Hierarchical and Sequential Monte Carlo Fusion
A method of unifying distributed analyses

Ryan Chan [1,3], Murray Pollock [1,3], Gareth Roberts [1,3], Petros Dellaportas [2,3]

[1] University of Warwick, [2] University College London, [3] The Alan Turing Institute

Monte Carlo Fusion

Provides theory to carry out perfect inference for the target:

$$\pi(\boldsymbol{x}) \propto f_1(\boldsymbol{x}) \cdots f_C(\boldsymbol{x}) = \prod_{c=1}^{C} f_c(\boldsymbol{x}) \tag{1}$$

Uses rejection sampling on the extended target distribution:

$$g(\boldsymbol{x}^{(1:C)}, \boldsymbol{y}) \propto \prod_{c=1}^{C} \left[f_c^2(\boldsymbol{x}^{(c)}) p_c(\boldsymbol{y} \mid \boldsymbol{x}^{(c)}) \cdot \frac{1}{f_c(\boldsymbol{y})} \right] \tag{2}$$

Let $p_c(\boldsymbol{y} \mid \boldsymbol{x}^{(c)})$ be the transition density of a stochastic process with stationary distribution $f_c^2(\boldsymbol{x})$, then (**??**) admits π as a marginal for $\boldsymbol{y}$.

Proposal distribution:

$$h(\boldsymbol{x}^{(1:C)}, \boldsymbol{y}) \propto \prod_{c=1}^{C} \left[f_c\left(\boldsymbol{x}^{(c)}\right) \right] \cdot \exp\left(-\frac{C \cdot \|\boldsymbol{y} - \bar{\boldsymbol{x}}\|^2}{2T} \right) \tag{3}$$

Dai et al. (2019) showed that under certain mild conditions,

$$\frac{g(\boldsymbol{x}^{(1:C)}, \boldsymbol{y})}{h(\boldsymbol{x}^{(1:C)}, \boldsymbol{y})} \propto \rho \cdot Q \tag{4}$$

Algorithm 1 Monte Carlo Fusion (Dai et al. 2019)

1. Initialise a value for $T > 0$
2. Simulate a proposal $\boldsymbol{y}$ from h:
   a) For $c = 1, \ldots, C$, simulate $\boldsymbol{x}_c \sim f_c(\boldsymbol{x})$ and calculate $\bar{\boldsymbol{x}}$
   b) Simulate $\boldsymbol{y} \sim \mathcal{N}_d(\bar{\boldsymbol{x}}, \frac{T \mathbb{I}_d}{C})$
3. Accept $\boldsymbol{y}$ as a sample from (**??**) with probability $\rho \cdot Q$

Example 1. Target $\pi(x) \propto e^{-\frac{x^4}{2}}$ and $f_c(x) \propto e^{-\frac{x^4}{2C}}$ for $c = 1, \ldots, C$ and $C = 4$.

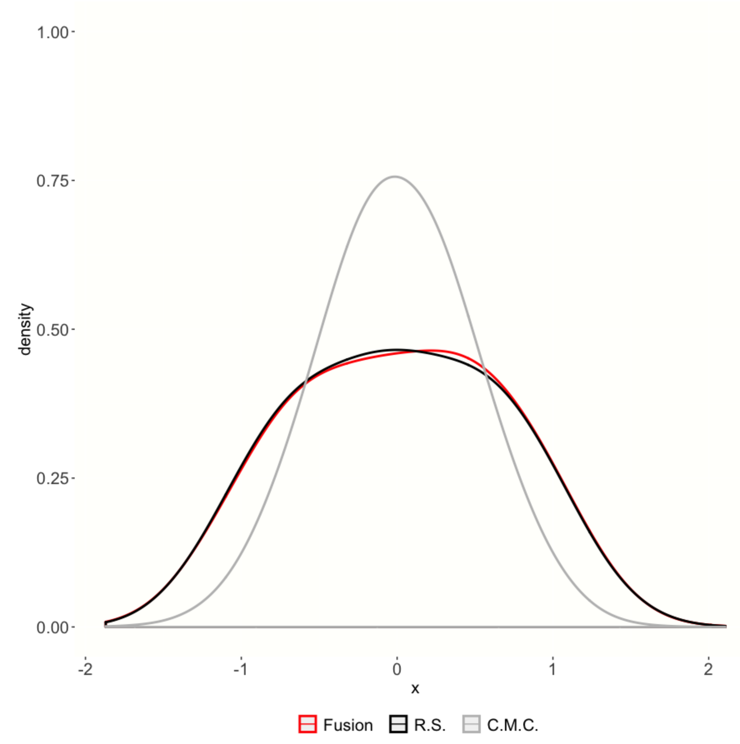

Figure 1: Kernel density fitting based on 10,000 realisations for density proportional to $e^{-\frac{x^4}{2}}$, based on different Monte Carlo methods - 1. black (standard rejection sampling), 2. red (Monte Carlo fusion), 3. grey (basic Consensus Monte Carlo (Scott et al. 2016))

