

Convexity, Simplexes, Probabilities

Matilde Marcolli

Ma148: Geometry of Information
Caltech, Fall 2021

Convexity is the mathematical property that describes the operation of **mixing**

- notation: \mathbb{E}^n Euclidean space in n -dimensions, \mathbb{A}^n affine space (usually over \mathbb{C} or \mathbb{R}), $\mathbb{P}^n = (\mathbb{A}^{n+1} \setminus \{0\})/\mathbb{G}_m$ projective space (quotient by multiplicative group \mathbb{C}^* or \mathbb{R}^*)
- over \mathbb{R} : line segment in \mathbb{E}^n

$$\ell_{\underline{x}_1, \underline{x}_2} = \{ \underline{x} = \lambda_1 \underline{x}_1 + \lambda_2 \underline{x}_2 \mid \lambda_i \geq 0, \lambda_1 + \lambda_2 = 1 \}$$

convex linear combinations of extremal points $\underline{x}_1, \underline{x}_2$

- barycentric coordinates: **simplex**

$$\Delta_{\underline{x}_1, \dots, \underline{x}_N} = \{ \underline{x} = \lambda_1 \underline{x}_1 + \lambda_2 \underline{x}_2 + \dots + \lambda_N \underline{x}_N \mid \lambda_i \geq 0, \sum_i \lambda_i = 1 \}$$

- **convex set**: each point in the set has unobstructed view of all other points in the set (line segment connecting them is entirely contained in the set)

- set $S \subset \mathbb{A}^n$ **convex hull** $H(S)$ smallest convex set in \mathbb{A}^n containing S
- if S is a finite set of points $H(S)$ is a **polytope**
- if S consists of $N + 1$ points not all on any $N - 1$ affine subspace, $H(S)$ is an N -simplex Δ_S (or $\Delta_N(S)$)
- C convex set in some ambient \mathbb{A}^N : dimension $\dim C = n$ if largest n for which C contains a simplex Δ_n
- slices $C \cap \mathbb{A}^k$ of convex sets by lower dimensional affine subspaces $\mathbb{A}^k \subset \mathbb{A}^N$ are also convex sets
- $C \subset \mathbb{A}^N$ convex set, **cone** $\hat{C} \subset \mathbb{A}^{N+1}$: take point $\underline{x} \in \mathbb{A}^{N+1}$ not in C and union \hat{C} of all half-line from \underline{x} passing through points of C (unbounded cone)
- cones have a **partial ordering** $\underline{x} \leq \underline{y}$ in \hat{C} iff $\underline{y} - \underline{x} \in \hat{C}$
- cone $\hat{C} \subset \mathbb{E}^N$, **dual cone**: $\hat{C}^\vee =$ linear functionals $f \in \mathbb{E}^{n^\vee} = \text{Hom}(\mathbb{E}^n, \mathbb{R})$ such that $f(\underline{x}) \geq 0$ for all $\underline{x} \in \hat{C}$
- with Euclidean metric identify $\mathbb{E}^n \simeq \mathbb{E}^{n^\vee}$, then cone is self-dual if $\hat{C} = \hat{C}^\vee$ in this identification

- topologically an n -dimensional convex set is always homomorphic to an n -ball (move along the rays from an interior point, center of the ball)
- **convex body** C a compact convex set (while convex cones \hat{C} always non-compact)
- convex bodies and convex cones over a convex body always have **extremal points** that are not convex combinations of other points (**pure points, pure states**)
- but not true for other convex cones (eg half-space)
- k -dimensional face $F \subset C$ of an n -dimensional convex set: if a point $\underline{x} \in C$ is in F and $\underline{x} = \lambda \underline{x}_1 + (1 - \lambda) \underline{x}_2$, for some $0 \leq \lambda \leq 1$, then $\underline{x}_1, \underline{x}_2$ also in F
- partial ordering by inclusion of the set of faces of a convex set (lattice of inclusions)

Facts about convex sets

- **Minkowski**: every convex body Y is convex hull $Y = C(X)$ of its set X of pure points
- **Carathéodory**: given $X \subset \mathbb{E}^n$ (so $\dim X \leq n$), any $\underline{x} \in H(X)$ is a convex combination of at most $n + 1$ points of X

$$\underline{x} = \sum_i \lambda_i \underline{x}_i \quad \sum_i \lambda_i = 1, \quad \lambda_i \geq 0$$

Note that here $\lambda_i, \underline{x}_i$ **depend on \underline{x}**

- case of simplex: all points written as a mixture in a **unique way** (this property characterizes simplices)
- number of k -faces in an n -simplex

$$\#\{k\text{-faces}\} = \binom{n+1}{k+1} = \binom{n+1}{n-k} = \#\{(n-k-1)\text{-faces}\}$$

lattice of faces is self-dual (in fact it is also a Boolean lattice)

- supporting hyperplanes of convex sets: $C \subset \mathbb{E}^n$ real hyperplane H that intersects C and such that all of C lies in the same half-space cut out by H
- **regular point** of C , point \underline{x} on ∂C such that there is only one support hyperplane through \underline{x}
- **regular support hyperplane** meets C at only one point
- **regular convex set** C all points and all support hyperplanes are regular
- a ball is regular a simplex is not
- **Hahn–Banach**: $C \subset \mathbb{E}^n$ convex body and $\underline{x}_0 \in \mathbb{E}^n \setminus C$ then $\exists f : \mathbb{E}^n \rightarrow \mathbb{R}$ linear functional such that $f(\underline{x}) > 0$ for all $\underline{x} \in C$ and $f(\underline{x}_0) < 0$ (think in terms of support hyperplanes)
- **convex functions** $f : C \rightarrow \mathbb{R}$ from C convex body such that

