Geometric statistical methods with tumour images

Karthik Bharath

School of Mathematical Sciences
University of Nottingham

Acknowledgements

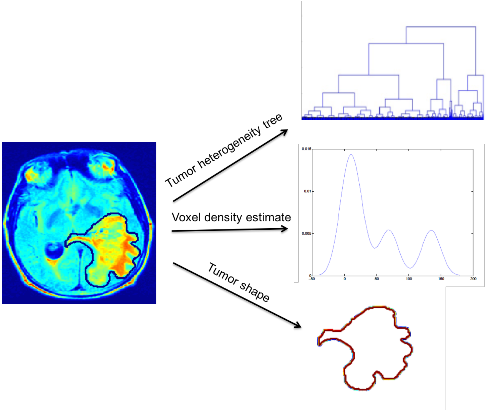
Joint work with:

- Sebastian Kurtek (Ohio State University);
- Dipak. K. Dey (University of Connecticut);
- Veera Baladandayuthapani (University of Michigan);
- Arvind Rao (University of Michigan).

Research partially supported by:

- National Science Foundation (NSF DMS 1613054);
- National Institutes of Health (NIH R01-CA214955);
- Bloomberg Data Science Grant.

Whole-tumour data objects from MR images



Data objects reside in non-Euclidean spaces with non-trivial geometries:

- Trees live on stratified spaces;
- Densities and tumour shapes (as curve) live on infinite-dimensional manifolds.

In this talk. . .

- I will provide an overview of the work based on the three representations, with little technical details.
- I will use a standard imaging dataset to explain ideas, and on which the methodologies were implemented.
- Focus will be on analysis of two sets of MR images pertaining to two patient groups:
  - Formal hypothesis test with trees;
  - Geometry-based clustering and PCA with densities and shapes.

The TCGA Dataset for Glioblastoma Multiforme (GBM)

- GBM is a morphologically heterogenous form of malignant brain cancer.
- Median survival time is about 12 months.

The TCGA Dataset for Glioblastoma Multiforme (GBM)

- GBM is a morphologically heterogenous form of malignant brain cancer.
- Median survival time is about 12 months.
- Images of 82 patients with histologically confirmed GBM and molecular data from The Cancer Genome Atlas (TCGA) database (`https://www.cancerimagingarchive.net/`).
- T1-post and T2-FLAIR images were registered spatially.
- The tumor region (slice with max surface area) was segmented with in-plane resolution of 1mm $\times$ 1mm.

TREES FROM TUMOUR IMAGES

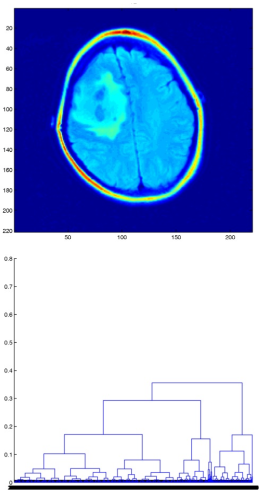
K. Bharath, D. Dey et al. Statistical Tests for Large Tree-structured Data. *Journal of the America Statistical Association*.(2017), 112, 1733-1743.

Binary trees from images

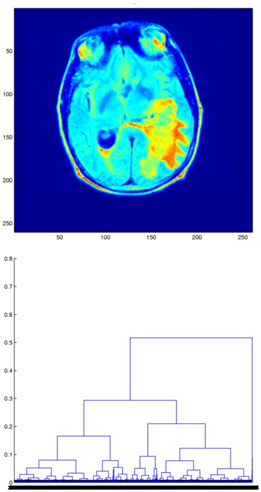
- Intra-tumor heterogeneity as 'clustering tendency' of pixels.
- Representation of groups of groups of groups of...... pixels: recursive partitioning of the set of pixels.

Leaves: Pixels
Internal nodes: clusters of pixels
Edge lengths: distance between clusters of pixels

Dataset: High variation in branch lengths



Long surviving ($\approx$ 60 months)



Short surviving ($\approx$ 5 months)

Issues with statistical models

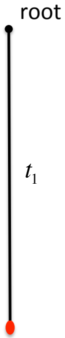
- Inference should not depend too much on choice of metrics and representations.
- Distributions should be unaffected by labelling scheme (Exchangeability).
- Hierarchical information between vertices should be preserved.
- Should be easy to simulate rich classes of binary trees.

