Approach II: energy conserving subsampling

Using unbiased estimators for the re-weightings/gradients/rejection steps (so the error won't scale with ε)

Target perturbed posterior:

$$\overline{\pi}_m(\theta, u) \propto \widehat{L}_m(\theta) p_\Theta(\theta) p_U(u),$$

For the Markov moves, perform Gibbs update using block Metropolis (within Gibbs; for subsampling indices, discrete and ill-suited for HMC) + HMC (again within Gibbs; for parameter vector)

1. $u | \theta, \vec{p}, y$
2. $\theta, \vec{p} | u, y$

$$\overline{\pi}_m(\theta, \vec{p}, u) \propto \exp\left(-\widehat{\mathcal{H}}(\theta, \vec{p})\right) p_U(u), \quad \widehat{\mathcal{H}}(\theta, \vec{p}) = \widehat{\mathcal{U}}(\theta) + \mathcal{K}(\vec{p})$$

$$\alpha_u = \min\left\{1, \frac{\widehat{L}_m(\theta^{(j-1)}; u')}{\widehat{L}_m(\theta^{(j-1)}; u^{(j-1)})}\right\}$$

$$\widehat{\mathcal{U}}(\theta) = -\log \widehat{L}_m(\theta) - \log p_\Theta(\theta) \quad \text{and } \mathcal{K}(\vec{p}) = \frac{1}{2}\vec{p}\,' M^{-1} \vec{p},$$

(then marginalize over momentum/indices for Θ samples)

Dang et al, Hamiltonian Monte Carlo with Energy Conserving Subsampling, 2019
Gunawan et al, Subsampling Sequential Monte Carlo for Static Bayesian Models, 2020
Tran et al, The Block Pseudo-Marginal Sampler, 2017

Approach II: energy conserving subsampling

$$\widehat{L}_m(\theta) = \exp\left(\widehat{\ell}_m(\theta) - \frac{1}{2}\widehat{\sigma}^2_m(\theta)\right)$$

$$\widehat{\ell}_m(\boldsymbol{\theta}) = \sum_{k=1}^{n} q_k(\boldsymbol{\theta}) + \frac{n}{m}\sum_{i=1}^{m} \ell_{u_j}(\boldsymbol{\theta}) - q_{u_j}(\boldsymbol{\theta}), \quad u_j \in \{1, \ldots, n\} \text{ iid}$$

$$q_k(\boldsymbol{\theta}) = \ell_k(\overline{\boldsymbol{\theta}}) + \nabla_{\boldsymbol{\theta}}\ell_k(\overline{\boldsymbol{\theta}})^\top (\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})^\top (\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}^\top}\ell_k(\overline{\boldsymbol{\theta}}))(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})$$

(quadratic/unimodality assumption)

$$\widehat{\sigma}^2_m(\theta) = \frac{n^2}{m^2}\sum_{i=1}^{m}\left(d_{u_i}(\theta) - \overline{d}_u(\theta)\right)^2, \quad \text{with } d_{u_i}(\theta) = \ell_{u_i}(\theta) - q_{u_i}(\theta)$$

$$\nabla_\theta \widehat{\ell}_m(\theta) = A(\theta^\star) + B(\theta^\star)(\theta - \theta^\star) + \frac{n}{m}\sum_{i=1}^{m}\left(\nabla_\theta \ell_{u_i}(\theta) - \nabla_\theta q_{u_i}(\theta)\right),$$

$$A(\theta^\star) \coloneqq \sum_{k=1}^{n} \nabla_\theta \ell_k(\theta^\star) \in \mathbb{R}^d \quad \text{and } B(\theta^\star) \coloneqq \sum_{k=1}^{n} H_k(\theta^\star) \in \mathbb{R}^{d\times d}$$

(and similarly for variance estimator gradient)