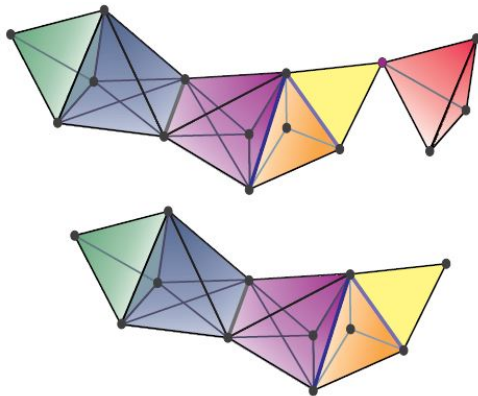
$$f(\lambda \underline{x} + (1 - \lambda) \underline{y}) \leq \lambda f(\underline{x}) + (1 - \lambda) f(\underline{y})$$

$$\forall \underline{x}, \underline{y} \in C \text{ and } \forall \lambda \in [0, 1]$$

Another way to think of simplices: **Simplicial Sets**

Simplicial Sets: intuition

Sets locally described by simplexes appropriately glued together along faces (subtleties: some simplexes can be degenerate etc.)



Idea: these sets locally have “mixing property” of simplexes (barycentric coordinates) but are not necessarily globally convex

Categories

- **Category:** \mathcal{C} with **objects** $Obj(\mathcal{C})$ and **morphisms** $\text{Hom}_{\mathcal{C}}(X, Y)$ for $X, Y \in Obj(\mathcal{C})$ (small categories: $Obj(\mathcal{C})$ and $\text{Hom}_{\mathcal{C}}(X, Y)$ are sets); composition

$$\text{Hom}_{\mathcal{C}}(X, Y) \times \text{Hom}_{\mathcal{C}}(Y, Z) \rightarrow \text{Hom}_{\mathcal{C}}(X, Z), \quad (\phi, \psi) \mapsto \psi \circ \phi$$

is associative; there is an identity morphism

$1_X \in \text{Hom}_{\mathcal{C}}(X, X)$ for every object X

- **Functors** $F : \mathcal{C} \rightarrow \mathcal{C}'$ between two categories, maps objects to objects $F : Obj(\mathcal{C}) \rightarrow Obj(\mathcal{C}')$ and morphisms to morphisms $F : \text{Hom}_{\mathcal{C}}(X, Y) \rightarrow \text{Hom}_{\mathcal{C}'}(F(X), F(Y))$ respecting composition and identities, $F(\phi \circ \psi) = F(\phi) \circ F(\psi)$ and $F(1_X) = 1_{F(X)}$

- **Natural transformations** of functors: $F, G : \mathcal{C} \rightarrow \mathcal{C}'$ functors, natural transformation $\eta : F \rightarrow G$
 - for all objects $X \in \text{Obj}(\mathcal{C})$ a morphism $\eta_X : F(X) \rightarrow G(X)$, $\eta_X \in \text{Hom}_{\mathcal{C}'}(F(X), G(X))$
 - for every morphism $\phi : X \rightarrow Y$, $\phi \in \text{Hom}_{\mathcal{C}}(X, Y)$, compatibility $\eta_Y \circ F(\phi) = G(\phi) \circ \eta_X$

$$\begin{array}{ccc}
 F(X) & \xrightarrow{\eta_X} & G(X) \\
 F(\phi) \downarrow & & \downarrow G(\phi) \\
 F(Y) & \xrightarrow{\eta_Y} & G(Y)
 \end{array}$$

Another way to think of simplices: **Simplex Category** (or Δ category)

- **Simplex Category**: objects in $Obj(\Delta)$ are totally ordered sets $[n] = \{0, 1, \dots, n\}$; morphisms in $\text{Hom}_{\Delta}([n], [m])$ are nondecreasing maps $f : [n] \rightarrow [m]$

$$f(i) \leq f(j) \quad \text{for } i \leq j$$

- this category “looks like a simplex”: this can be seen by showing that the morphisms are generated by a set of faces and degeneracies morphisms (all morphisms are finite compositions of those)

Simplex Category: Faces and Degeneracies

- D_i increasing injection $[n] \rightarrow [n+1]$ not taking value i ; S_i nondecreasing surjection taking value i twice
- morphisms are generated by maps $D_i : [n] \rightarrow [n+1]$ and $S_i : [n+1] \rightarrow [n]$

$$D_i[0, \dots, n] = [0, \dots, \hat{i}, \dots, n], \quad S_i[0, \dots, n] = [0, \dots, i, i, \dots, n]$$

- in Δ^{op} the D_i become face maps $d_i : [n+1] \rightarrow [n]$ and S_i the degeneracy maps $s_i : [n] \rightarrow [n+1]$

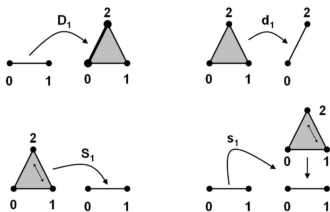


image of degeneracy s_1 degenerate 2-simplex image of collapse map S_1

- generators in Δ faces $D_i^n : [n] \rightarrow [n+1]$ and degeneracies $S_i^n : [n+1] \rightarrow [n]$

- relations

$$D_i^{n+1} \circ D_j^n = D_{j+1}^{n+1} \circ D_i^n \quad i \leq j$$

$$S_j^n \circ S_i^{n+1} = S_i^n \circ S_{j+1}^{n+1} \quad i \leq j$$

$$S_j^n \circ D_i^{n+1} = \begin{cases} D_i^n \circ S_{j-1}^{n-1} & i < j \\ \text{Id}_n & i = j \text{ or } i = j + 1 \\ D_{i-1}^n \circ S_j^{n-1} & j + 1 < i \end{cases}$$

