02/05/19

Ezra Miller:

- Goal: statistics where sample space is a stratified space X
  e.g. trees, shapes, persistence diagrams (inf. dimensional — not covered), diffusion tensors
- ordinary statistics: object = vector, sample space = vector space
  methods: mean, variance, PCA, LLN, CLT
- stratified stat.: all of these → raise fundamental problems in geometric probability
- Def. A manifold is a second-countable, Hausdorff topological space M s.t. every point $x \in M$ has a neighborhood $\cong$ open ball in $\mathbb{R}^n$.
- chart: $\pi_\alpha : U_\alpha \hookrightarrow M$ (open ball in $\mathbb{R}^d$), $M_{\alpha\beta} = \pi_\alpha(U_\alpha) \cap \pi_\beta(U_\beta)$

$\pi_\beta^{-1}(U_\beta)$

$\rightsquigarrow \pi_{\alpha\beta} : \pi_\alpha^{-1}(M_{\alpha\beta}) \to \pi_\beta^{-1}(M_{\alpha\beta})$, $\pi_{\alpha\beta}(x) = \pi_\beta^{-1} \circ \pi_\alpha(x)$

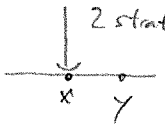
$\pi_{\alpha\beta}$ topological / smooth / analytic / algebraic $\rightsquigarrow$ topological / smooth / analytical manifold / algebraic variety

{all π_α} = atlas

- data examples: angles (circle), rotations ($SO(3)$), lines ($\mathbb{RP}^{n-1}$), subspaces ($Gr_{k,n}$)

- stratified spaces : **Def** A *topologically stratified space* is a Hausdorff topological space X that is a disjoint union $X = M_1 \cup M_2 \cup \dots \cup M$ of manifolds called *strata* s.t.
  0) each stratum has finitely many connected components
  1) $M_1 \cup \dots \cup M_k$ is closed in X $\forall k \leq l$
  2) $\forall x, y \in M_i$: $\exists$ homeomorphism $\varphi : X \to X$ with
     - $\varphi(M_k) = M_k$ $\forall k$
     - $\varphi(x) = \varphi(y)$

     ↝ points "look the same" on the same stratum

     ↝ not: 

- examples :

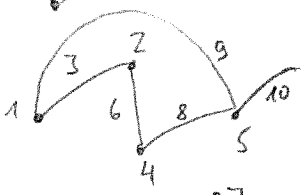
  dim 0 : finite set with discrete topology

  ↝ 1 disconnected stratum or lots of connected strata or ...

  ↝ same space but different stratifications

  dim 1 : graphs ↝ first vertices (condition 1) then edges

  e.g.  not  (but without 3 ok)

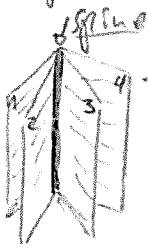
  

  e.g. trees : connected, no cycles 

  or spider $\mathcal{Y}$, or star :  ← no vertices at the end

  , $d=0$: spider

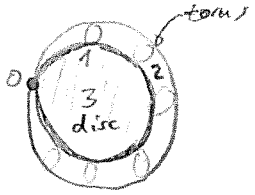
  higher dim. : e.g. open book = spider $\times \mathbb{R}^d$

  $d=1$: 

  = spine ∪ open pages
  M_0 $M_1, \dots, M_k$

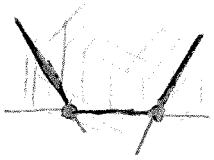
GM example :

Goresky & MacPherson (?) ↝ book



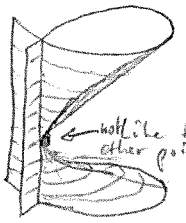
dim	0	1	2
#	1	0	1	2

polyhedron : intersection of finitely many half-spaces in $\mathbb{R}^d$



dim	0	1	2
#	2	3	1

Whitney cusp:



$sing(X) = \{\}$

but $X = sing(X) \cup (X \setminus sing(X))$

is not a (top.) stratification

shape spaces, eg. face recognition ($d=3$)

n (landmarks), labelled

d [| | | |] $= A \in \mathbb{R}^{d \times n}$, A' same shape if $A' = gA$,

$g \in G$ = group of transformations, e.g. G = rigid motions

or G could have rotations, translations, projective transformations, reflections

$\rightsquigarrow$ $X = G \backslash \mathbb{R}^{d \times n}$ ~ (typically) algebraic variety $\rightsquigarrow$ stratified (see below)

after removing unstable orbits (keep stable and semi-stable orbits)