A simple distribution on binary trees

- Consider a non-homogeneous Poisson process with rate $\lambda(t) = t$.
- Let $t_1, t_2, \ldots,$ be inter-event times.

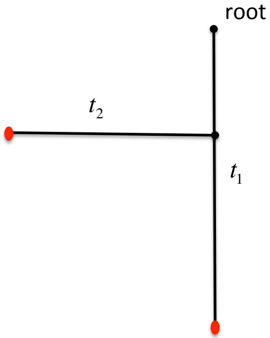
A simple distribution on binary trees

- Consider a non-homogeneous Poisson process with rate $\lambda(t) = t$.
- Let $t_1, t_2, \ldots,$ be inter-event times.



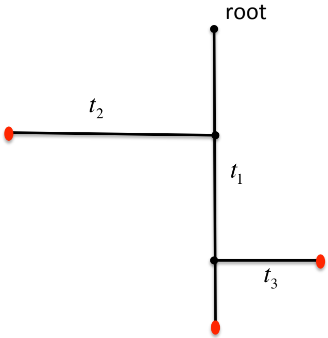
A simple distribution on binary trees

- Consider a non-homogeneous Poisson process with rate $\lambda(t) = t$.
- Let $t_1, t_2, \ldots,$ be inter-event times.



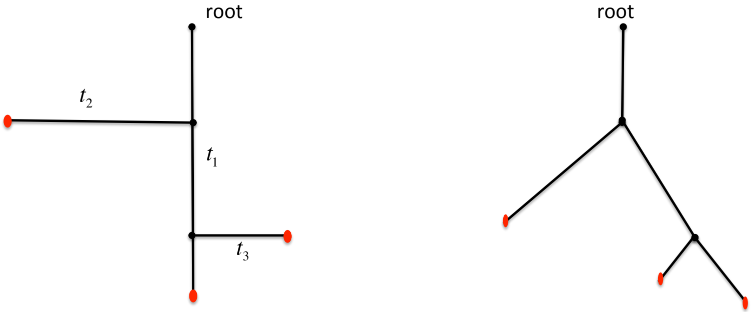
A simple distribution on binary trees

- Consider a non-homogeneous Poisson process with rate $\lambda(t) = t$.
- Let $t_1, t_2, \ldots,$ be inter-event times.



A simple distribution on binary trees

- Consider a non-homogeneous Poisson process with rate $\lambda(t) = t$.
- Let $t_1, t_2, \ldots,$ be inter-event times.



$t_1 + t_2 + t_3$ = total path length of binary tree

A simple distribution on binary trees

With n inter-event times, a binary tree $\tau(n)$ with n leaves or terminal nodes, $2n$ vertices and $2n-1$ edges is constructed.

Proposition

From the properties of the Poisson process with rate t, $\tau(n)$ can be given the density

$$f(\tau(n)) = \underbrace{\left[\prod_{i=1}^{n-1} \frac{1}{2i-1}\right]^{-1}}_{\text{topological info}} \underbrace{\frac{1}{2}^{n-1} s e^{-s^2/2}}_{\text{branch length info}}, \quad s = \sum_{i=1}^{2n-1} t_i.$$

- $f(\cdot)$ is exchangeable with respect to the branch lengths.

A simple distribution on binary trees

With n inter-event times, a binary tree $\tau(n)$ with n leaves or terminal nodes, $2n$ vertices and $2n-1$ edges is constructed.

Proposition

From the properties of the Poisson process with rate t, $\tau(n)$ can be given the density

$$f(\tau(n)) = \underbrace{\left[\prod_{i=1}^{n-1} \frac{1}{2i-1}\right]^{-1}}_{\text{topological info}} \underbrace{\frac{1}{2}^{n-1} s e^{-s^2/2}}_{\text{branch length info}}, \quad s = \sum_{i=1}^{2n-1} t_i.$$

- $f(\cdot)$ is exchangeable with respect to the branch lengths.
- Removal of a leaf from $\tau(n)$ results in a tree a with density $f(\tau(n-1))$.

A simple distribution on binary trees

With n inter-event times, a binary tree $\tau(n)$ with n leaves or terminal nodes, $2n$ vertices and $2n-1$ edges is constructed.