Dang et al, Hamiltonian Monte Carlo with Energy Conserving Subsampling, 2019
Gunawan et al, Subsampling Sequential Monte Carlo for Static Bayesian Models, 2020
Tran et al, The Block Pseudo-Marginal Sampler, 2017

Energy conserving subsampling

* **Loglikelihood approximations**:
Exact: -2.3136094233276503
Linear approximation: -2.2970251653956764
Quadratic approximation: -2.3156459493569908
*** Log-gradient approximation**:
Exact: [3.72314258]
Approximation: [3.5916315]
[test_approximation]

* **Loglikelihood estimator**:
- *Using control variates:*
  Estimator: -6.449208155054759
  Variance: 2.163409155741166e-17
- *Without control variates:*
  Estimator: -3.5046659638556283
  Variance: 2.355537773910069
- *Exact:* -6.449593115362749

*** Likelihood estimator:**
- Estimator (control variates): 0.0015817741928654402
- Estimator (no control variates): 0.009256448355579287
- Exact: 0.0015811653897750168

*** Ratio of tempered likelihoods:**
- Estimator (control variates): 0.05599973066625483
- Estimator (no control variates): 0.09219095195518667
- Exact: 0.055990096510776965

[test_estimators]

*** Loglikelihood gradients…**
- *Using control variates:*
  Gradient estimator: [-138.8016446]
  Calculated estimator gradient: [-138.80175163]
  Exact estimator gradient: [-138.80175163]
  Gradient approximation: [-138.56621602]
- *Without control variates:*
  Gradient estimator: [-219.4163191]
  Calculated estimator gradient: [32.05689257]
  Exact estimator gradient: [32.05689257]
- *Exact gradient*: [-146.78710099]

*** Variance gradients…**
- *Using control variates:*
  Exact variance estimator gradient: [-0.00021405]
  Variance gradient estimator: [-0.00021405]
- *Without control variates:*
  Exact variance estimator gradient: [502.94642335]
  Variance gradient estimator: [502.94642335]

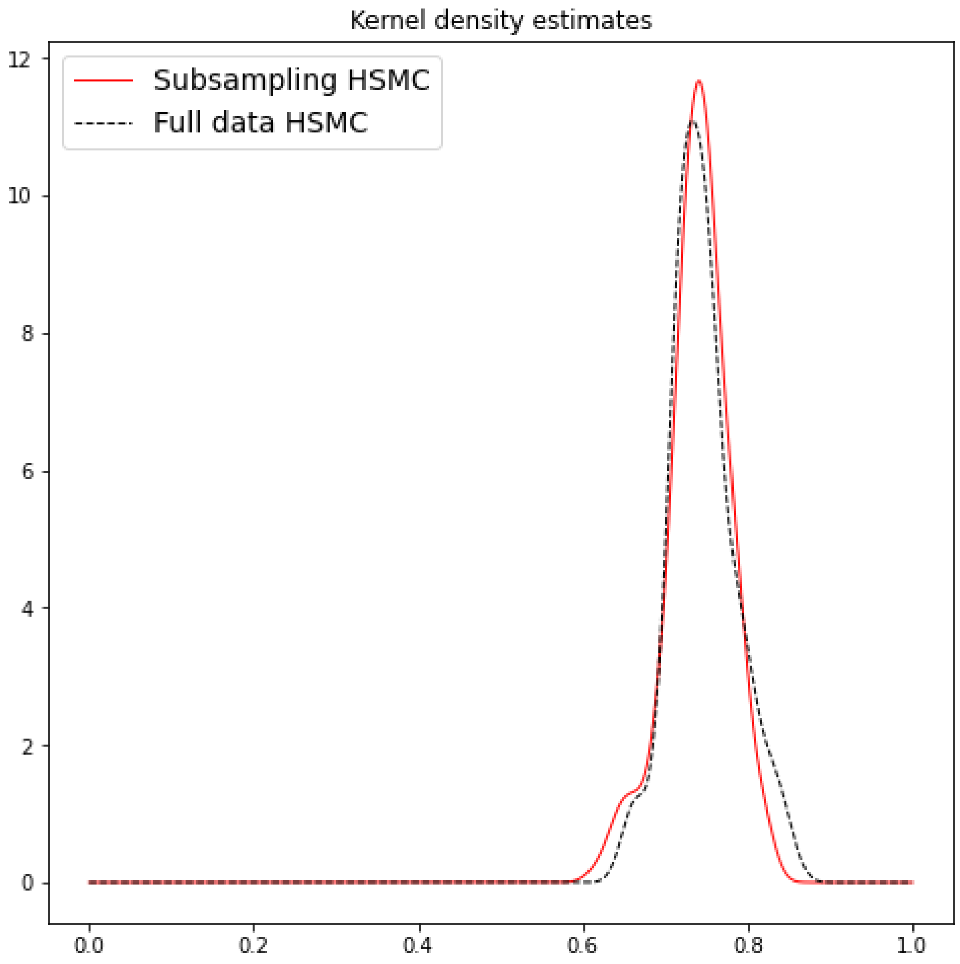
[test_gradient_estimator]