Problem: Fusion becomes inefficient as the number of sub-posteriors increases.

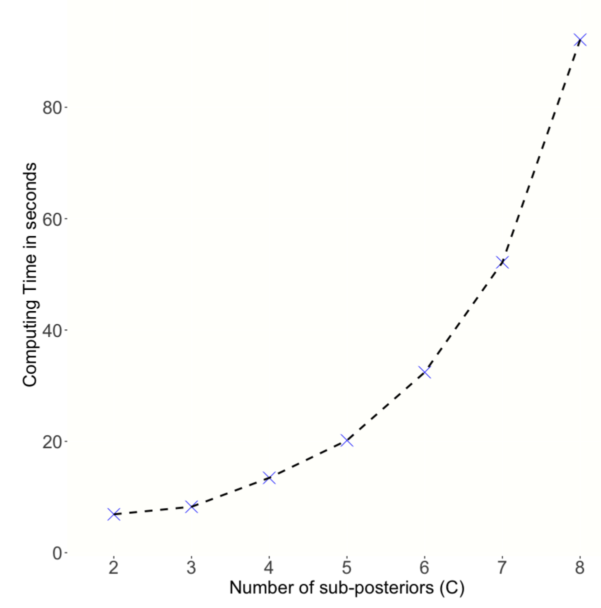

Figure 2: Run-times of standard MC fusion for Example 1 for varying C

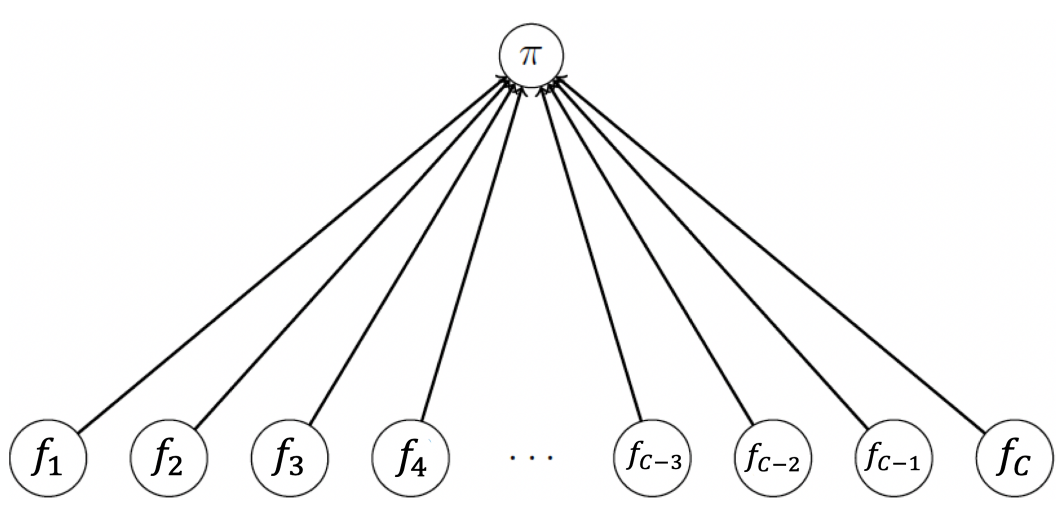

Figure 3: The 'fork-and-join' approach

Reference

Dai, H. M. Pollock, and G. Roberts. 2019. Monte Carlo Fusion. To be published in *Journal of Applied Probability*. [Preprint]. Available at: https://arxiv.org/abs/1901.00139

Acknowledgements: RC is funded by The Alan Turing Institute Doctoral Studentship, under the EPSRC grant EP/N510129/1.