Proposition

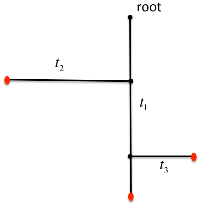
From the properties of the Poisson process with rate t, $\tau(n)$ can be given the density

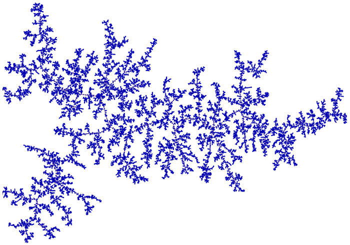
$$f(\tau(n)) = \underbrace{\left[\prod_{i=1}^{n-1} \frac{1}{2i-1}\right]^{-1}}_{\textit{topological info}} \underbrace{\frac{1}{2}^{n-1} s e^{-s^2/2}}_{\textit{branch length info}}, \quad s = \sum_{i=1}^{2n-1} t_i.$$

- $f(\cdot)$ is exchangeable with respect to the branch lengths.
- Removal of a leaf from $\tau(n)$ results in a tree a with density $f(\tau(n-1))$.
- $s =$ Sum of branch lengths characterizes f, and captures overall distance between pixel clusters (and is Gamma distributed).

Aldous'[1] Continuum Random Tree (CRT)

As $n \to \infty$,





(http://www.normalesup.org/ kortchem/english.html)

- CRT is invariant (to offspring distribution) limit of trees from (conditioned) Galton–Watson branching process.
- Binary trees from Poisson model 'finite-dimensional subtrees' of CRT. This allows generalisations to non-binary trees.

[1] D. Aldous. The Continuum Random Tree III. (1993). 21, 248-289

Invariant GoF test for binary trees: two-sample

Suppose $\boldsymbol{\tau(n)} = (\tau(n_1), \ldots, \tau(n_p))$ and $\boldsymbol{\eta(m)} = (\eta(m_1), \ldots, \eta(m_q))$ are independent samples of binary trees from π_τ and π_η respectively.

Theorem

Let r_j denote the sum of the branch lengths of $\eta(m_j)$, and without loss of generality assume that $\sum_{i=1}^{p} s_i > \sum_{j=1}^{q} r_j$. Then, the critical function

$$\psi(\mathbf{n}, \mathbf{m}, \alpha) = \begin{cases} 1 & \text{if } \frac{\sum_{i=1}^{p} s_i}{\sum_{j=1}^{q} r_j} > \left(\frac{\sum_{i=1}^{p} n_i}{\sum_{j=1}^{q} m_j}\right) F_{1-\alpha, 2\sum_{i=1}^{p} n_i, 2\sum_{j=1}^{q} m_j}; \\ 0 & \text{otherwise}, \end{cases}$$

where $F_{\alpha,a,b}$ is the αth percentile of an F distribution with a and b degrees of freedom, for testing $H_0 : \pi_\tau = \pi_\eta = f$, is such that $E_{H_0}\psi(\mathbf{n}, \mathbf{m}, \alpha) = \alpha$. The test is invariant to the permutation of the leaves.

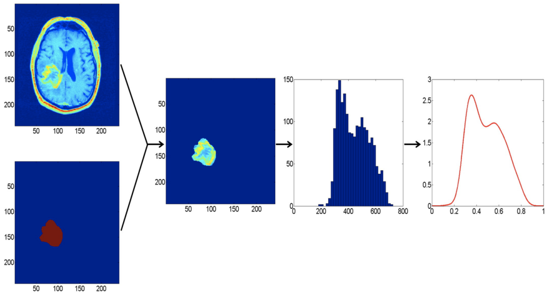
Two-sample test to detect heterogeneity

- From binary trees obtained from hierarchical clustering, we chose binary subtrees by randomly selecting a subset of leaves and constructing their Least Common Ancestor trees (see paper for details).

Two-sample test to detect heterogeneity

- From binary trees obtained from hierarchical clustering, we chose binary subtrees by randomly selecting a subset of leaves and constructing their Least Common Ancestor trees (see paper for details).
- Using the survival times, we created two groups of patients: those with survival times $\leq$ 12 months and those $>$12 months.
- Differences in groups was detected by LCA-based test at 1% significance level.
- Naive Bayes classifier with the likelihood from LCA trees, provided 69% classification accuracy.

