Problematic confidence: The Role of Uncertainty when Data is estimated from Data

DALI workshop on Hidden Gems, San Sebastian 03.09.2019

Aasa Feragen

Section for Image Analysis & Computer Graphics
Technical University of Denmark

afhar@dtu.dk

Danmarks Tekniske Universitet

The questions asked in this talk are inspired by Anton Mallasto's upcoming PhD thesis

Ingrid (3 months) also helped!

Case: Estimating structural brain connectivity

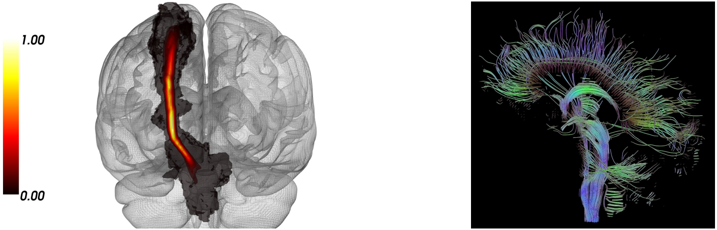


Figure: **Left:** Voxel-wise heatmap, NeuroImage'16. **Right:** Estimated white-matter trajectories (figure by Thomas Schultz).

- Estimate white-matter bundle trajectories from local tangential directions
- Data is noisy and has low resolution; tracked axons so small, we only observe population level behavior
- $\rightsquigarrow$ Our estimates are wrong! How wrong?

Case: Estimating structural brain connectivity

Population analysis routinely assumes that our best estimate is correct. Example:

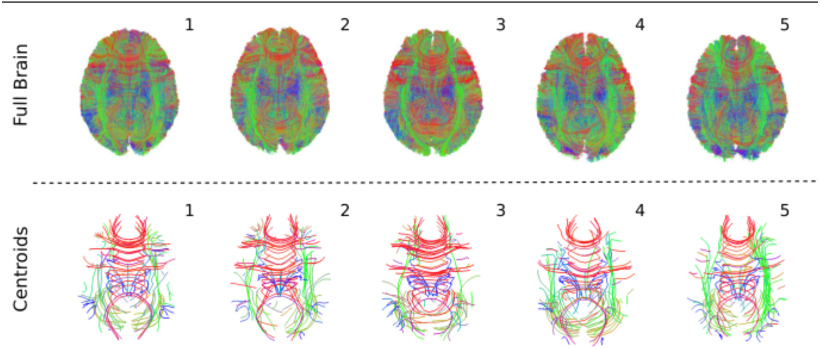
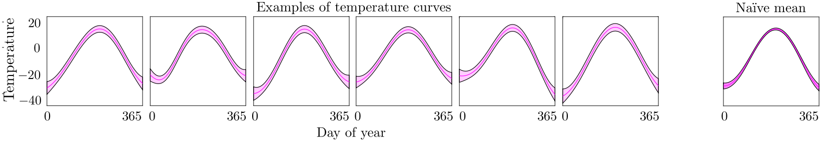


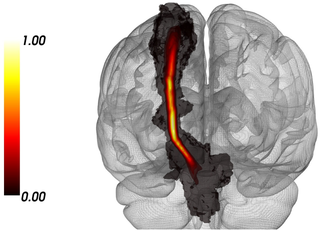
Figure: From Garyfalldis et al, Front Neurosci'12

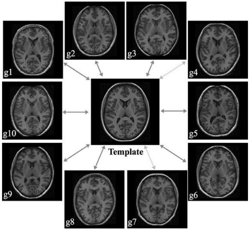
Ignoring uncertainty in population analysis is a problem

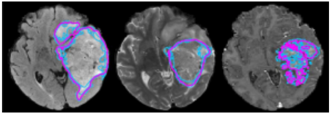


- **Left:** Small sample of uncertain yearly temperature curves represented as Gaussian Processes (GPs) from a Siberian metereological station.
- **Bottom right:** The mean and pointwise standard deviation of the mean temperature curves (the best estimates).
  This is what we routinely do in imaging and many other domains.