Hierarchical and Sequential Monte Carlo Fusion

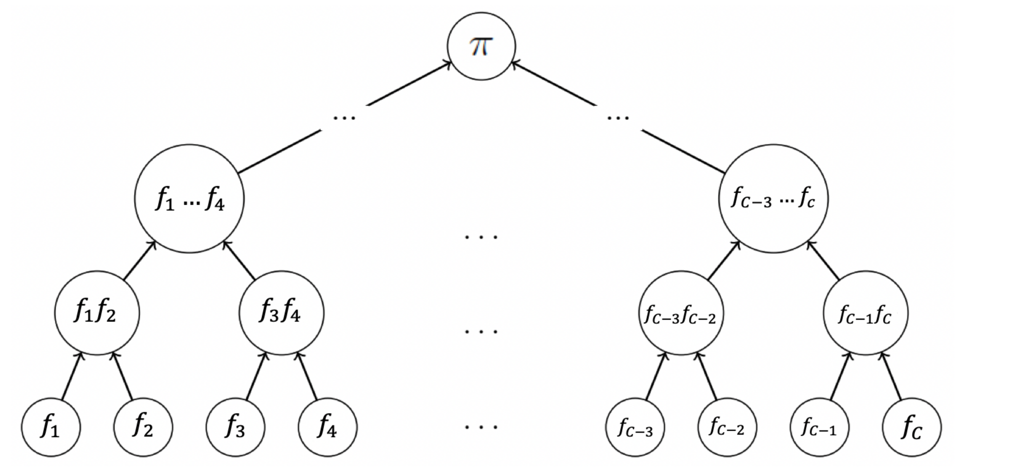

Figure 4: The hierarchical approach

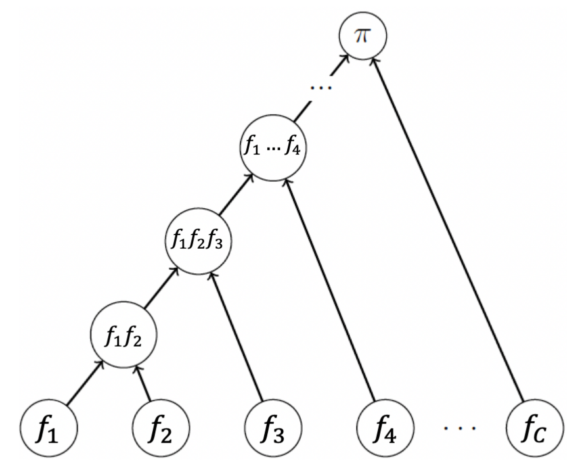

Figure 5: The sequential approach

Tempering for Monte Carlo Fusion

The *power-tempered target distribution*: $\pi_\beta(\boldsymbol{x}) \propto [\pi(\boldsymbol{x})]^\beta$, for $\beta \in (0, 1]$

$$\pi(\boldsymbol{x}) = \pi(\boldsymbol{x})^{\frac{1}{\beta} \cdot \beta} = \prod_{i=1}^{\frac{1}{\beta}} \pi_\beta(\boldsymbol{x}) \quad \text{if } \frac{1}{\beta} \in \mathbb{Z} \tag{5}$$

Useful to consider when:

- sampling from multi-modal densities (let $f_c = \pi_\beta(\boldsymbol{x})$ for $c = 1, \ldots, \frac{1}{\beta}$)
- sub-posteriors conflict / have little overlapping support

Example 2. Target $\pi(x) \propto f_1 f_2$, where $f_1 \sim \mathcal{N}(-8, 1)$ and $f_2 \sim \mathcal{N}(2, 0.5)$.
Problem: Lack of overlapping support between f_1 and f_2 - fusion is inefficient.
Solution: Perform fusion on f_1^β and f_2^β to obtain samples for π^β and then use hierarchical or sequential Monte Carlo fusion to get samples for π.