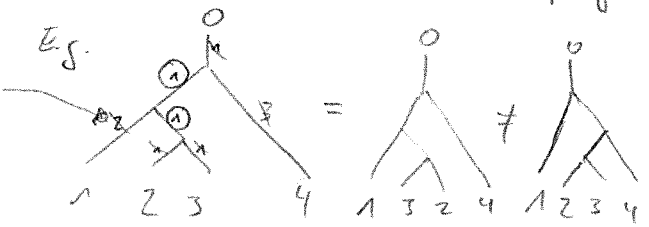
e.g. $\Sigma_d^n = \mathbb{R}^{d \times n} \setminus \{$all columns equal$\} / $ Similarities

$d = 2$: $\Sigma_2^n \cong \mathbb{CP}^{n-2}$, else ($d > 2$): "non-free shapes"

Def. A phylogenetic n-tree is a tree with $n+1$ labelled leaves (vertices of degree 1) $\rightsquigarrow$ lower dim. strata

and edge lengths > 0 on internal edges (no pendant edges)

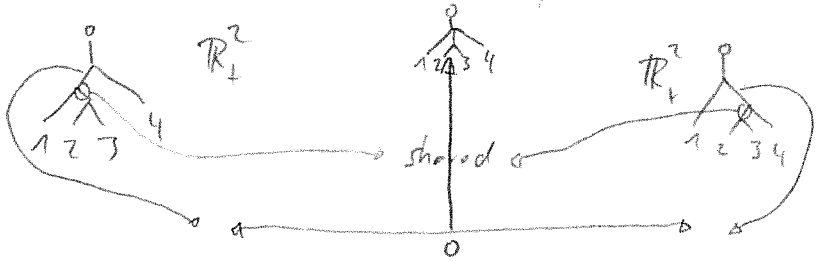
$\overline{T}_n = \{n\text{-trees}\}$



$\rightsquigarrow$ removing edge gives partition of leave set into two connected components

$\rightsquigarrow$ fixed "combinatorics": specify internal edge lengths, eg. $\mathbb{R}_+^2$

($\mathbb{R}_+ = \mathbb{R}_{>0}$, $\overline{\mathbb{R}}_+ = \mathbb{R}_{\geq 0}$) $\rightsquigarrow$ orthants $\mathbb{R}_+^2$ are manifolds



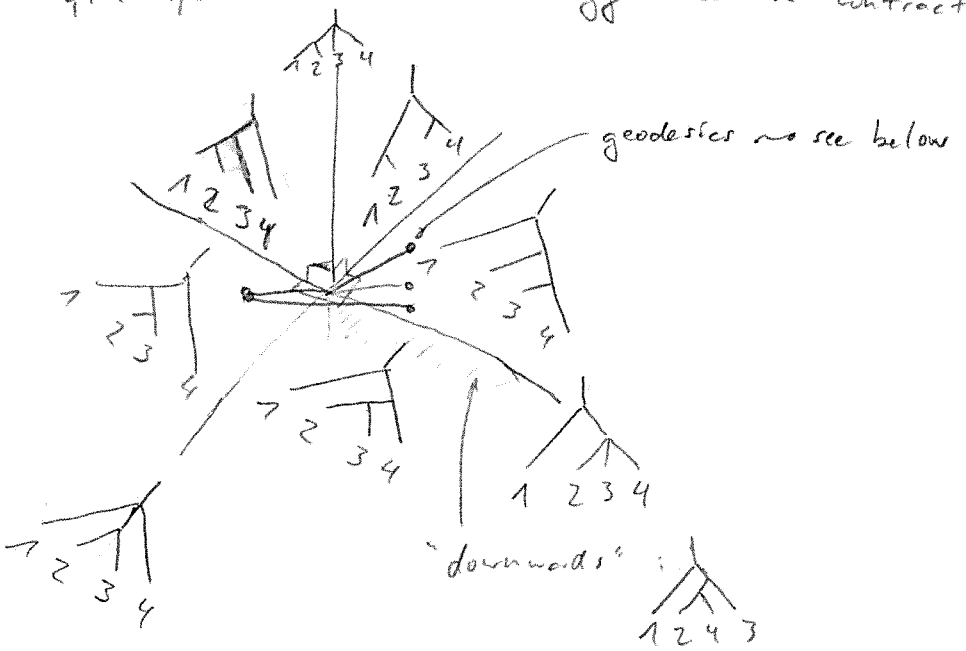
(Billera-Holmes-Vogtmann 1998)

facets of $\overline{T}_n$: orthants $\overline{\mathbb{R}}_+^{n-2}$ in bijection with binary tree topologies (indexed by)

faces of $\overline{T}_n$: orthants $\overline{\mathbb{R}}_+^d$ for $d \leq n-2$ # of edges

tree T: $x_T \in \mathbb{R}_+^{e(T)}$ edges $\hat{=}$ splits $\rightsquigarrow$ strata, $Q_T \subseteq \overline{Q}_{T'} \Leftrightarrow T_{|S} = $ contraction of T'

