Density function estimation using ergodic recursion

Erik Bodin[1], Zhenwen Dai[2], Neill D. F. Campbell[3], Carl Henrik Ek[4]

[1]University of Bristol [2]Spotify Research [3]University of Bath [4]University of Cambridge

Summary

- Physically motivated likelihood functions and priors often give rise to complicated or degenerate inference problems, often requiring weeks of computation and a portfolio of methods even in relatively low dimension.
- We propose a **fast**, **robust** and **simple-to-use** algorithm, for general Bayesian computation and inference, applicable to a wide set of problems.
  - Robust to degeneracies in the density function, such as discontinuities, nonsmoothness, and zero density regions.
  - Does not require tuning.
  - Does not rely on gradients.
  - Asymptotically **exact**.
- Enables down-stream tasks, such as
  - Bayesian parameter inference,
  - evidence estimation,
  - expectations of functions,
  - fast, constant-time sampling,
  - conditional and marginal distribution estimation,
  - estimation of divergences and mutual information.

Methodology

$$\mathcal{P}(\boldsymbol{\theta}) = \frac{f(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}\in\Omega} f(\boldsymbol{\theta})d\boldsymbol{\theta}} = \frac{f(\boldsymbol{\theta})}{Z},$$

where $f : \Omega \rightarrow \mathbb{R}_+$ is an explicitly unknown density function defined on a bounded sample space $\Omega \subset \mathbb{R}^D$, which can be evaluated for any $\boldsymbol{\theta} \in \Omega$. Z is the unknown normalizing constant.

Representation of $\hat{f}$

- Search-tree over partitions of the sample space,
- constituting a Riemann sum.

Iterative construction of $\hat{f}$

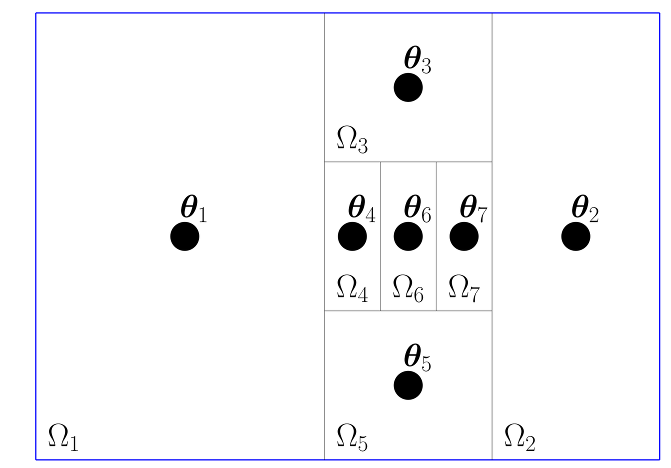
- Addresses sample-efficiency by the prioritization-order of refining partitions.
- Uses an exploration-exploitation trade-off which considers all density observations seen so far.
- The time complexity of acquiring a new partition at step t is $\mathcal{O}(\log N_t + U_t \log U_t)$.

  In practice, decision-times are fractions of a millisecond even after millions of partitions.

Input: Density function f defined over Ω with unknown normalization constant Z

Output: Approximation $\hat{f}$, $\hat{Z}$, as specified by the produced partitioning Π_T

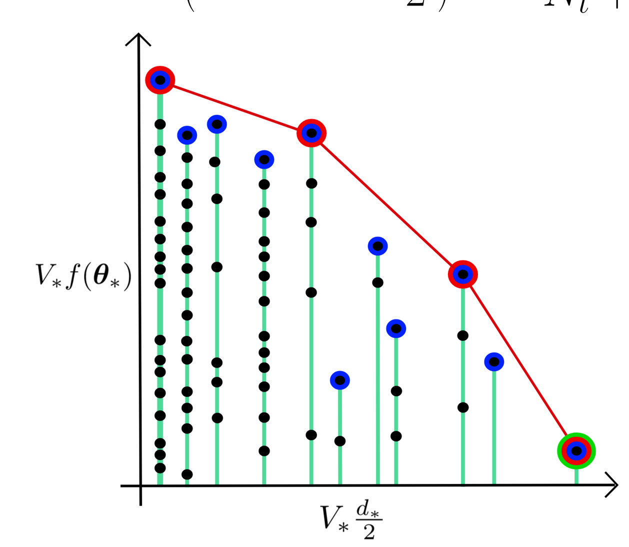
set t = 1 and initial partitioning $\Pi_1 = \{\Omega\}$;
while $N_t \leq N_{max}$ **do**
  $\{\Omega_i\}$ = to_divide $[\Pi_t]$;
  divide each partition Ω_i, each one resulting in $\{\Omega_j\}$ new partitions;
  add all sets of $\{\Omega_j\}$ into Π_t remove the divided partitions $\{\Omega_i\}$ from Π_t;
  set $t \rightarrow t+1$ and update data structures;
end
return Π_T;



