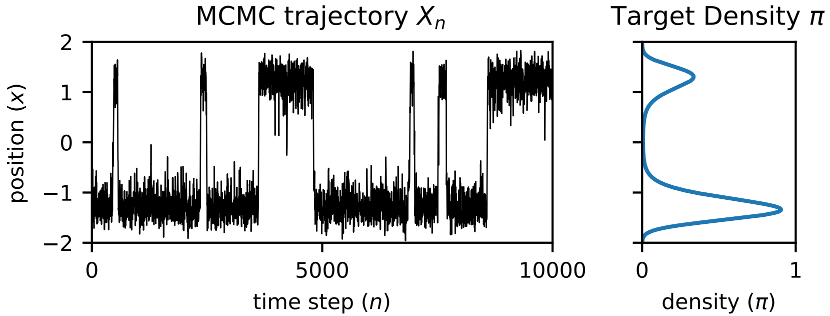
Compute sample from *posterior distribution*.

Markov Chain Monte Carlo (MCMC)

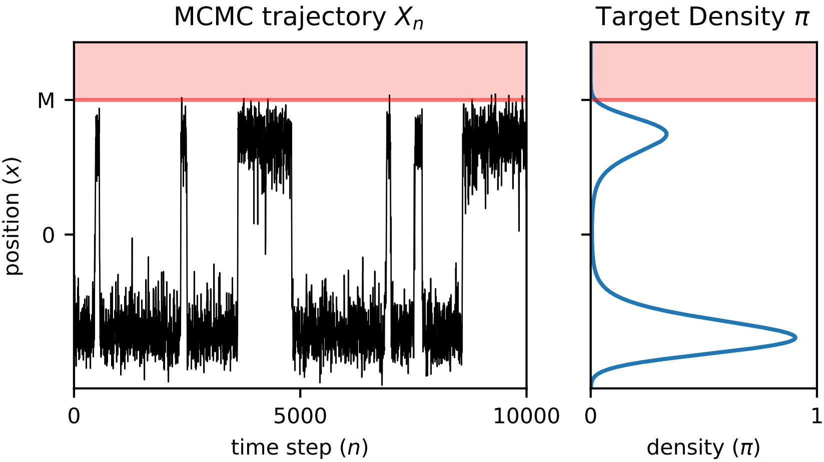
Goal: Compute $\pi(g) := \int g(x)\pi(dx)$.

MCMC Method: Choose Markov chain X_n so that

$$\underbrace{\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} g(X_n) = \pi(g).}_{\text{“}X_n \text{ samples } \pi.\text{”}}$$



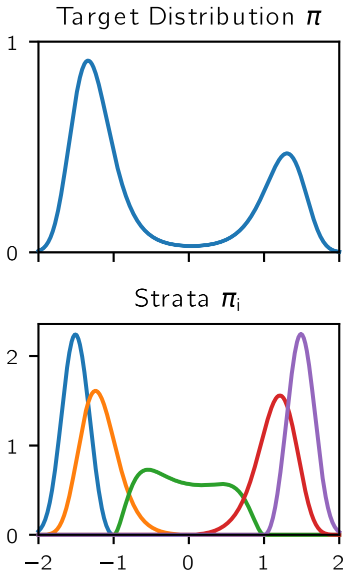
Difficulties with MCMC



Multimodality: Multimodality $\Longrightarrow$ slow convergence

Tails: Need large sample to compute small probabilities, e.g. $\pi([M, \infty))$.

Sketch of Stratified MCMC



1. Choose family of *strata*, i.e. distributions π_i whose supports cover support of target π.
2. Sample strata by MCMC.
3. Estimate $\pi(g)$ from samples of strata.

Typical Strata: $\pi_i(dx) \propto \psi_i(x)\pi(dx)$
for "localized" ψ_i.

Why Stratify?

- Strata may be *unimodal*, even if π is *multimodal*
- Can concentrate sampling in *tail*

History of Stratification

Surveys: [Russian census, late 1800s], [Neyman, 1937]
Bayes factors: [Geyer, 1994]
Selection bias models: [Vardi, 1985]

Free energy: [Umbrella Sampling, 1977], [WHAM, 1992], [MBAR, 2008]
Ion channels: [Berneche, et al, 2001]
Protein folding: [Boczko, et al, 1995]

Problems:

1. WHAM/MBAR are complicated iterative methods ...
2. No clear *story* explaining benefits of stratification.
3. Stratification underappreciated as a *general* strategy.
4. Need good *error bars* for adaptivity.

