

1 Introduction

Consider a closed curve:

$$C : \mathbb{S} \rightarrow \mathbb{R}^2$$

Shapes should be invariant across several transformations including translation, scaling, rotation, and reparameterizaion. Specifically, the shape C is defined to be the equivalence class

$$C = \{\sigma OC(\gamma(t)) + a, \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \sigma > 0\}$$

Consider the curve $C_j^m : \mathbb{S} \rightarrow \mathbb{R}^2, j = 1, \dots, n_m$ which is only observe on some region (don't want to use \mathbb{R} here...) $S_j \subset \mathbb{S}$. The goal here is to impute the missing part of each shape. Then perform some type of complete shape analysis.

2 Notation

$$C : \mathbb{S} \rightarrow \mathbb{R}^2$$

In general, we have a missingness function like this. This allows for us to be missing the x or y value individually at a given point along the curve. We aren't really going to deal with that case here so we are dealing with something a little more specific.

$$R : \mathbb{S} \rightarrow \{0, 1\}^2$$

We are really dealing with this:

$$R : \mathbb{S} \rightarrow \{\{0, 0\}, \{1, 1\}\}$$

This means that at any given location that we are missing a value, both the x and y values are missing.

Here R is a missingness indicator function. It is 1 when the value of C is unobserved and 0 otherwise.

The set of fully observed shapes is the set \mathcal{C}^o and contains n_o shapes; the set of partially observed shapes is the set \mathcal{C}^m and contains n_m shapes. Together $(\mathcal{C}^o, \mathcal{C}^m) = \mathcal{C}$

Next we want to consider the joint distribution of the curves with the missingness indicator.

$$P(C, R) = P(C_{obs}, C_{mis}, R)$$

Bayesian approach: Construct a model for

$$P(C_{obs}, C_{mis}, R | \theta, \phi)$$

where θ is a vector of parameters associated with modeling the data and ϕ are parameters associated with the model of the missing data mechanism.

We then seek this distribution:

$$\int \int P(C_{mis}|\theta, \phi)P(\theta, \phi|C_{obs}, C_{mis}, R)d\theta d\phi = P(C_{mis}|C_{obs}, R)$$

Random draws from the distribution $P(C_{mis}|C_{obs}, R)$ can then be used to fill in the missing part of the shape.

Each partially observed shape can be completed M times. This yields M sets of completed shapes.

3 Combining

3.1 Shape Analysis

Let's say we wanted to ask a statistical question where the answer was a shape (e.g What is the average shape?). I propose to answer this question, you first find the karcher mean within each of the M completed data sets and then you find the karcher mean of the M means across the completed data sets. This is analagous to Rubin's combining rules.

It would be nice to try to come up with something like a "95 percent shape shadow". Like a band around the tooth showing the uncertainty around the mean shape.

3.2 Traditional Analysis

There are other types of analysis, like classification, where the results might be probabilities or other traditional statistical quantities. I believe that these can be combined across imputatations using Rubin's classic combining rules.

4 Approximating $P(C_{mis}|C_{obs}, R)$

The distribution $P(C_{mis}|C_{obs}, R)$ can be approximated non-parametrically by employing a hot deck type procedure similar to the idea of predictive mean matching in traditional missing data imputation.

We first perform a matching step. Consider a shape $C_i^m \in \mathcal{C}^m$ and a

What if we took all the complete shapes and aligned them

We can represent shapes using finite representations.

An issue here though is that these are really just one realization from an equivalence class.

MAR here means that the proability of missingness is a function of only the observes part of the shape. This seems unrealistic in our setting.