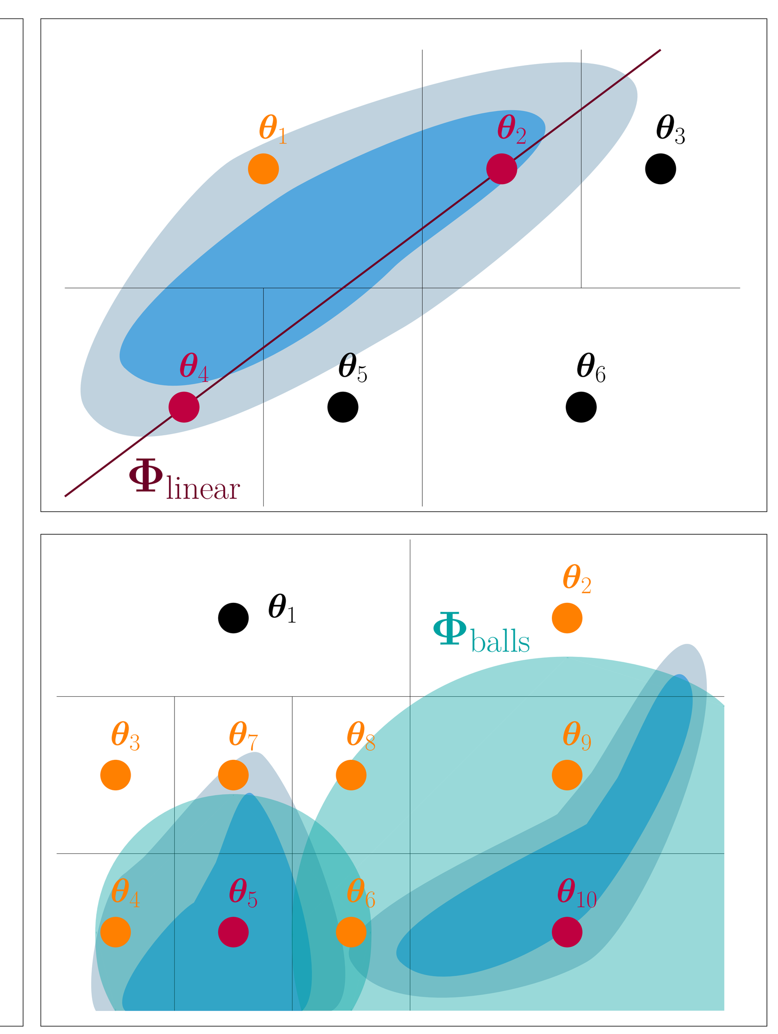
Partition division criteria

$$V_k \cdot \left(f(\boldsymbol{\theta}_k) + \bar{K}\frac{d_k}{2}\right) \geq V_i \cdot \left(f(\boldsymbol{\theta}_i) + \bar{K}\frac{d_i}{2}\right), \quad \forall\, i \in [1, N_t]$$

and $$V_k \cdot \left(f(\boldsymbol{\theta}_k) + \bar{K}\frac{d_k}{2}\right) \geq \beta \frac{\hat{Z}}{N_t + 1}.$$



$$\bar{K}_i^{\text{upper}} = 2\frac{V_{i+1}f(\boldsymbol{\theta}_{i+1}) - V_i f(\boldsymbol{\theta}_i)}{V_i d_i - V_{i+1}d_{i+1}}$$