History of Stratification

Surveys: [Russian census, late 1800s], [Neyman, 1937]
Bayes factors: [Geyer, 1994]
Selection bias models: [Vardi, 1985]

Free energy: [Umbrella Sampling, 1977], [WHAM, 1992], [MBAR, 2008]
Ion channels: [Berneche, et al, 2001]
Protein folding: [Boczko, et al, 1995]

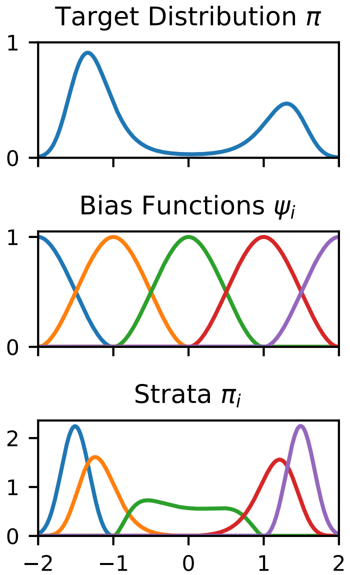
Problems:

1. WHAM/MBAR are complicated iterative methods ...
2. No clear *story* explaining benefits of stratification.
3. Stratification underappreciated as a *general* strategy.
4. Need good *error bars* for adaptivity.

BvK, et al: Propose Eigenvector Method for Umbrella Sampling, develop *story*, *error bars*, stratification for *dynamical quantities* ...

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]



- **Bias Functions:** $\{\psi_i\}_{i=1}^{L}$ with

$$\sum_{i=1}^{L} \psi_i(x) = 1 \text{ and } \psi_i(x) \geq 0.$$

Note: User chooses bias functions.

- **Weights:** $z_i = \pi(\psi_i)$

- **Strata:** $\pi_i(dx) = z_i^{-1}\psi_i(x)\pi(dx)$

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]

Goal: Write $\pi(g)$ in terms of averages over strata $\pi_i(dx) = \frac{\psi_i(x)\pi(dx)}{z_i}$.

First, decompose $\pi(g)$ as weighted sum:

$$\pi(g) = \int g(x) \underbrace{\sum_{i=1}^{L} \psi_i(x)}_{\psi_i\text{'s sum to one}} \pi(dx)$$

$$= \sum_{i=1}^{L} z_i \int g(x) \underbrace{\frac{\psi_i(x)\pi(dx)}{z_i}}_{\pi_i(dx)} = \sum_{i=1}^{L} z_i \pi_i(g).$$

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]

Goal: Write $\pi(g)$ in terms of averages over strata $\pi_i(dx) = \frac{\psi_i(x)\pi(dx)}{z_i}$.

First, decompose $\pi(g)$ as weighted sum:

$$\pi(g) = \int g(x) \underbrace{\sum_{i=1}^{L} \psi_i(x)}_{\psi_i\text{'s sum to one}} \pi(dx)$$

$$= \sum_{i=1}^{L} z_i \int g(x) \underbrace{\frac{\psi_i(x)\pi(dx)}{z_i}}_{\pi_i(dx)} = \sum_{i=1}^{L} z_i \pi_i(g).$$

How to express weights $z_i = \pi(\psi_i)$ as averages over strata?

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]

Goal: Write $\pi(g)$ in terms of averages over strata $\pi_i(dx) = \frac{\psi_i(x)\pi(dx)}{z_i}$.

First, decompose $\pi(g)$ as weighted sum: $\pi(g) = \sum_{i=1}^{L} z_i \pi_i(g)$.

How to express weights $z_i = \pi(\psi_i)$ *as averages over strata?*

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]

Goal: Write $\pi(g)$ in terms of averages over strata $\pi_i(dx) = \frac{\psi_i(x)\pi(dx)}{z_i}$.