- any nondecreasing maps $f : [n] \rightarrow [m]$ is a finite composition of D_i^k 's and S_j^ℓ 's

Simplicial Sets: rigorous definition

- **Simplicial set**: functor $X : \Delta^{op} \rightarrow \mathcal{S}$ to the category of sets (contravariant functor from Δ)
- this means collection of sets $X_n = X([n])$ images of objects of Δ -category endowed with face and degeneracy maps: n -skeleta of the simplicial set

Realization

- **Realization of a simplex:** $|\Delta^n|$ the geometric simplex realization of combinatorial $\Delta^n = [n]$
- Standard simplex $|\Delta^n|$ is simplex $\Delta_n = \Delta_{e_0, \dots, e_n}$ of standard o.n. basis in \mathbb{E}^{n+1}

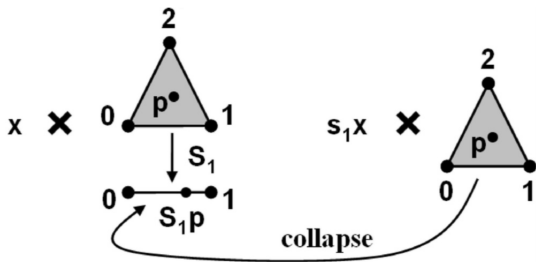
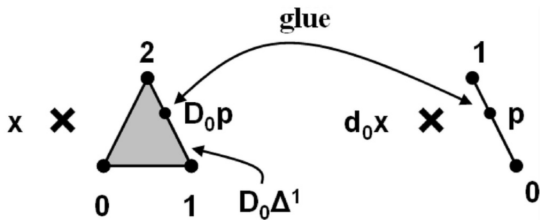
$$|\Delta^n| = \Delta_n = \{x \in \mathbb{E}^{n+1} \mid x_i \geq 0 \text{ and } \sum_{i=0}^{n+1} x_i = 1\}$$

- **Realization of X**

$$|X| := \sqcup_n (X_n \times |\Delta^n|) / \sim$$

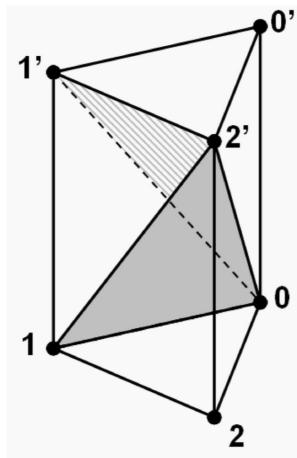
modulo equivalence relation $(x, S_i(t)) \sim (s_i(x), t)$ and $(x, D_i(t)) \sim (d_i(x), t)$

- interpret as recipe for gluing the geometric simplexes $|\Delta^n|$ together according to the combinatorial scheme prescribed by the X_n so that faces and degeneracies match



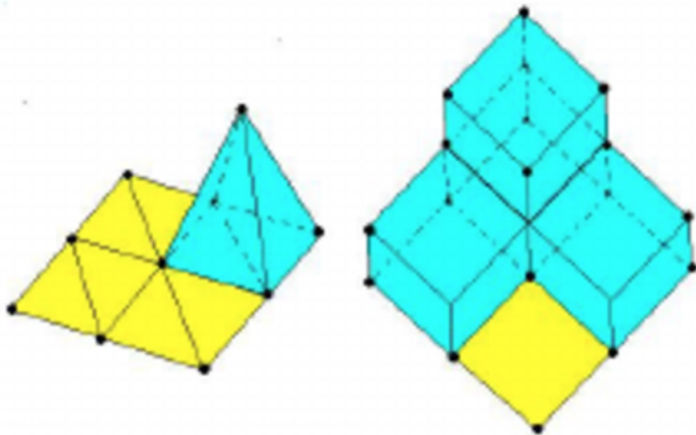
- **Nerve:** simplicial sets from categories
 - category Cat of small categories with functors as morphisms, nerve functor $\mathcal{N} : \text{Cat} \rightarrow \Delta\mathcal{S}$ to the category of simplicial sets $\Delta\mathcal{S} = \text{Func}(\Delta^{op}, \mathcal{S})$
 - for a small category \mathcal{C} the nerve $\mathcal{N}(\mathcal{C})$ has a 0-simplex (vertex) for each object of \mathcal{C} , a 1-simplex (edge) for each morphism, a 2-simplex for each composition of two morphisms, a k -simplex for every chain of k composable morphisms
 - face maps: composition of two adjacent morphisms at the i -th place of a k -chain $d_i : \mathcal{N}_k(\mathcal{C}) \rightarrow \mathcal{N}_{k-1}(\mathcal{C})$ and degeneracies are insertions of the identity morphism at an object in the chain

- **Products:** product of simplexes is not a simplex but can be decomposed as a union of simplexes



Cubes behave better than simplexes with respect to products

Simplicial and cubical complexes



- **Cubical sets in topology**

- \mathcal{I} unit interval as combinatorial structure consisting of two vertices and an edge connecting them
- $|\mathcal{I}| = [0, 1]$ geometric realization: unit interval as topological space (subspace of \mathbb{R})
- \mathcal{I}^n for the **n -cube** as combinatorial structure and $|\mathcal{I}^n| = [0, 1]^n$ its geometric realization
- \mathcal{I}^0 a single point
- **face maps** $\delta_i^a : \mathcal{I}^n \rightarrow \mathcal{I}^{n+1}$, for $a \in \{0, 1\}$ and $i = 1, \dots, n$

$$\delta_i^a(t_1, \dots, t_n) = (t_1, \dots, t_{i-1}, a, t_i, \dots, t_n)$$

- **degeneracy maps** $s_i : \mathcal{I}^n \rightarrow \mathcal{I}^{n-1}$

$$s_i(t_1, \dots, t_n) = (t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n)$$