PROBABILITY DENSITIES FROM TUMOUR IMAGES

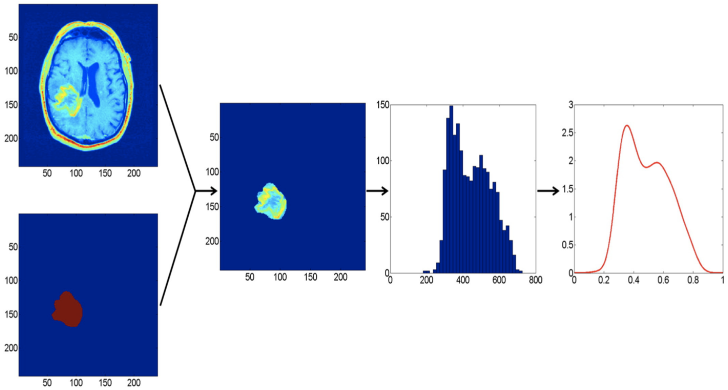


*S. Kurtek and K. Bharath. Bayesian Sensitivity Analysis with Fisher–Rao metric. *Biometrika*. (2015), 102, 616.

*K. Bharath et al. DEMARCATE: Density-based Magnetic Resonance Image Clustering for Assessing Tumor Heterogeneity in Cancer. *NeuroImage*. (2016). 12, 132-143.

*A. Saha, K. Bharath and S. Kurtek. A Geometric Variational Approach to Bayesian Inference. Minor Revision with *Journal of the America Statistical Association*.

Voxel density representation



- Captures intra-tumour heterogeneity.
- Commonly used with numerical summaries.

Hilbert space geometry of the space of PDFs

- Parametric families of densities is not appropriate.
- However, the nonparametric family $\mathcal{P} = \left\{ f : \mathbb{R} \to \mathbb{R}^+ \cup \{0\} : \int_{\mathbb{R}} f(x)dx = 1 \right\}$, is a (non-linear) Banach manifold.

Hilbert space geometry of the space of PDFs

- Parametric families of densities is not appropriate.
- However, the nonparametric family $\mathcal{P} = \left\{ f : \mathbb{R} \to \mathbb{R}^{+} \cup \{0\} : \int_{\mathbb{R}} f(x)dx = 1 \right\}$, is a (non-linear) Banach manifold.
- The nonparametric Fisher-Rao metric is:

$$\langle\langle \delta f_1, \delta f_2 \rangle\rangle_f = \int_{\mathbb{R}} \delta f_1(x) \delta f_2(x) \frac{1}{f(x)} dx.$$

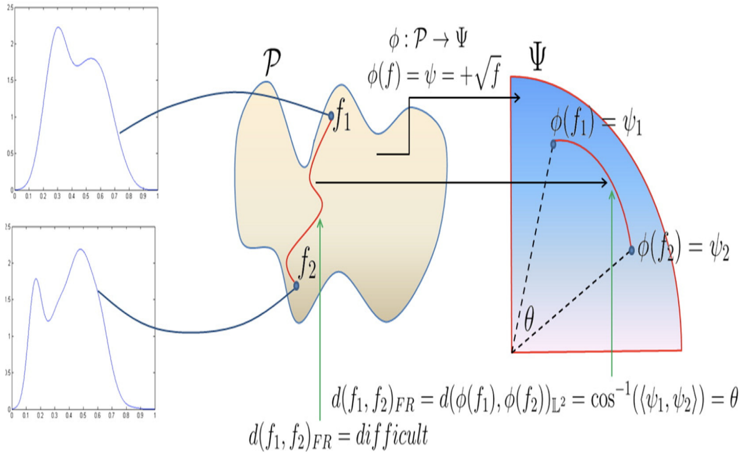
Hilbert space geometry of the space of PDFs

- Parametric families of densities is not appropriate.
- However, the nonparametric family $\mathcal{P} = \left\{f : \mathbb{R} \to \mathbb{R}^{+} \cup \{0\} : \int_{\mathbb{R}} f(x)dx = 1\right\}$, is a (non-linear) Banach manifold.
- The nonparametric Fisher-Rao metric is:

$$\langle\langle \delta f_1, \delta f_2 \rangle\rangle_f = \int_{\mathbb{R}} \delta f_1(x) \delta f_2(x) \frac{1}{f(x)} dx.$$

- The metric is invariant to one-to-one transformations, but distance is difficult to compute.
- Under the mapping $f \mapsto +\sqrt{f}$, we move to the positive orthant of the unit $\mathbb{L}^2$ sphere , and Fisher–Rao metric transforms to the usual $\mathbb{L}^2$ metric.