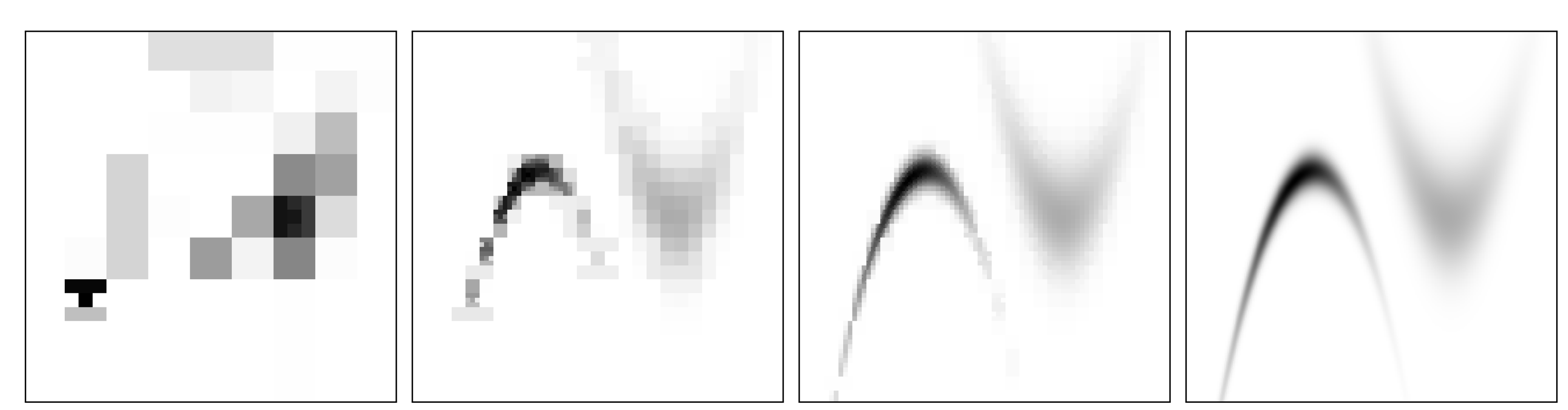


Figure: Partitions after 53, 303, 2 101, and 100 003 density evaluations, respectively.

Experiments

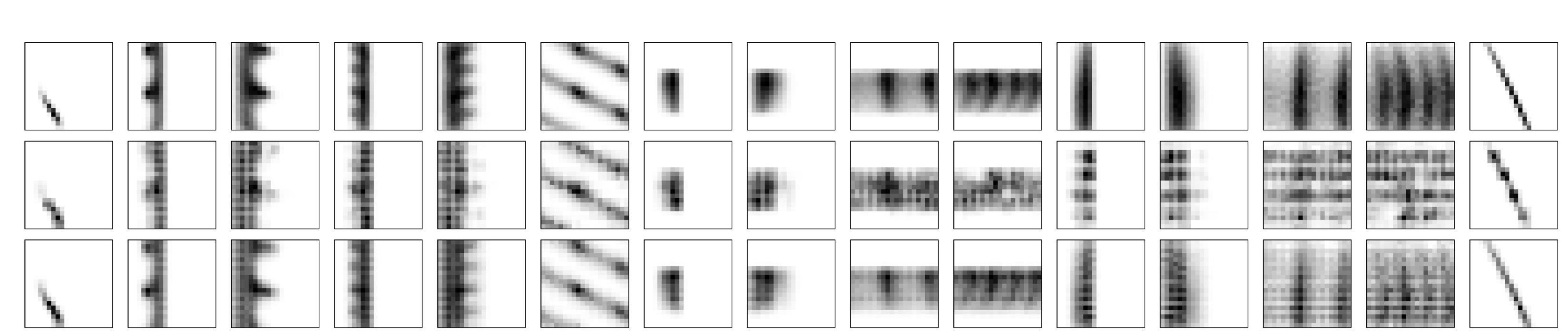


Figure: Parameter estimation for gravitational-wave physics.