- **cubical relations** for $i < j$

$$\delta_j^b \circ \delta_i^a = \delta_i^a \circ \delta_{j-1}^b \quad \text{and} \quad s_i \circ s_j = s_{j-1} \circ s_i$$

and relations

$$\delta_i^a \circ s_{j-1} = s_j \circ \delta_i^a \quad i < j$$

$$s_j \circ \delta_i^a = 1 \quad i = j$$

$$\delta_{i-1}^a \circ s_j = s_j \circ \delta_i^a \quad i > j$$

- **Cube category**: \mathfrak{C} has objects \mathcal{I}^n for $n \geq 0$ and morphisms generated by the face and degeneracy maps δ_i^a and s_i
- **Cubical set**: functor $C : \mathfrak{C}^{op} \rightarrow \mathcal{S}$ to the category of sets.
- notation: $C_n := C(\mathcal{I}^n)$

- variant of the cube category \mathfrak{C}_c with additional degeneracy maps $\gamma_i : \mathcal{I}^n \rightarrow \mathcal{I}^{n-1}$ called connections

$$\gamma_i(t_1, \dots, t_n) = (t_1, \dots, t_{i-1}, \max\{t_i, t_{i+1}\}, t_{i+2}, \dots, t_n)$$

satisfying relations

$$\gamma_i \gamma_j = \gamma_j \gamma_{i+1}, \quad i \leq j; \quad s_j \gamma_i = \begin{cases} \gamma_i s_{j+1} & i < j \\ s_i^2 = s_i s_{i+1} & i = j \\ \gamma_{i-1} s_j & i > j \end{cases}$$

$$\gamma_j \delta_i^a = \begin{cases} \delta_i^a \gamma_{j-1} & i < j \\ 1 & i = j, j+1, a = 0 \\ \delta_j^a s_j & i = j, j+1, a = 1 \\ \delta_{i-1}^a \gamma_j & i > j+1. \end{cases}$$

- role of degeneracy maps: maps s_i identify opposite faces of a cube, additional degeneracies γ_i identify adjacent faces

- cubical set with connection: functor $C : \mathcal{C}_c^{op} \rightarrow \mathcal{S}$ to the category of sets
- category of cubical sets has these functors as objects and natural transformations as morphisms
- so morphisms given by collection $\alpha = (\alpha_n)$ of morphisms $\alpha_n : C_n \rightarrow C'_n$ satisfying compatibilities $\alpha \circ \delta_i^a = \delta_i^a \circ \alpha$ and $\alpha \circ s_j = s_j \circ \alpha$ (and in the case with connection $\alpha \circ \gamma_i = \gamma_i \circ \alpha$)
- **cubical nerve** $\mathcal{N}_{\mathcal{C}}\mathcal{C}$ of a category \mathcal{C} is the cubical set with

$$(\mathcal{N}_{\mathcal{C}}\mathcal{C})_n = \text{Fun}(\mathcal{I}^n, \mathcal{C})$$

with \mathcal{I}^n the n -cube seen as a category with objects the vertices and morphisms generated by the 1-faces (edges), and $\text{Fun}(\mathcal{I}^n, \mathcal{C})$ is the set of functors from \mathcal{I}^n to \mathcal{C}

- when working with cubical sets with connection **homotopy equivalent to simplicial** nerve
- R. Antolini, *Geometric realisations of cubical sets with connections, and classifying spaces of categories*, Appl. Categ. Structures 10 (2002), no. 5, 481–494.

Simplexes and Probabilities

- X finite set with cardinality $N + 1 = \#X$
- Probability measure P on X : to all $x \in X$ assign $P_x \geq 0$ with normalization $\sum_{x \in X} P_x = 1$
- Set $\mathcal{P}(X)$ of all probability measures on X is a copy of the simplex Δ_N

$$\mathcal{P}(X) = \{(P_x)_{x \in X} \mid P_x \geq 0 \text{ and } \sum_{x \in X} P_x = 1\} \simeq \Delta_N,$$

where vertices of simplex Δ_N are probabilities $P = \delta^{(x)}$ with $\delta_x^{(x)} = 1$ and $\delta_y^{(x)} = 0$ for all $y \neq x$

- choose a bijection $X \simeq \{0, \dots, N\}$ then

$$\mathcal{P}(X) = \Delta_N = \{\underline{x} = (x_i)_{i=0}^N \in \mathbb{R}_+^{N+1} \mid \sum_{i=0}^N x_i = 1\}$$

Partial ordering

- transformations of probability distributions: transformations of \mathbb{R}^{N+1} preserving positivity and ℓ_1 -norm of non-negative vectors
- for $\underline{x} \in \mathbb{R}^{N+1}$ let \underline{x}^\downarrow be the vector obtained from \underline{x} by a permutation of coordinates so that in decreasing order

$$x_0^\downarrow \geq x_1^\downarrow \geq x_2^\downarrow \geq \dots \geq x_N^\downarrow$$