Hilbert space geometry of the space of PDFs



Distance-based clustering of images with PDFs

- Equipped with a computable distance d_{FR} on $\mathcal{P}$, the (sample) Frechét mean $\hat{f}_{mean}$ can be defined as

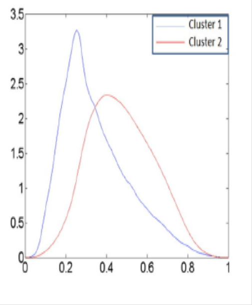
$$\text{argmin}_{\mu \in \mathcal{P}} \sum_{i=1}^{n} d_{FR}^2(f_i, \mu).$$

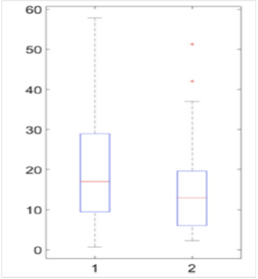
Distance-based clustering of images with PDFs

- Equipped with a computable distance d_{FR} on $\mathcal{P}$, the (sample) Frechét mean $\hat{f}_{mean}$ can be defined as

$$\mathrm{argmin}_{\mu \in \mathcal{P}} \sum_{i=1}^{n} d_{FR}^2(f_i, \mu).$$

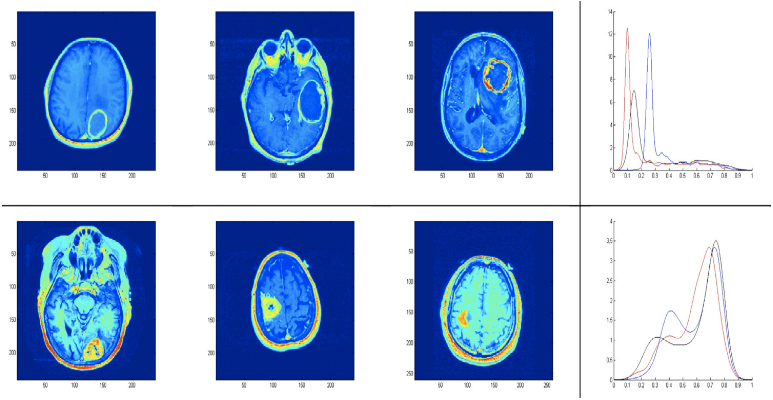
- k-means clustering with $k = 2$ can be carried out.





Left: Cluster 1
Right: Cluster 2

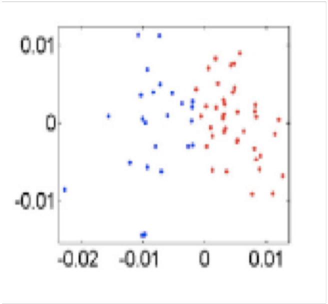
Distance-based clustering of images with PDFs

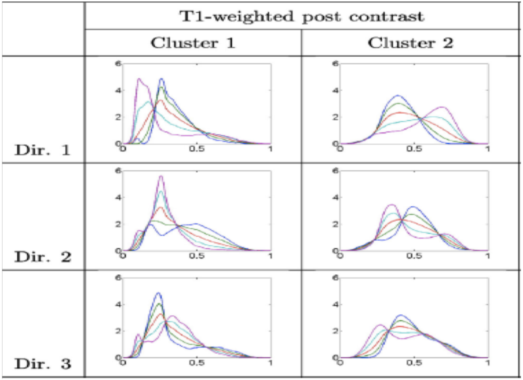


Top: 3 images from Cluster 1; Bottom: 3 images from Cluster 2.

Visualising clustering results through PCA

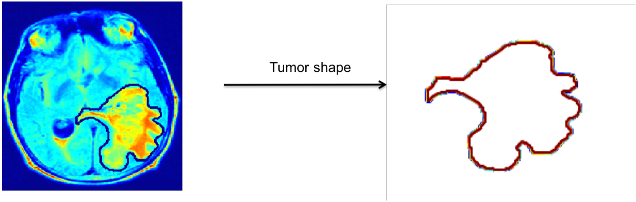
At the tangent space of $\hat{f}_{mean}$, the sample covariance operator can be estimated, enabling PCA.





Details and further results in the papers.