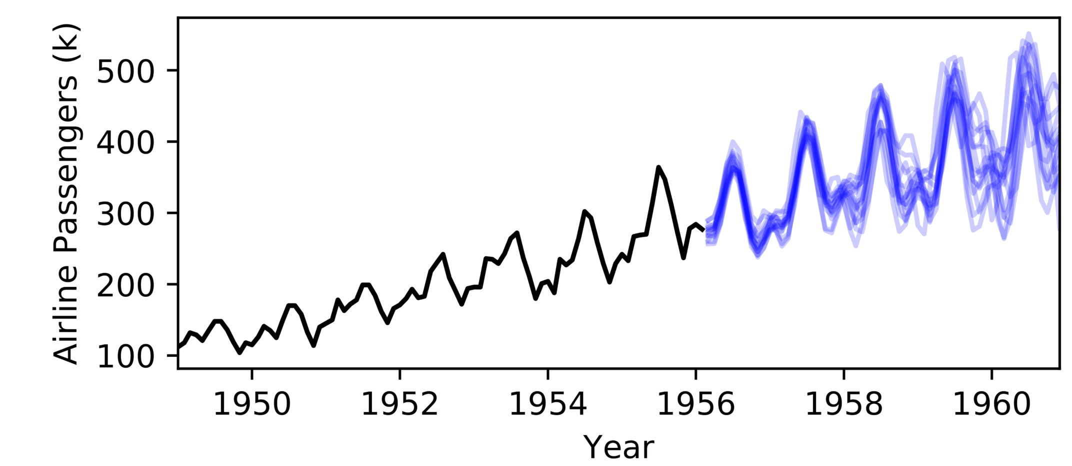
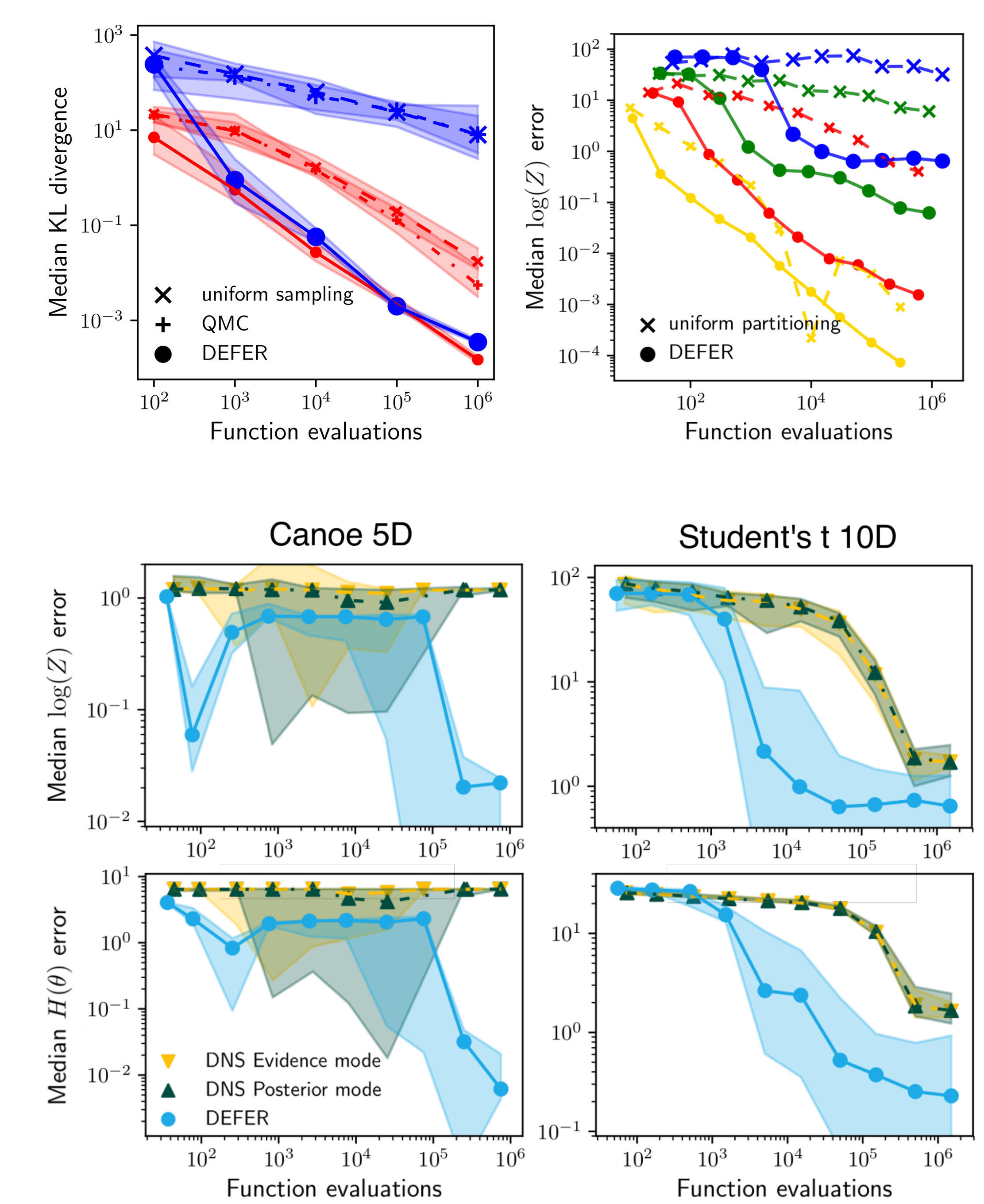


Figure: GP regression using the Spectral Mixture Kernel.



Contact Information

www.linkedin.com/in/stigerikbodin
mail@erikbodin.com