First, decompose $\pi(g)$ as weighted sum: $\pi(g) = \sum_{i=1}^{L} z_i \pi_i(g)$.

How to express weights $z_i = \pi(\psi_i)$ as averages over strata?

$$z_j = \pi(\psi_j) = \sum_{i=1}^{L} z_i \pi_i(\psi_j) \quad \Longleftrightarrow \quad \underbrace{z^T = z^T F}_{\text{eigenproblem}}, \text{ where } \underbrace{F_{ij} = \pi_i(\psi_j)}_{\text{overlap matrix}}.$$

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]

Goal: Write $\pi(g)$ in terms of averages over strata $\pi_i(dx) = \frac{\psi_i(x)\pi(dx)}{z_i}$.

First, decompose $\pi(g)$ as weighted sum: $\pi(g) = \sum_{i=1}^{L} z_i \pi_i(g)$.

To express weights $z_i = \pi(\psi_i)$ as averages over strata,

$$\underbrace{z^T = z^T F}_{\text{eigenproblem}}, \text{ where } \underbrace{F_{ij} = \pi_i(\psi_j)}_{\text{overlap matrix}}.$$

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]

Goal: Write $\pi(g)$ in terms of averages over strata $\pi_i(dx) = \frac{\psi_i(x)\pi(dx)}{z_i}$.

First, decompose $\pi(g)$ as weighted sum: $\pi(g) = \sum_{i=1}^{L} z_i \pi_i(g)$.

To express weights $z_i = \pi(\psi_i)$ as averages over strata,

$$\underbrace{z^T = z^T F}_{\text{eigenproblem}}, \text{ where } \underbrace{F_{ij} = \pi_i(\psi_j)}_{\text{overlap matrix}}.$$

Why does eigenproblem determine z?

1. F is stochastic; z is a probability vector.
2. If F irreducible, z is *unique* solution of eigenproblem.

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]

Recall: $\pi(g) = \sum_{i=1}^{L} z_i \pi_i(g)$, and $z^T = z^T F$ for $F_{ij} = \pi_i(\psi_j)$.

EMUS Algorithm:

1. Choose bias functions ψ_i and processes X_n^i sampling the strata.
2. Compute $\bar{g}_i := \frac{1}{N_i} \sum_{n=1}^{N_i} g(X_n^i)$ to estimate $\pi_i(g)$.
3. Compute $\bar{F}_{ij} := \frac{1}{N_i} \sum_{n=1}^{N_i} \psi_j(X_n^i)$ to estimate F.
4. Solve eigenproblem $\bar{z}^T = \bar{z}^T \bar{F}$ to estimate weights z.
5. Output $g^{\mathrm{EM}} = \sum_{i=1}^{L} \bar{z}_i \bar{g}_i$.

Eigenvector Method for Umbrella Sampling (EMUS)

[BvK, et al]

Recall: $\pi(g) = \sum_{i=1}^{L} z_i \pi_i(g)$, and $z^T = z^T F$ for $F_{ij} = \pi_i(\psi_j)$.

EMUS Algorithm:

1. Choose bias functions ψ_i and processes X_n^i sampling the strata.
2. Compute $\bar{g}_i := \frac{1}{N_i} \sum_{n=1}^{N_i} g(X_n^i)$ to estimate $\pi_i(g)$.
3. Compute $\bar{F}_{ij} := \frac{1}{N_i} \sum_{n=1}^{N_i} \psi_j(X_n^i)$ to estimate F.
4. Solve eigenproblem $\bar{z}^T = \bar{z}^T \bar{F}$ to estimate weights z.
5. Output $g^{\mathrm{EM}} = \sum_{i=1}^{L} \bar{z}_i \bar{g}_i$.

Key Point: Simplicity of EMUS enables analysis of stratification.

EMUS Analysis: Outline

1. **Sensitivity of g^{EM} to sampling error.**

2. Dependence of sampling error on choice of strata.

3. Stories involving multimodality and tails.

Quantifying Sensitivity to Sampling Error I

- For F irreducible and stochastic, let $z(F)$ be the unique solution of

$$z(F)^T = z(F)^T F.$$

- $\mathbf{P}_i^F[\tau_j < \tau_i]$: probability of hitting j before i, conditioned on starting from i, for a Markov chain on $1, \ldots, L$ with transition matrix F.