- for $\underline{x}, \underline{y} \in \mathbb{R}^{N+1}$ let $\underline{y} \succcurlyeq \underline{x}$ iff

$$\sum_{i=0}^k x_i^\downarrow \leq \sum_{i=0}^k y_i^\downarrow, \quad \forall 0 \leq k \leq N$$

where for probabilities the last one is

$$\sum_{i=0}^N x_i^\downarrow = \sum_{i=0}^N y_i^\downarrow = 1$$

- have $\underline{x} \succeq \underline{x}$ and if $\underline{x} \succeq \underline{y}$ and $\underline{y} \succeq \underline{z}$ then $\underline{x} \succeq \underline{z}$
- don't have $\underline{y} \succeq \underline{x}$ and $\underline{x} \succeq \underline{y}$ implies $\underline{y} = \underline{x}$ (only up to a permutation), but when written in the form \underline{y}^\downarrow and \underline{x}^\downarrow implies equal
- so get partial ordering on these
- there are pairs $\underline{x}, \underline{y}$ for which neither $\underline{x} \succeq \underline{y}$ nor $\underline{y} \succeq \underline{x}$
- there is a smallest element given by the uniform distribution

$$\underline{x}_{(N+1)} = \left(\frac{1}{N+1}, \dots, \frac{1}{N+1} \right)$$

- largest: $N + 1$ probabilities with one 1 entry and all other entries 0

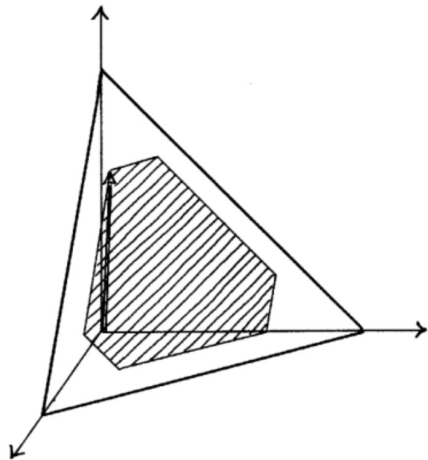
- **Shur convex functions** $\underline{x} \succeq \underline{x} \Rightarrow f(\underline{x}) \geq f(\underline{y})$
- examples: elementary symmetric functions

$$s_0(\underline{x}) = 1, \quad s_1(\underline{x}) = x_1 + \cdots + x_N,$$

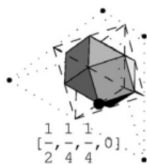
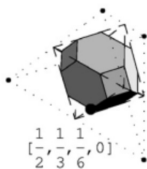
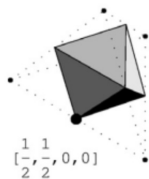
$$s_2(\underline{x}) = \sum_{i < j} x_i x_j, \quad \cdots, \quad s_N(\underline{x}) = x_1 \cdots x_N.$$

- set of vectors majorized by a given one is a convex set

$$\underline{x} \succeq \underline{y} \text{ and } \underline{x} \succeq \underline{z} \Rightarrow \underline{x} \succeq \lambda \underline{y} + (1 - \lambda) \underline{z} \quad \forall \lambda \in [0, 1]$$



convex set of probabilities \underline{x} inside Δ_3 majorized by a given vector:
pure points of the convex set permutations of entries of the vector
(from Ingemar Bengtsson, Karol Zyczkowski "The geometry of quantum states", Cambridge 2007)



examples inside Δ_4 : probabilities majorized by given vector (from Ingemar Bengtsson, Karol Zyczkowski "The geometry of quantum states", Cambridge 2007)

Stochastic Matrices

- $S = (S_{ij}) \in M_{N \times N}(\mathbb{C})$ with $S_{ij} \geq 0$ and $\sum_{i=1}^N S_{ij} = 1$
(columns are probability measures)
- $S_{ij} \geq 0$: preserves positivity; $\sum_{i=1}^N S_{ij} = 1$: preserves ℓ_1 -norm
- **bistochastic** if $\sum_{j=1}^N S_{ij} = 1$ as well
- bistochastic matrices also fix the point $\underline{x}_{(N)}$: they cause contraction of the probability simplex toward the center
- **Hardy–Littlewood–Pólya**: $\underline{x} \succeq \underline{y} \Rightarrow \exists S$ bistochastic with $\underline{y} = S\underline{x}$
- bistochastic matrices form a semigroup; only ones with a bistochastic inverse are permutation matrices
- **Birkhoff theorem**: set of $N \times N$ bistochastic matrices is a convex polytope with pure points the $N!$ permutation matrices

Perron-Frobenius Theorem (strictly positive case)

- $A = (A_{ij})$ with $A_{ij} > 0$
- \exists positive eigenvalue $\lambda_A = \rho(A)$ (spectral radius) with all other eigenvalues $|\lambda| < \lambda_A$
- eigenvalue λ_A is simple
- eigenvector $\underline{v}_A = (v_{A,i})$, with $A \underline{v}_A = \lambda_A \underline{v}_A$, entries $v_{A,i} > 0$
- same for transpose A^T : left eigenvector \underline{w}_A with $\underline{w}_A A = \lambda_A \underline{w}_A$ and with $w_{A,i} > 0$ and with

$$\underline{w}_A^T \cdot \underline{v}_A = 1$$

- these are only eigenvectors with positive entries (up to scalar multiples)
- limit

$$\lim_{k \rightarrow \infty} \frac{A^k}{\lambda_A^k} = \underline{v}_A \cdot \underline{w}_A^T.$$

- column sums

$$\min_i \sum_j A_{ij} \leq \lambda_A \leq \max_i \sum_j A_{ij}$$

Perron-Frobenius Theorem (non-negative case)