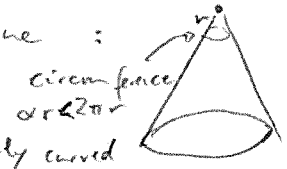
$\bar{O}_T = \bar{O}_{T'} \cap \bar{O}_{T''}$ $\Leftrightarrow$ T is the biggest common contraction



diffusion tensors : positive semidefinite matrices of size 3

stratification by subsets of eigenvalues that are equal

ice-cream cone :



$\cong \mathbb{R}^2$

$\rightsquigarrow$ positively curved

Local geometry away from apex

$\alpha > 2\pi$, e.g. $5\frac{\pi}{2}$ $\rightsquigarrow$ negatively curved : Kale

contained in $\mathcal{T}_4$, see above

- Tubular neighbourhood Theorem : X is topologically stratified

$\Rightarrow$ a tubular neighbourhood of each stratum S is a fiber bundle over S (i.e. it is locally a cross-product of S and N_s) with fiber N_S where N_S is homeomorphic to a cone over a stratified space L with $\dim L = \dim N_S - 1$.

$\rightsquigarrow$ tangent cone at $p \in S$: $T_p X = \underbrace{T_p S}_{\text{tangent manifold}} \times \underbrace{N_S X}_{\text{normal slice} \cong N_S}$

even more : the tangent bundle is locally trivial

e.g. open book : on page $N_p S$ = one point = cone over $\emptyset$

on spine : $N_p S$ = spider = cone over k points

- Def. $\dim X = \max_k (\dim M_k)$
- Suppose $p \in S$ stratum of dim. $d-1$ where $d = \dim X$.

  $N_pX = ?$ $T_pX = ?$ $\quad T_pX = \underbrace{T_pS}_{\dim: d-1} \times \underbrace{N_pX}_{1}$

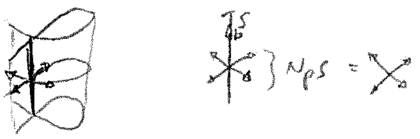
Thm. If M_i and M_j are strata, $M_i \cap \overline{M_j} \neq \emptyset \Rightarrow M_i \subseteq \overline{M_j}$

Cor. Strata are partially ordered: $M_i \leq M_j \Leftrightarrow M_i \subseteq \overline{M_j}$

$p \in \overline{S'} \Rightarrow \overline{S'} \geq S \leadsto N_pX$ is cone over a finite set (conn. components of intersection of tub. neigh. with d-dim strata containing S) $\leadsto$ spider

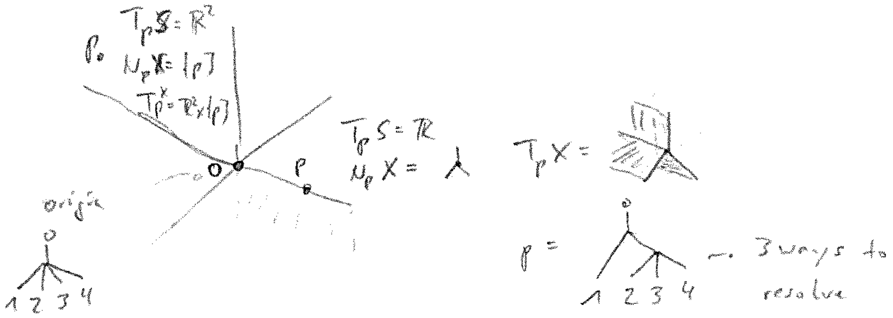
$\leadsto T_pX$: d-dim. open book $\leadsto$ universal for points in co-dim. 1 strata

e.g.

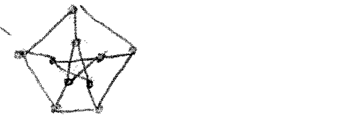


therefore at the Whitney cusp the point must be its own stratum

e.g. J_4



$T_0 J_4 = N_0 J_4 =$ cone over Peterson graph



- Metrics, smoothness etc.:

  "Riemannian stratified space" when properly defined should require top. strat. geodesic* metric space s.t. each stratum is Riemannian

  ["smooth stratified space" $\leadsto$ see [Pflaum] (book)]

  e.g. icecream cone / Kale: apex has to be its own stratum to get Riemannian strat. space

  exercise: embed J_3 into kale K_α for $\alpha > 3\pi \longrightarrow$ 3 angles each larger than π

*geodesic has to be required (at least intrinsic), else extrinsic metric on λ would be allowed

- **Def.** A Whitney stratified space is a top. strat. space X embedded in a vector space where limits of secant lines joining strat. M_i and M_j with $M_i \subseteq \bar{M}_j$ are contained in limits of tangent planes to M_j as points converge to M_i

e.g. real or complex alg./semianalytic/subanalytic varieties

Thm. Whitney stratification $\Rightarrow$ triangulation exists

- X strat. metric space

Def. A prob. distr. ρ on X has Fréchet function

$$F(y) = \int_X d(x,y)^2 \rho(dx)$$

ρ has Fréchet mean $\operatorname{argmin}_{y \in X} F(y)$.