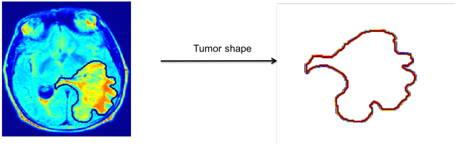
Shapes from tumour images



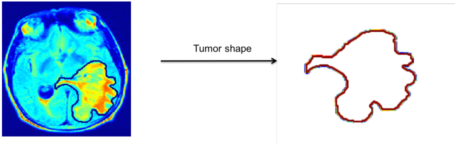
*K. Bharath et al. Radiologic Image-based Statistical Shape Analysis of Brain Tumors. *Journal of the Royal Statistical Society–Series C*. (2018+).

*K. Bharath and S. Kurtek. Distribution and Sampling of Warps Maps for Curve Alignment. Minor Revision with *Journal of the America Statistical Association*.

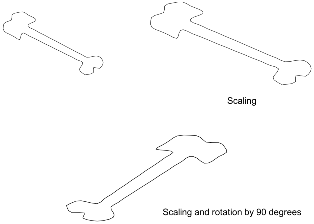
Representation of tumour shapes



Representation of tumour shapes



- Absolutely continuous parameterised closed curve: $C : \mathbb{S}^1 \to \mathbb{R}^2$.
- The notion of a shape of such a curve requires invariances to transformations that represent nuisance information, such as translations, rotations and reparameterisations of C.



Representation of tumour shapes

- $\Gamma := \{\gamma : S \to S$ is an orientation-preserving diffeomorphism$\}$
- $SO(2)$ is the rotation group in $\mathbb{R}^3$

The shape of a parameterized curve C is defined to be the equivalence class

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

The set of shapes is the quotient space under the actions of relevant transformation groups, and is an infinite-dimensional non-linear manifold.

Riemannian geometric framework

The Square-Root Transform (SRT)[2]: Consider the bijective (mod translations) transform

$$C \mapsto Q_C := \frac{C'}{\|C'\|}.$$

- Removes scale and translation variation;

[2] A. Srivastava and E Klassen. *Functional and Shape Data Analysis*. (2016). Springer, NY.

Riemannian geometric framework

The Square-Root Transform (SRT)[2]: Consider the bijective (mod translations) transform

$$C \mapsto Q_C := \frac{C'}{\|C'\|}.$$

- Removes scale and translation variation;
- Allows the definition of a valid, computable Riemannian distance between tumours which is invariant to rotations, translations, scalings and reparameterisations:

$$d(C_1, C_2) := \inf_{(\gamma, O) \in \Gamma \times SO(2)} \|Q_{C_1} - O Q_{C_2}(\gamma)\sqrt{\gamma'}\|.$$

[2] A. Srivastava and E Klassen. *Functional and Shape Data Analysis*. (2016). Springer, NY.

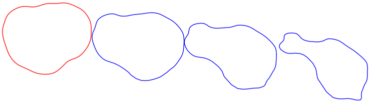
Visualising shape deformations

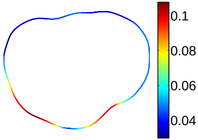
Matching tumour shape on Left with the one on the Right. Middle: is shape 2 after reparameterisation.

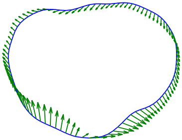
Geodesic path between tumour shape (a) and tumour shape (b).

PCA on tumour shape space

- The distance allows us to compute sample Frechét mean and perform PCA on the set of tumour shapes.
- Following the deformation vector field along the geodesic in the direction of decreased survival gives:

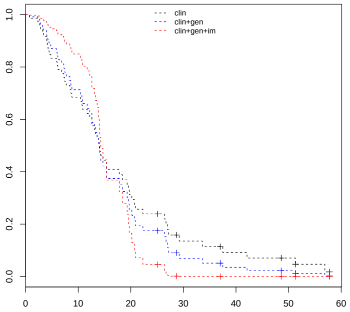






Cox survival model with tumor shape predictors (PC regression)

Model	C-index
Clinical predictors only	0.652
Clinical + Genetic predictors	0.722
Clinical+Genetic+Shape PCs	0.859



— Clinical
— Clinical + Genetic
— Clinical + Genetic + Shape

Sundries

All papers are available on arXiv.

Email: Karthik.Bharath@nottingham.ac.uk

- C++ code for trees available on Github page pkambadu/DyckPath.
- For shapes and PDFs:
  - R package fdasrvf, maintained by Derek Tucker (jdtuck@sandia.gov).
  - Matlab stand-alone programs available on Sebastian Kurtek's (kurtek.1@stat.osu.edu) website at OSU.