- **Theorem** [BvK, et al]**:**

$$\frac{1}{2}\frac{1}{\mathbf{P}_i^F[\tau_j < \tau_i]} \leq \max_{m=1,\ldots L}\left|\frac{\partial \log z_m}{\partial F_{ij}}(F)\right| \leq \frac{1}{\mathbf{P}_i^F[\tau_j < \tau_i]} \leq \frac{1}{F_{ij}}.$$

- Led to new perturbation bounds for Markov chains [BvK, et al].

Quantifying Sensitivity to Sampling Error II

Assumption: CLT holds for MCMC averages:

$$\sqrt{N_i}(\bar{g}_i - \pi_i(g)) \xrightarrow{d} \mathbf{N}(0, \underbrace{\mathrm{C}\left(\bar{g}_i\right)}_{\text{asymptotic variance}}).$$

Theorem [BvK, et al]**:** $\sqrt{N}\left(g^{\mathrm{EM}} - \pi(g)\right) \xrightarrow{d} \mathbf{N}\left(0, \mathrm{C}\left(g^{\mathrm{EM}}\right)\right)$, where

$$\frac{\mathrm{C}\left(g^{\mathrm{EM}}\right)}{\mathrm{var}_\pi(g)} \lesssim \sum_{i=1}^{L} \underbrace{\left(\sum_{\substack{j \neq i \\ F_{ij} > 0}} \frac{1}{\mathbf{P}_i^F[\tau_j < \tau_i]^2}\right)}_{\text{sensitivity to error in } \bar{F}} \times \underbrace{\frac{\sum_{j=1}^{L} \mathrm{C}\left(\bar{F}_{ij}\right)}{\kappa_i}}_{\text{error in } \bar{F}} + z_i^2 \frac{\mathrm{C}\left(\bar{g}_i\right)}{\kappa_i}.$$

Notation: N is total sample size, with $N_i = \kappa_i N$ from π_i.

EMUS Analysis: Outline

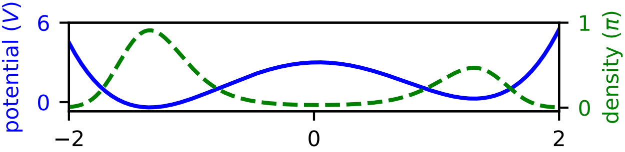
1. Sensitivity of g^{EM} to sampling error.

2. **Dependence of sampling error on choice of strata.**

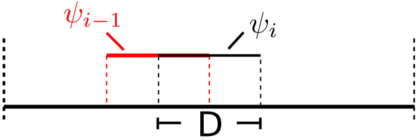
3. Stories involving multimodality and tails.

Dependence of Sampling Error on Strata I

Write $\pi(dx) = Z^{-1}\exp(-V(x)/\varepsilon)$ for some *potential V*:



Assume bias functions ψ_i piecewise constant:

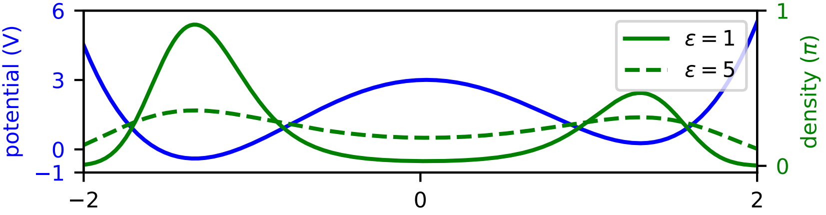


Assume X_t^i is overdamped Langevin with reflecting boundaries:

$$dX_t^i = -\underbrace{\nabla V(X_t^i)dt}_{\text{gradient descent}} + \underbrace{\sqrt{2\varepsilon}dB_t^i}_{\text{noise}} + \text{reflecting BCs}$$

Dependence of Sampling Error on Strata II

Let $\pi(dx) = Z^{-1}\exp(-V(x)/\varepsilon)$ for some *potential V*:



Theorem **[BvK, et al]:** For overdamped Langevin with reflecting BCs,

$$\frac{\mathrm{C}(\bar{g}_i)}{\mathrm{var}_{\pi_i}(g)} \lesssim \underbrace{\frac{D^2}{\varepsilon}}_{\text{diffusion scaling}} \times \underbrace{\exp\left(\frac{\max_{\mathrm{supp}\,\pi_i} V - \min_{\mathrm{supp}\,\pi_i} V}{\varepsilon}\right)}_{\text{Arrhenius}}.$$

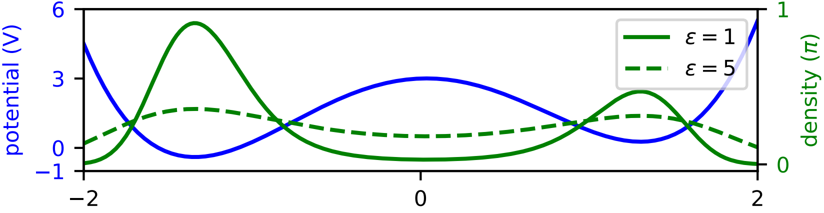
Notation: D is diameter of support of π_i.

EMUS Analysis: Outline

1. Dependence of sampling error on choice of strata.
2. Sensitivity of g^{EM} to sampling error.
3. **Stories involving multimodality and tails.**

EMUS and Multimodality

Let $\pi(dx) = Z^{-1} \exp(-V(x)/\varepsilon)$ for double well V:

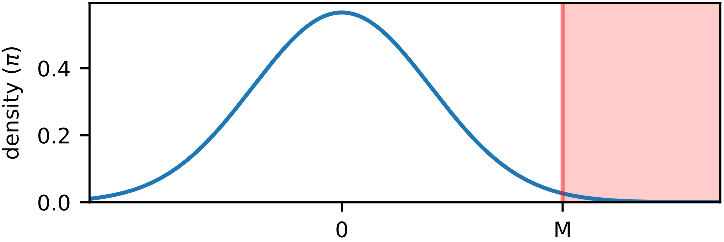


Asymptotic variance of naïve MCMC grows *exponentially* as $\varepsilon \downarrow 0$.

Theorem **[BvK, et al]:**
For right choice of strata ($L \propto \varepsilon^{-1}$),
asymptotic variance of EMUS estimate g^{EM} grows *polynomially* as $\varepsilon \downarrow 0$.

EMUS and Tails

Goal: Compute $\pi([M,\infty)) = \int_M^\infty \pi(dx)$.

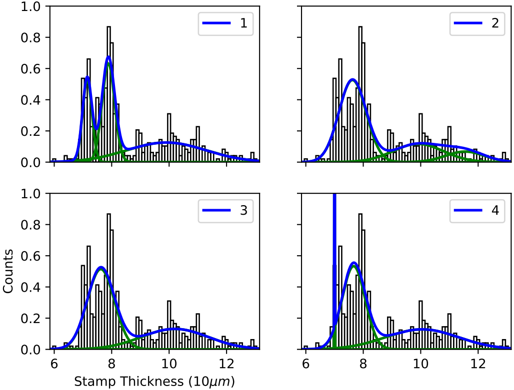


For a broad class of distributions π,
relative asymptotic variance of MCMC grows *exponentially* as $M \uparrow \infty$.

Theorem [BvK, et al]:
For right choice of strata,
relative asymptotic variance of EMUS grows *polynomially* as $M \uparrow \infty$.

Example: EMUS for Bayesian Inference

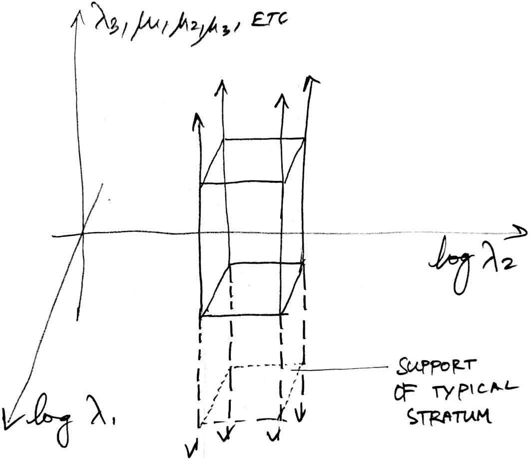
- **Goal:** Fit set of thicknesses of 485 stamps by mix of 3 Gaussians:



- **Parameters:** means $\mu_1 \leq \mu_2 \leq \mu_3$, precisions $\lambda_1, \lambda_2, \lambda_3$, weights, etc
- **Bayesian method**: Define *posterior distribution* on parameter space.