- usual stat.: data wiggles $\Rightarrow$ mean wiggles

strat. stat.: mean can stick

e.g. 

$a = b = c = 1 \Rightarrow \mu_{\text{mean}} = 0$

$$F(\rho) = a^2 + b^2 + c^2$$

$$F(\rho_\varepsilon) = F(\rho) - 2\varepsilon a + 2\varepsilon b + 2\varepsilon c + 3\varepsilon^2$$

$$\rightsquigarrow F(\rho_\varepsilon) - F(\rho) = 2\varepsilon(\underbrace{b+c-a}_{>0 \text{ if } a < b+c \text{ for small } \varepsilon}) + 3\varepsilon^2$$

higher dim.: e.g. open book

- What is a law of large numbers (LLN)?

$X_1, \dots, X_n$ ind. rand. var. with values in X distributed according to ρ

LLN: $\bar{X}_n \longrightarrow \mu(\rho)$ as $n \to \infty$.

$\mu(\rho)$ is sticky $\Rightarrow$?

Fix X = open book = K = spider where $K = \mathbb{R}^d$ is spine. X is CAT(0) $\Rightarrow$ unique Fréchet mean

An integrable prob. distr. ρ has 3 possibilities for mean $\mu(\rho)$:

1) $\mu(\rho) \notin K$
2) $\mu(\rho) \in K$ and $\mu(\rho') \in K$ for all ρ' "near" ρ
3) $\mu(\rho) \in K$ but $\mu(\rho') \notin K$ for some ρ' "arbitrarily near" ρ

Def. Corresponding to these cases, $\mu(\rho)$ is 1) non-sticky, 2) sticky, 3) partly sticky (there is also a perturbation which makes it sticky)

Thm [SAMSI WG] iid $X_1, X_2, \dots$ in $X \sim p \Rightarrow \bar{X}_n \to \mu(p)$ as $n \to \infty$

Moreover: • $\mu(p)$ is not sticky or partly sticky ((1)+(3)) & supp p spreads over 3 pages $\Rightarrow \exists$ page L and rand. $N \in \mathbb{N}$ s.t. $\bar{X}_n \in \bar{L}$ $\forall n \geq N$ a.s.

• $\mu(p)$ sticky $\Rightarrow \exists$ rand. $N \in \mathbb{N}$ s.t. $\bar{X}_n \in K$ $\forall n \geq N$ a.s.

• What is a central limit theorem (CLT)?

$X_1, X_2, \dots \overset{iid}{\sim} p$, $\bar{X}_n \xrightarrow{\mathcal{D}} \delta_{\mu(p)}$

idea: rescale $\sqrt{n}\,\bar{X}_n \xrightarrow{\mathcal{D}} \tilde{p}$ limiting distr. (classical: $\tilde{p}$ gaussian)

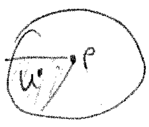
Thm [SAMSI WG] A CLT holds for square-int. p holds with limiting distribution

1) gaussian on $\mathbb{R}^{d+1} = T_{\mu(p)}X = T_{\mu(p)}L$

2) gaussian on $\mathbb{R}^{d} = K = T_{\mu(p)}K$

3) half gaussian on L plus its reflected projection to K ($\frac{1}{2}$ * gaussian on $T_{\mu(p)}K$)

• Def. Fix a metric space X and a top set $\mathcal{P}$ of (int) prob. distr. on X as well as a subset $K \subseteq X$.

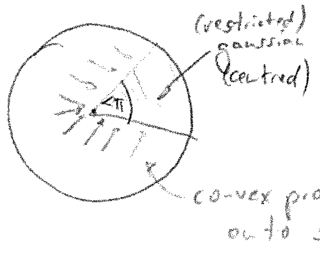


The mean of $\mu(p)$ of $p \in \mathcal{P}$ sticks to K if every neighborhood U of p in $\mathcal{P}$ contains a nonempty open set $U' \subseteq U$ with $\mu(p') \subseteq K$ $\forall p' \in U'$.

! • Example (Huckemann, Hattingly, Nolen, Miller):

X = isolated planar hyperbolic singularity (Kale with $\alpha > 2\pi$)

$\Rightarrow$ CLT has limiting distribution:



⇝ induces further stratification (boundary of sector)