- $A = (A_{ij})$ with $A_{ij} \geq 0$
- assume **irreducible**: $\forall i, j \exists m \geq 1$ with $(A^m)_{ij} > 0$
- equivalent: associate to A a graph G_A with N vertices and oriented edge from i to j iff $A_{ij} > 0$, then A irreducible means that G_A is *strongly connected*: given any two vertices there is an oriented path from the first to the second
- **period** h_A of A is greatest common divisor of the m such that $(A^m)_{ii} > 0$ (independent of i if A irreducible)

- **Perron-Frobenius:** \exists eigenvalue $\lambda_A = \rho(A) > 0$ simple
- \exists left/right eigenvectors $\underline{w}_A, \underline{v}_A$ with positive entries
- these are the only eigenvectors that are positive (up to scalar multiples)
- there are h_A complex eigenvalues λ with $|\lambda| = \lambda_A$, each $\lambda = \lambda_A \zeta$ (ζ root of 1) and simple
- column sums

$$\min_i \sum_j A_{ij} \leq \lambda_A \leq \max_i \sum_j A_{ij}$$

Stochastic matrices

- S stochastic matrix, columns sum estimate of Perron-Frobenius theorem shows $\lambda_S = 1$
- Perron-Frobenius eigenvector is an invariant measure $S \underline{v}_S = \underline{v}_S$
- stochastic matrices as linear maps = Markov maps
- Markov chain = sequence of Markov maps

Note: can also consider stochastic $N \times M$ matrices with $S_{ij} \geq 0$ and $\sum_{j=1}^M S_{ij} = 1$ for $i = 1, \dots, N$

Categories of finite probabilities

- Category of finite probability spaces with stochastic matrices \mathcal{FP}
- objects $\text{Obj}(\mathcal{FP})$ are pairs (X, P) of a finite set X with a probability measure P
- morphisms $S \in \text{Hom}_{\mathcal{FP}}((X, P), (Y, Q))$ are stochastic $(\#Y \times \#X)$ -matrices S with
 - $S_{yx} \geq 0$, for all $x \in X, y \in Y$
 - $\sum_{y \in Y} S_{yx} = 1$ for all $x \in X$
 - the probability measures are related by $Q = S P$

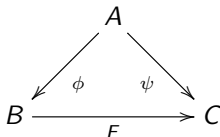
Note: objects (X, P) of \mathcal{FP} can be thought of as fuzzy sets, with P_x value at $x \in X$ of the membership function of the fuzzy set

- **Category of stochastic maps** $\mathbf{FinStoch}$ (Baez–Fritz)
- objects in $\mathbf{Obj}(\mathbf{FinStoch})$ are finite sets X
- morphisms $\mathbf{Hom}_{\mathbf{FinStoch}}(X, Y)$ are stochastic maps (assign a probability on Y to each point in X): stochastic matrices

$$S = (S_{yx}) \quad S_{yx} \geq 0, \quad \sum_{y \in Y} S_{yx} = 1, \quad \forall x \in X.$$

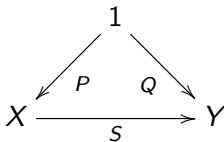
Under Category

- category \mathcal{C} and object $A \in \text{Obj}(\mathcal{C})$
- **under category** $A \downarrow \mathcal{C}$ (also denoted A/\mathcal{C}):
 - objects $\text{Obj}(A \downarrow \mathcal{C})$ are morphisms $\phi : A \rightarrow B$ of \mathcal{C} with source A
 - morphisms $F \in \text{Hom}_{A \downarrow \mathcal{C}}(\phi : A \rightarrow B, \psi : A \rightarrow C)$ are commuting triangles



Categories \mathcal{FP} and FinStoch

- \mathcal{FP} is the under category $1 \downarrow \text{FinStoch}$
- stochastic maps $P : 1 \rightarrow X$ are probability distributions on X
- so objects of $1 \downarrow \text{FinStoch}$ are pairs (X, P)
- triangles



are stochastic matrices S with $SP = Q$

Category of finite probability measures FinProb

- objects in $\text{Obj}(\text{FinProb})$ are pairs (X, P) of a finite set with a probability measure
- morphisms $f \in \text{Hom}_{\text{FinProb}}((X, P), (Y, Q))$ are functions $f : X \rightarrow Y$ that are measure preserving

$$Q_y = \sum_{x \in f^{-1}(y)} P_x.$$

- relation between FinProb and \mathcal{FP} : subcategory FinProb where for each x there is a unique $y = y(x)$ such that $S_{yx} > 0$: then stochastic matrix must have $S_{yx} = 1$, so S is a (single valued) function $f : X \rightarrow Y$ that is measure preserving

Pointed Sets and \mathcal{FP}

- category \mathcal{FS}_* of **finite pointed sets**: objects (X, x_0) finite sets with a base point; morphisms $f : (X, x_0) \rightarrow (Y, y_0)$ functions between sets $f : X \rightarrow Y$ that preserve base points, $f(x_0) = y_0$
- \mathcal{FS}_* as subcategory of \mathcal{FP} : objects (X, δ_{x_0}) and morphisms stochastic maps $S : (X, \delta_{x_0}) \rightarrow (Y, \delta_{y_0})$ with $S_{yx} = \chi_{f^{-1}(y)}(x)$, with χ the indicator function,

$$\delta_{y_0, y} = \sum_x S_{yx} \delta_{x_0, x} = \sum_{x \in f^{-1}(y)} \delta_{x_0, x}$$

Category \mathcal{S}_* of pointed sets (not necessarily finite) similarly defined