1. Use Monte Carlo fusion to obtain samples for $\pi^\beta \propto f_1^\beta f_2^\beta$ with $\beta = \frac{1}{8}$:

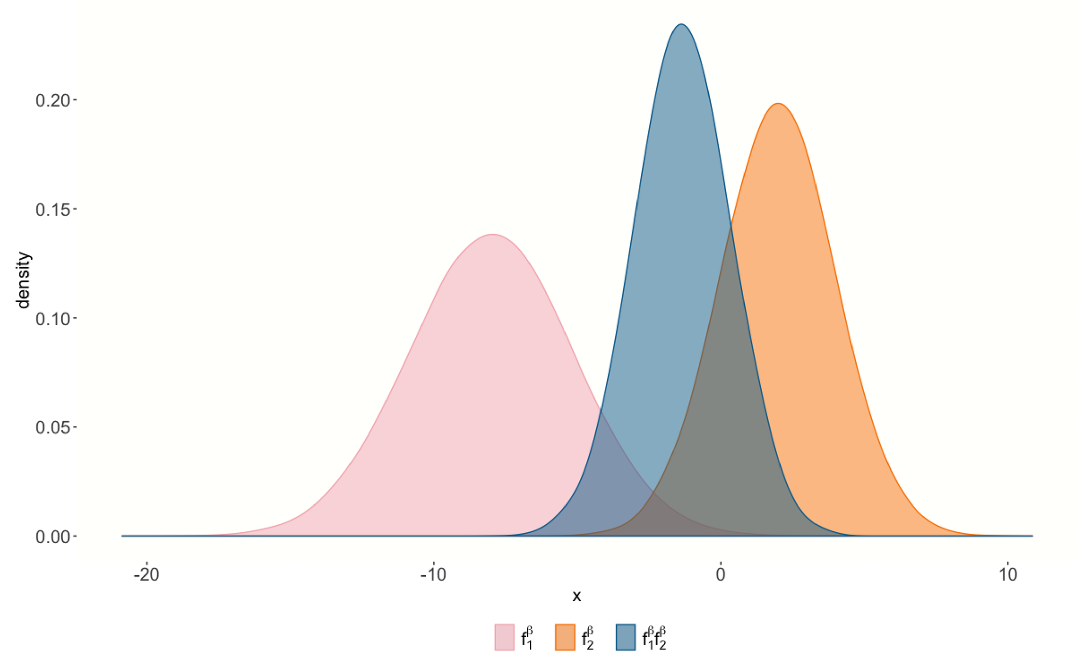

Figure 6: Kernel density fitting based on $10,000$ realisations for density proportional to $f_1^\beta f_2^\beta$ (blue) and the true density curves for f_1^β (pink) and f_2^β (orange) in Example 2

2. Use hierarchical or sequential Monte Carlo fusion and samples for π^β to obtain samples for $\pi \propto f_1 f_2$

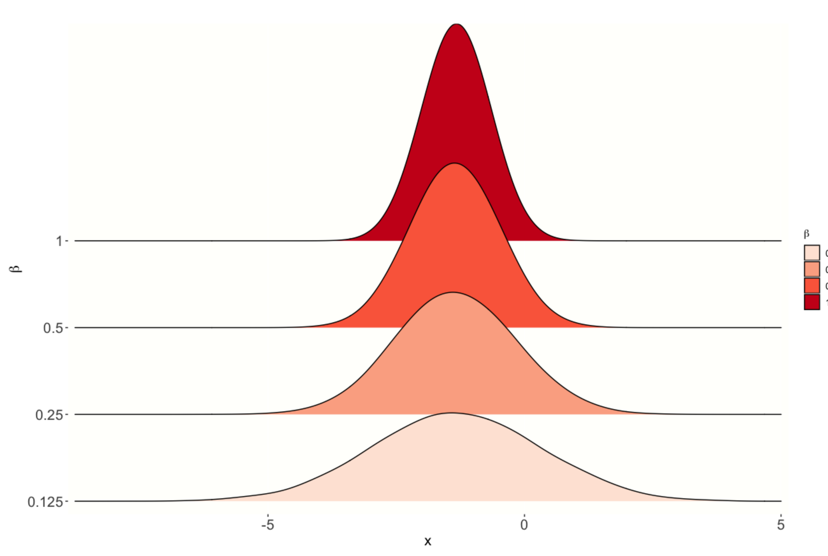

Figure 7: Kernel density fitting for π^β (hierarchical Monte Carlo fusion)