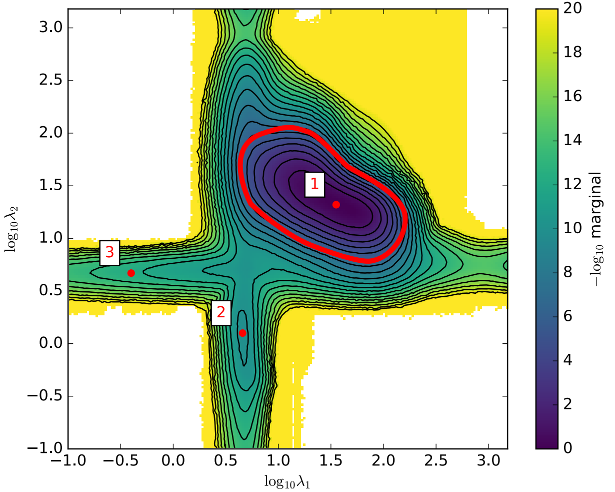
Example: EMUS for Bayesian Inference

- **Parameters:** means $\mu_1 \le \mu_2 \le \mu_3$, precisions $\lambda_1, \lambda_2, \lambda_3$, weights, etc
- **Objective:** Compute marginal in $\log_{10} \lambda_1$ and $\log_{10} \lambda_2$.
- **Strata**: Cylinders over grid of regions in $\log_{10} \lambda_1$, $\log_{10} \lambda_2$ plane:



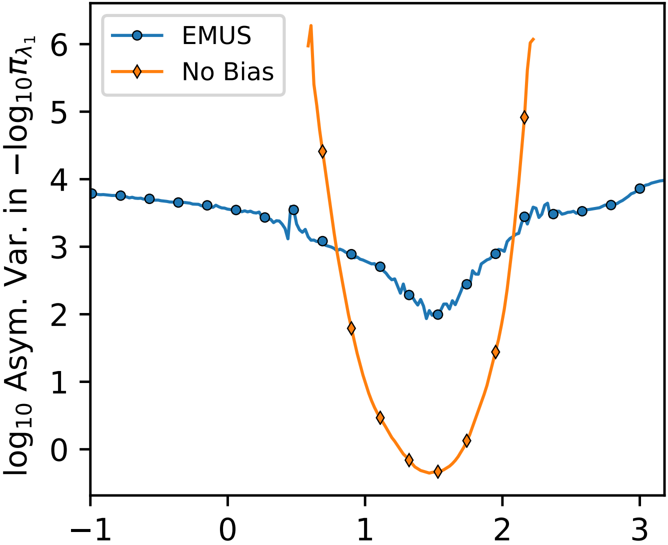
Example: EMUS for Bayesian Inference

- **Parameters:** means $\mu_1 \leq \mu_2 \leq \mu_3$, precisions $\lambda_1, \lambda_2, \lambda_3$, weights, etc
- **Objective:** Compute marginal in $\log_{10} \lambda_1$ and $\log_{10} \lambda_2$.



Example: EMUS for Bayesian Inference

Asymptotic variances of EMUS vs. unbiased MCMC for marginal in $\log \lambda_1$:



Conclusions

- We present and analyze EMUS, a stratified MCMC method, and we derive practical error bars for EMUS estimator [BvK et al, JCP, 2016].
- Our analysis required development of new perturbation estimates for stochastic matrices [BvK et al, SIMAX, 2015].
- We clearly identify classes of problems for which stratification is beneficial, and we propose novel applications in statistics [BvK et al, 2019+].
- We analyze and improve a stratification method for computing dynamical quantities [BvK et al, SIREV, 2017].
- **Ongoing Work:** Convergence of NEUS, automatic methods for determining strata, comparison with other rare event sampling methods, . . .