Note: can think of finite probability (X, P) as a formal convex combination of pointed sets

$$(X, P) = \sum_{x \in X} P_x (X, x)$$

Category of probabilistic pointed sets \mathcal{PS}_*

- objects in $\text{Obj}(\mathcal{PS}_*)$ are convex combinations of pointed sets

$$\Lambda X = \sum_i \lambda_i (X_i, x_i),$$

where $\Lambda = (\lambda_i)$ with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$ and $X = \{(X_i, x_i)\}$ a finite collection of pointed sets

- morphisms $\Phi \in \text{Mor}_{\mathcal{PS}_*}(\Lambda X, \Lambda' X')$ are pairs $\Phi = (S, F)$

① S is a stochastic map with $S\Lambda = \Lambda'$

② $F = (F_{ji})$ is a collection of probabilistic pointed maps
 $F_{ji} : (X_i, x_i) \rightarrow (X'_j, x'_j)$

- probabilistic pointed map F_{ji} is a finite set $\{F_{ji,a}\}$ of pointed maps $F_{ji,a} : (X_i, x_i) \rightarrow (X'_j, x'_j)$ together with a set of probabilities $\mu_a^{(ji)}$ with $\sum_a \mu_a^{(ji)} = S_{ji}$

- **composition of morphisms** $\Phi = (S, F) : \Lambda X \rightarrow \Lambda' X'$ and $\Phi' = (S', F') : \Lambda' X' \rightarrow \Sigma Y$ given by

$$\Phi' \circ \Phi = (S' \circ S, F' \circ F)$$

- $S' \circ S$ product of stochastic matrices
- $F' \circ F = \{(F' \circ F)_{ki}\}$ with set $(F' \circ F)_{ki} = \{F'_{kj,a} \circ F_{ji,b}\}$ and probabilities $\mu_a^{(kj)} \mu_b^{(ji)}$
- with $\sum_{a,b,j} \mu_a^{(kj)} \mu_b^{(ji)} = \sum_j S'_{kj} S_{ji} = (S' \circ S)_{ki}$
- probability associated set $(F' \circ F)_{ki}$ in $F' \circ F$ is $(S' \circ S)_{ki}$

Embedding of category \mathcal{FP} in \mathcal{PS}_*

- mapping $\Lambda = (\lambda_i)$ to the set $\Lambda_\star = \sum_i \lambda_i(\{\star_i\}, \star_i)$
- morphisms $S\Lambda = \Lambda'$ to $\Phi = (S, \mathbf{1})$ with $\mathbf{1} = \{1_{ji}\}$ with probabilities S_{ji}
- **forgetful functor** from \mathcal{PS}_* to \mathcal{FP} maps ΛX to finite probability Λ and morphism $\Phi = (S, F)$ to stochastic matrix S

Probabilistic categories \mathcal{PC}

- from pointed sets \mathcal{S}_* to probabilistic pointed sets \mathcal{PS}_*
- more general similar procedure starting from a given category \mathcal{C}
- \mathcal{PC} category with objects in $\text{Obj}(\mathcal{PC})$ given by formal finite convex combinations

$$\Lambda C = \sum_i \lambda_i C_i,$$

with $\Lambda = (\lambda_i)$ with $\sum_i \lambda_i = 1$ and $C_i \in \text{Obj}(\mathcal{C})$

- morphisms $\Phi : \Lambda C \rightarrow \Lambda' C'$, $\Phi \in \text{Hom}_{\mathcal{PC}}(\Lambda C, \Lambda' C')$ given by pairs $\Phi = (S, F)$
 - S a stochastic matrix with $S\Lambda = \Lambda'$
 - $F = \{F_{ab,r}\}$ finite collection of morphisms $F_{ab,r} : C_b \rightarrow C'_a$ with assigned probabilities μ_r^{ab}
 - probabilities satisfy $\sum_r \mu_r^{ab} = S_{ab}$

interpret F as mapping of C_a to C'_b by choosing randomly in the set $\{F_{ab,r}\}$ with probability μ_r^{ab} of choosing $F_{ab,r}$

Wreath product interpretation

- construction of \mathcal{PC} from \mathcal{C} can be seen as a wreath product $\mathcal{FP} \wr \mathcal{C}$ of the category \mathcal{C} with the category \mathcal{FP} of finite probabilities
- it has the effect of rendering the category \mathcal{C} probabilistic (objects are a mixture of objects and morphisms are randomly chosen with a stochastic map of the mixtures)

Zero Object

- Zero object $0 \in \text{Obj}(\mathcal{C})$: for all $X \in \text{Obj}(\mathcal{C})$ there is a unique morphism $0 \rightarrow X$ and a unique morphism $X \rightarrow 0$
- Zero objects are unique up to unique isomorphism
- points \star are zero objects in the category \mathcal{S}_* of pointed sets
- singletons $(\star, 1)$ are zero objects in the category \mathcal{FP}
 - unique morphism $\hat{Q} : (\{x\}, 1) \rightarrow (Y, Q)$ given by $\hat{Q}_{yx} = Q_y$
 - unique morphism $\hat{1} : (Y, Q) \rightarrow (\{x\}, 1)$ given by $\hat{1}_{xy} = 1$ for all $y \in Y$

$$1 = \sum_{y \in Y} Q_y = \sum_{y \in Y} \hat{1}_{xy} Q_y$$

- singletons $X = (\{x\}, x)$ with $\Lambda = 1$ are zero objects in \mathcal{PS}_*
 - unique morphism $\Phi = (S, F) : (\{x\}, x) \rightarrow \sum_i \lambda_i (X_i, x_i)$ with $S = \hat{\Lambda}$ (unique morphism in \mathcal{FP} from $(\{x\}, 1)$ to Λ) and $F = (F_i)$ with $F_i : x \mapsto x_i$ with probability λ_i
 - unique morphism $\Phi = (S, F) : \Lambda X \rightarrow (\{x\}, x)$ where $S = \hat{1}_\Lambda$ (unique morphism in \mathcal{FP} from Λ to $(\{x\}, 1)$) and $F = (F_i)$ with $F_i : (X_i, x_i) \rightarrow (\{x\}, x)$ the constant function with probability 1
- same argument: if \mathcal{C} has a zero object then \mathcal{PC} has a zero object given by the zero object of \mathcal{C} with $\Lambda = 1$