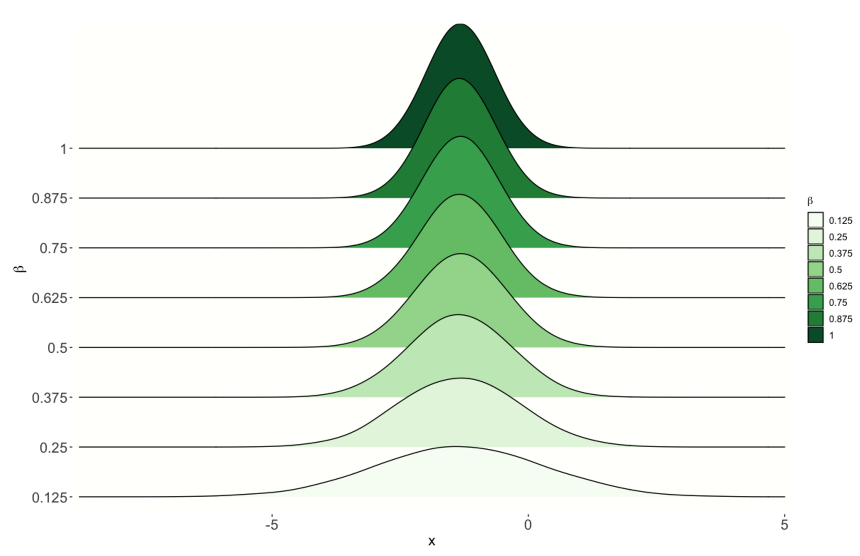

Figure 8: Kernel density fitting for π^β (sequential Monte Carlo fusion)

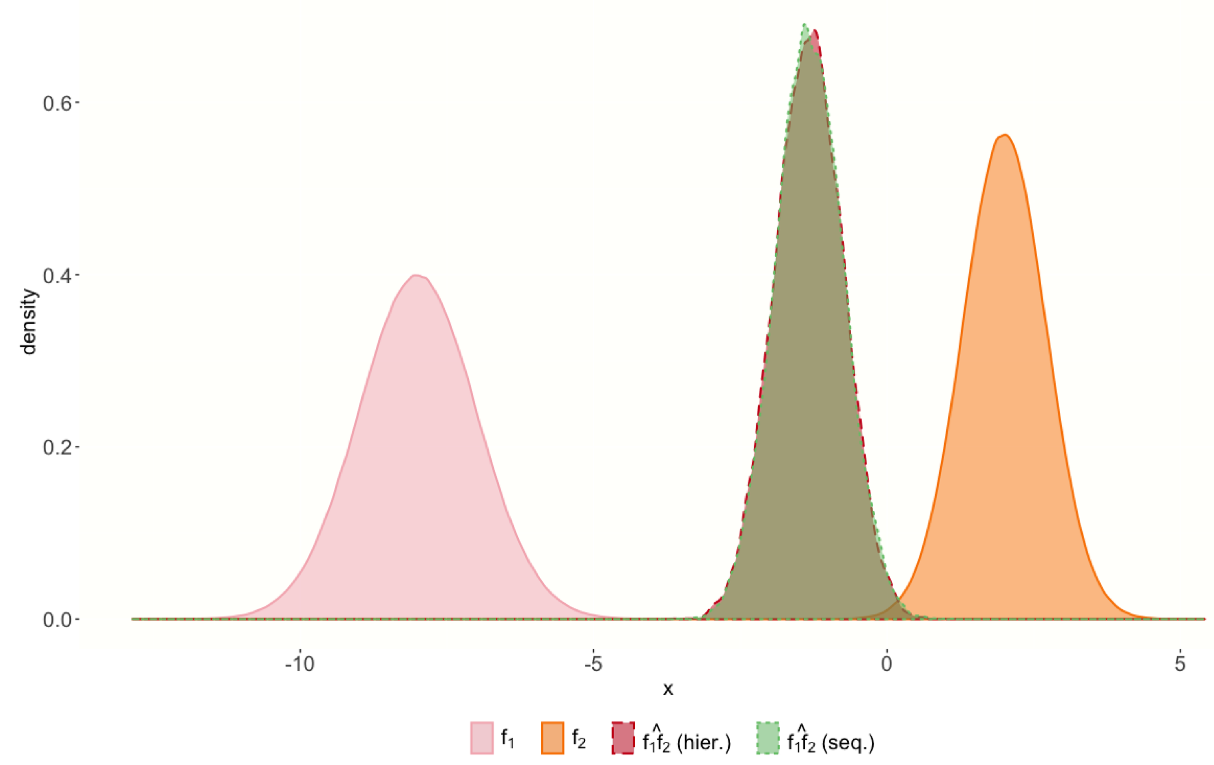

Figure 9: Kernel density fitting based on $10,000$ realisations for density proportional to π based on hierarchical fusion (red dashed), sequential fusion (green dotted), and the true density curves for f_1 (pink) and f_2 (orange) in Example 2