Ignoring uncertainty in population analysis is a **common** problem



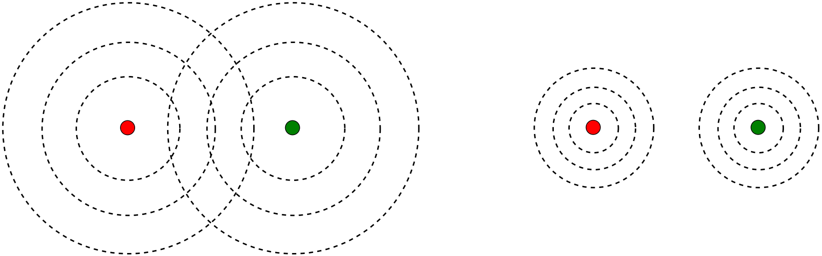




- In most population analysis in medical imaging, the "data" is estimated from data
- Algorithms that quantify uncertainty are starting to appear
- Incorporating uncertainty in population analysis is not currently tackled

Ignoring uncertainty in population is a potentially **serious** problem

Underestimating uncertainty can lead to incorrect "significant" differences

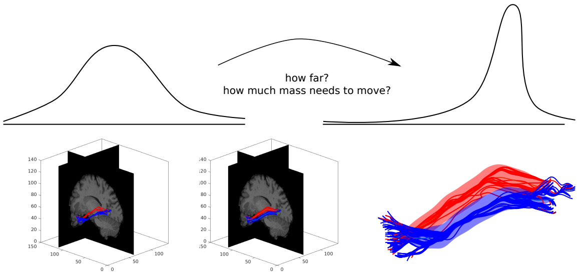


Conclusion: If your final output includes an uncertainty, then it needs to be propagated from the uncertainty of your "data"

Hidden gem? First steps.

Goal:

- Point out the problem: Learning from uncertain curves in the form of Gaussian Processes (GPs) (important: links to FDA)
- First solution: Distance based learning based on Wasserstein geometry of GPs



Mallasto and **F**, *Learning from uncertain curves*, NeurIPS'17

Main results

Geometric:

- For two GPs f and g, their Wasserstein distance is the limit of the Wasserstein distance between their finite-dimensional GD approximations.
- The Wasserstein geodesic between f and g can be approximated arbitrarily well by the Wasserstein geodesic in the space of finite-dimensional GD approximations.
- The Wasserstein *barycenter* (weighted mean) of a set of GPs is the limit of the Wasserstein barycenter of their finite-dimensional GD approximations.

Main results

Geometric:

- For two GPs f and g, their Wasserstein distance is the limit of the Wasserstein distance between their finite-dimensional GD approximations.
- The Wasserstein geodesic between f and g can be approximated arbitrarily well by the Wasserstein geodesic in the space of finite-dimensional GD approximations.
- The Wasserstein *barycenter* (weighted mean) of a set of GPs is the limit of the Wasserstein barycenter of their finite-dimensional GD approximations.

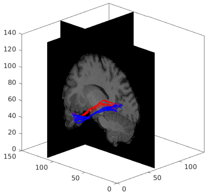
Computational:

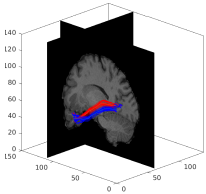
- Distances and geodesics for GDs have analytical formulas (Takatsu'11)
- Barycenters for GDs can be computed via iterative optimization schemes (essentially SGD)

Main results

Practical:

- This allows us to use distance-based learning:
  - mean GPs (cluster means)
  - hierarchical clustering
  - permutation tests for equal means





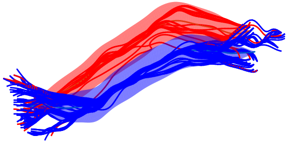


Figure: Clustering, hypothesis testing between populations of GPs

Main results

- With weighted means for GPs, we can perform kernel regression (GP-valued output).

Figure: Kernel regression on GPs: Predicted temperature curve distributions over 30 Russian weather stations in the period 1940-2009

But those are only first steps...

Limitations:

- Only enables distance-based learning; a bit limited
- Choice of distance is always ad hoc

But those are only first steps...

Limitations:

- Only enables distance-based learning; a bit limited
- Choice of distance is always ad hoc

Other options?

- Bayesian – forced to either make strict model assumptions, or learn how to sample from nontrivial distributions.
- What about "embedding-based" learning such as kernel methods or neural networks?

But those are only first steps...

Limitations:

- Only enables distance-based learning; a bit limited
- Choice of distance is always ad hoc

Other options?

- Bayesian – forced to either make strict model assumptions, or learn how to sample from nontrivial distributions.
- What about "embedding-based" learning such as kernel methods or neural networks?

Challenge: Complex data

- Nonlinear constraints and invariances – data often lives on manifold or similar
- Nonlinear processing steps, and variance needs to enter with the raw data

But those are only first steps...

Limitations:

- Only enables distance-based learning; a bit limited
- Choice of distance is always ad hoc

Other options?

- Bayesian – forced to either make strict model assumptions, or learn how to sample from nontrivial distributions.
- What about "embedding-based" learning such as kernel methods or neural networks?

Challenge: Complex data

- Nonlinear constraints and invariances – data often lives on manifold or similar
- Nonlinear processing steps, and variance needs to enter with the raw data

Leaving the question still: **How do we properly handle uncertainty when our data is estimated from data?**