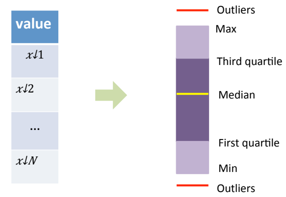
Statistical Graphics: Curve Boxplot

CPSC 547 Presentation
Ken Lau

Boxplot

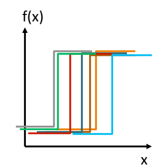


What: Data

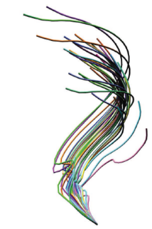
- 1D quantitative attribute is easy
- How to deal with 2D-3D curves?
  - Functional Data
  - Isocontours
  - Streamlines/Pathlines

Examples

Functional Data

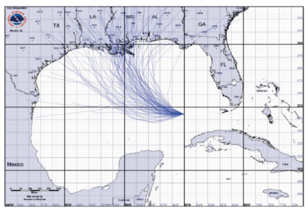


Helical Curves

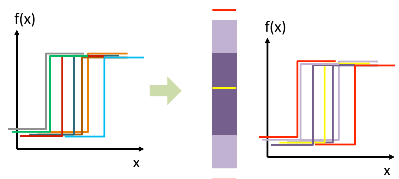


Examples

50 Simulated Hurricane Tracks



Big Picture



What: Derived

- Functional Band Depth
  - Measure of centrality for ensemble of functions
  - Sorting in higher dimensions

Computing the Band

- Given an ensemble of n functions:

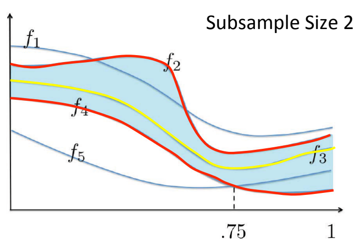
$$\{f_1(x), f_2(x), \cdots, f_n(x)\}$$

- Repeatedly subsample j of them and define the band as:

$$B(f_1(x), \cdots, f_j(x)) = \{(x,y) : x \in \mathscr{D}, y \in \mathscr{R}, \min_{k=1\cdots,j} f_k(x) \leq y \leq \max_{k=1,\cdots,j} f_k(x)\}.$$

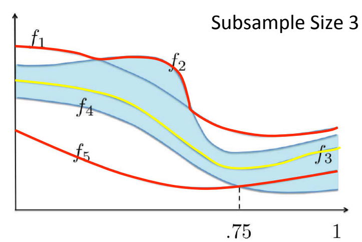
Computing the Band

Subsample Size 2



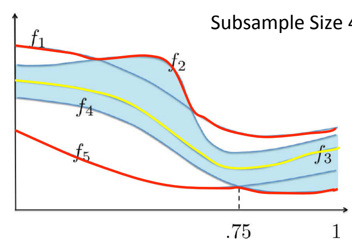
Computing the Band

Subsample Size 3



What: Derived

Subsample Size 4



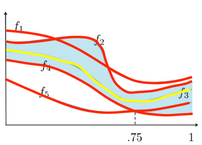
Band Depth

- Equation:

$$BD_j(g(x)) = \text{Prob}[g(x) \subset B(f_{i_1}(x), \cdots, f_{i_j}(x))], \quad 1 \leq i_1 \leq \cdots, \leq i_j \leq n$$

- Probability of a curve falling between the Band with subsample size j
- Compute BD for every function
- Highest band depth value is assigned the median

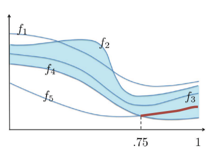
Subsample Size 2



Function	BD
f1	0
f2	0
f3	0.6
f4	0
f5	0

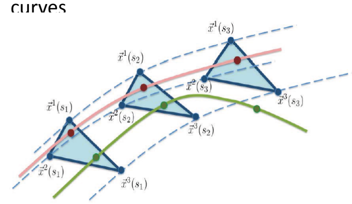
Modified Band Depth

- Includes portion of time a function lies inside the random band
- Provides more reliable results and prevents degeneracy



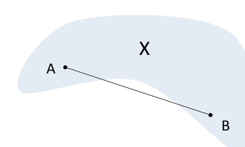
Generalization to multivariate Curves

- Convex hull formed by all subsampled curves



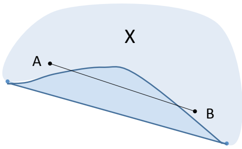
Convex Hull

Non-convex Set



Convex Hull

- The convex hull is the smallest convex set that contains all the points X
- The blue line completes the convex hull for X



How: Encode

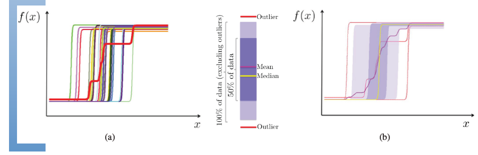
- Derive univariate boxplot using band depth values

How: Encode

- Colour encoding corresponding to univariate boxplot
- Constructive Solid Geometry
  - Union operator to represent the highest 50% band depth scores

Why: Tasks

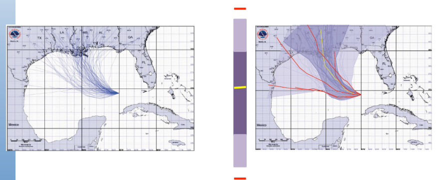
- Quantify information within ensemble members
- Detect hidden patterns
  - Not captured by point-wise mean
  - Ex. Red curve on left and Purple on right



Ensemble of Hurricane Tracks

Direct Ensemble Visualization

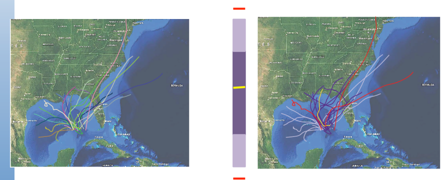
Curve Boxplot Visualization



Hurricane Tracks Gulf of Mexico

Direct Ensemble Visualization

Curve Boxplot Visualization



Qualitative Result Inquiry

- 10 Hurricane experts at NHC
- Users expressed appreciation/satisfaction in quantitative interpretation
- Outlier visualization useful for experts

Strengths

- Non-parametric method
  - No model assumptions
  - Method to visualize variability among complex ensemble members
- Minimizes visual clutter and focus on important information

Weaknesses

- Non-parametric method
  - Scalability issue:

$$(\#repititions)(\#ensembles)\binom{\#ensembles}{5}(\#points)^{\#dimensions}$$

- Unable to handle multimodal distribution
- Needs user study targeted at public

Critique and Other

- Provides both quantitative and qualitative information
  - Most central curve consistent with physical and simulation constraints
  - Aggregation by point-wise mean is less reliable
- Generalize to different data types?
  - Time series
  - Networks/Trees

Questions?