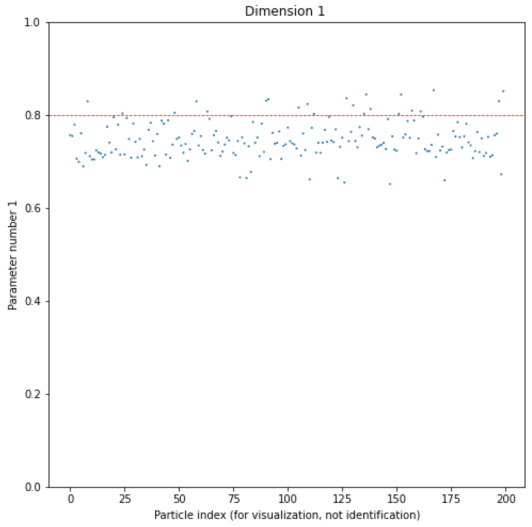
This doesn't work under same conditions as before (offline times with *tmax <=100* and flat prior on *]0,1[*) because the approximations are too off (too large distance between expansion center and points); the distribution tends to collapses into a single particle (the central one) or a few (for higher densities)

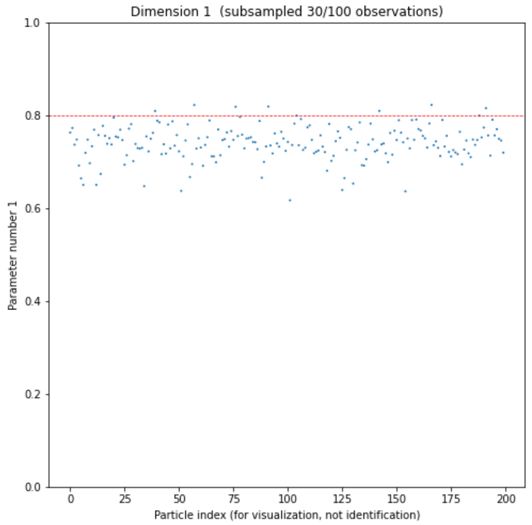
Energy conserving subsampling

It does work if:

1. *tmax* is set lower, e.g. 5:







Subsampled **30/100** observations

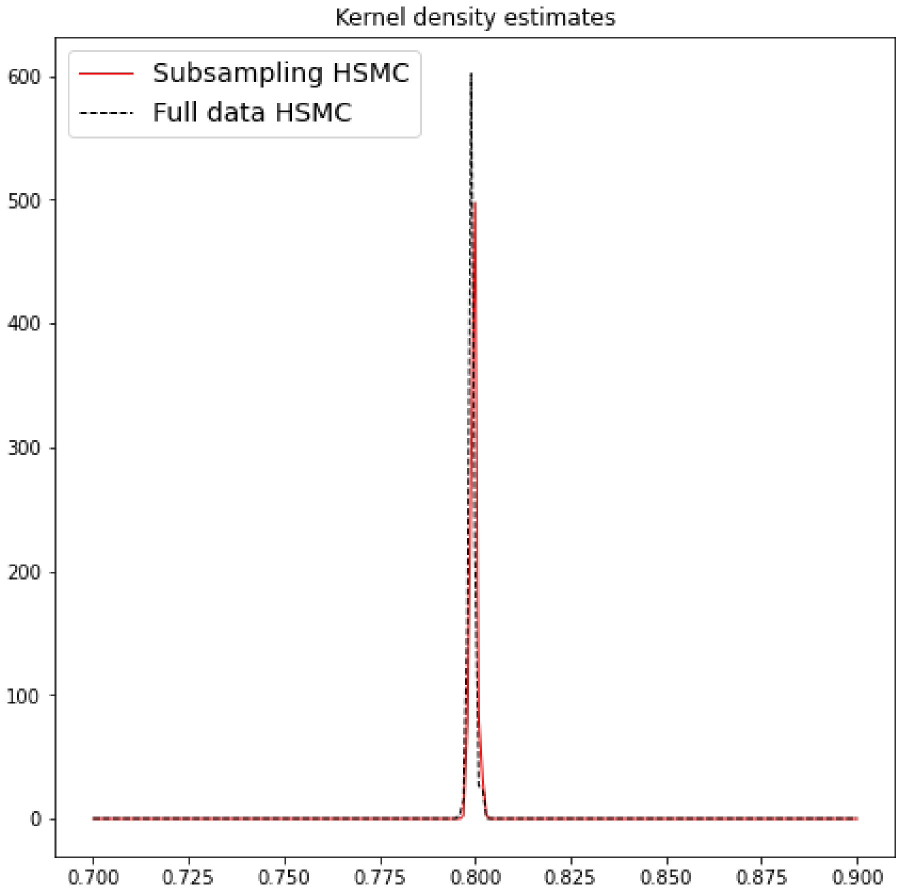
Final (corrected) standard deviations:

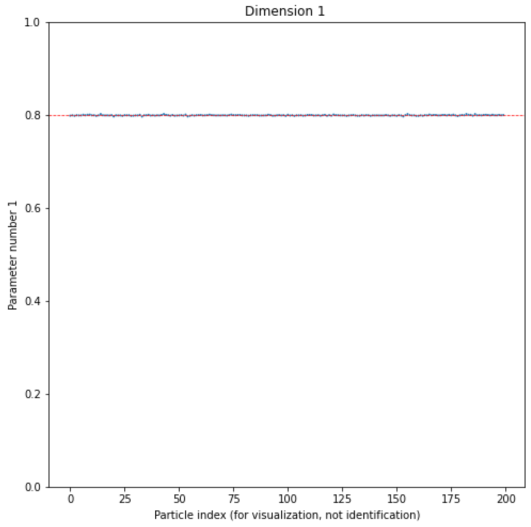
- Subsampling: 0.037
- Full data: 0.039

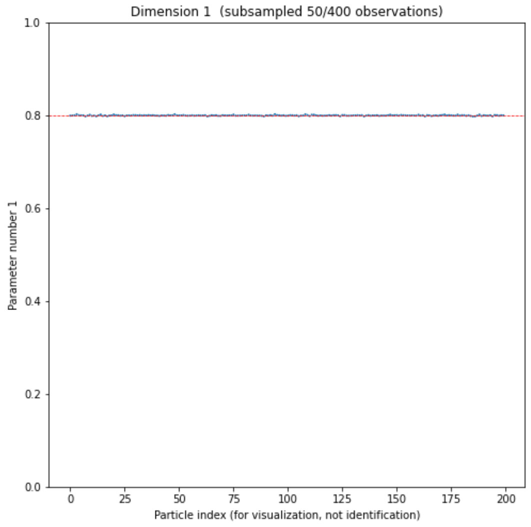
Energy conserving subsampling

Or:

2. The prior is narrower, e.g.]0.7,0.9[:







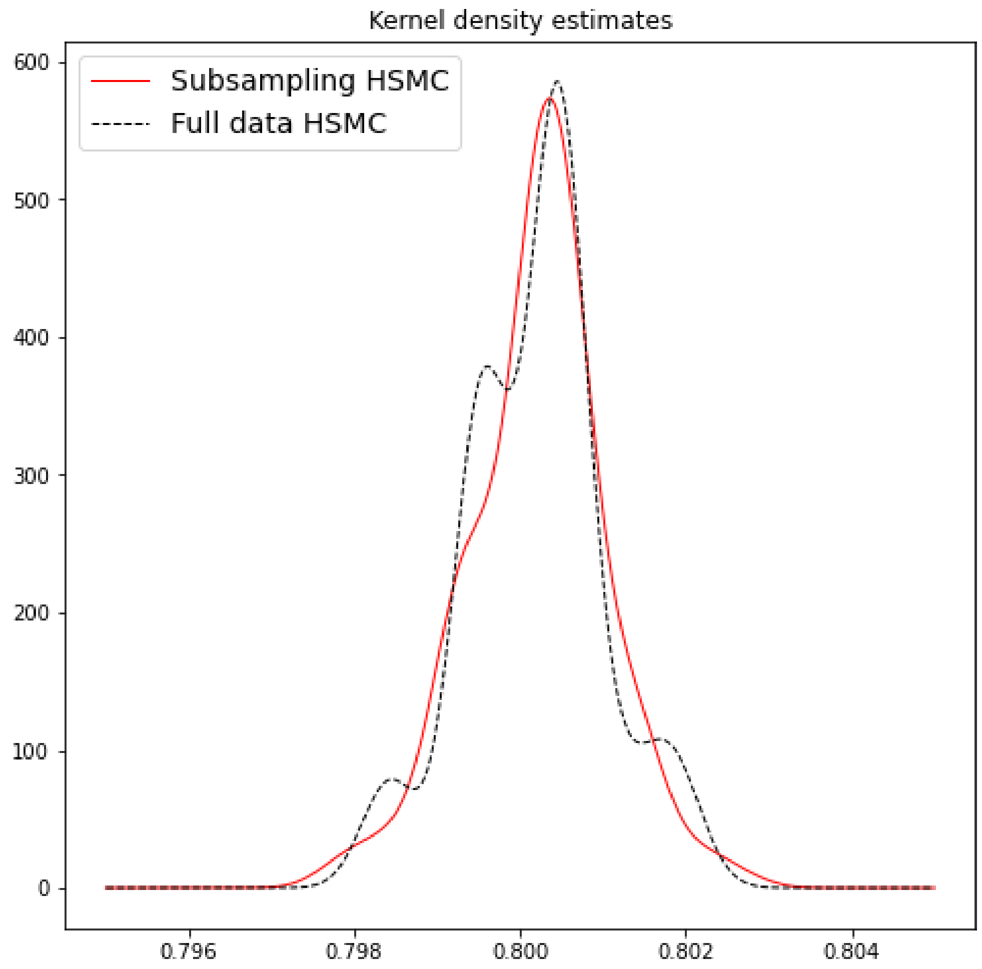
Subsampled **50/400** observations

Final (corrected) standard deviations:
- Subsampling: 0.00086
- Full data: 0.00082

(could also be a combination of **1.** and **2.**, e.g. a full data warm up, then add progressively longer times instead of tempering as in SIR)

Energy conserving subsampling

A different run, closer up:



Subsampled **50/400** observations

Final (corrected) standard deviations:

- Subsampling: 0.000819
- Full data: 0.000817