Categorical Sum (Coproduct)

- category of pointed sets \mathcal{S}_* : coproduct of (X, x) and (X', x')

$$(X, x) \vee (X', x') := (X \sqcup X' / x \sim x', x \sim x')$$

- universal property of coproduct: \exists morphisms $\iota_X, \iota_{X'}$ such that for any given morphisms $f : (X, x) \rightarrow (Y, y)$ and $f' : (X', x') \rightarrow (Y, y)$

$$\begin{array}{ccccc} & & (Y, y) & & \\ & \nearrow f & \uparrow \exists! f \vee f' & \nwarrow f' & \\ (X, x) & \xrightarrow{\iota_X} & (X, x) \vee (X', x') & \xleftarrow{\iota_{X'}} & (X', x') \end{array}$$

- category of probabilistic pointed sets \mathcal{PS}_* : coproduct of $\Lambda X = \sum_{i=1}^N \lambda_i(X_i, x_i)$ and $\Lambda' X' = \sum_{j=1}^M \lambda'_j(X'_j, x'_j)$

$$\Lambda X \amalg \Lambda' X' := \sum_{ij} \lambda_i \lambda'_j (X_i, x_i) \vee (X'_j, x'_j)$$

with coproducts $(X_i, x_i) \vee (X'_j, x'_j)$ of pointed sets

- induces coproduct on \mathcal{FP}

$$(X, P) \amalg (X', P') = (X \times X', P \cdot P')$$

statistically independent probabilities

universal property of coproduct in \mathcal{PS}_*

- morphisms $\Phi = (S, F) : \Lambda X \rightarrow \Sigma Y$ and $\Phi' = (S', F') : \Lambda' X' \rightarrow \Sigma Y$ with $\Sigma Y = \sum_k \sigma_k (Y_k, y_k)$
- $S\Lambda = \Sigma$ and $S'\Lambda' = \Sigma$
- $F = \{f_{ka,r}\}_{r=1}^N$ with probabilities $\sum_r \mu_r^{(ka)} = S_{ka}$ and $F' = \{f'_{ka',r'}\}_{r'=1}^M$ with probabilities $\sum_{r'} \mu_{r'}^{(ka')} = S'_{ka'}$
- construct $F \vee F'$ as set $\{f_{ka,r} \vee f'_{ka',r'}\}$ of pointed maps from coproducts of pointed sets with probabilities $\sigma_k^{-1} \mu_r^{(ka)} \mu_{r'}^{(ka')}$ for $\sigma_k \neq 0$ and $M^{-1} \mu_r^{(ka)} + N^{-1} \mu_{r'}^{(ka')}$ for $\sigma_k = 0$
- then set $(\Phi \amalg_\lambda \Phi') = (S \amalg S', F \vee F')$ with

$$(S \amalg_\lambda S')_{k,(a,a')} = \begin{cases} \sigma_k^{-1} \cdot S_{k,a} \cdot S'_{k,a'} & \sigma_k \neq 0 \\ S_{k,a} + S'_{k,a'} & \sigma_k = 0. \end{cases}$$

- to see universality holds, need to construct morphisms $\Psi : \Lambda X \rightarrow \Lambda X \amalg \Lambda' X'$ and $\Psi' : \Lambda' X' \rightarrow \Lambda X \amalg \Lambda' X'$, such that, for any Φ, Φ' morphisms to ΣY diagram commutes

$$\begin{array}{ccccc}
 & & \Sigma Y & & \\
 & \nearrow \Phi & \uparrow \Phi \amalg \Phi' & \nwarrow \Phi' & \\
 \Lambda X & \xrightarrow{\Psi} & \Lambda X \amalg \Lambda' X' & \xleftarrow{\Psi'} & \Lambda' X' .
 \end{array}$$

- take $\Psi = (\mathcal{I}, \mathcal{F})$ and $\Psi' = (\mathcal{I}', \mathcal{F}')$ with morphisms $\mathcal{I} \in \text{Mor}_{\mathcal{FP}}(\Lambda, \Lambda \cdot \Lambda')$, $\mathcal{I}' \in \text{Mor}_{\mathcal{FP}}(\Lambda', \Lambda \cdot \Lambda')$ given by

$$(\mathcal{I})_{(b,b'),a} = \delta_{ab} \lambda'_{b'} \quad \text{and} \quad (\mathcal{I}')_{(b,b'),a'} = \delta_{a'b'} \lambda_b.$$

- these satisfy $\mathcal{I}\Lambda = \Lambda \cdot \Lambda'$ and $\mathcal{I}'\Lambda' = \Lambda \cdot \Lambda'$
- probabilistic pointed maps $\mathcal{F} = (\mathcal{F}_{(b,b'),a} = \delta_{ab} \mathcal{F}_{b'b})$, $\mathcal{F}' = (\mathcal{F}'_{(b,b'),a'} = \delta_{a'b'} \mathcal{F}'_{bb'})$ inclusions of pointed sets

$$(X_b, x_b) \xrightarrow{\mathcal{F}_{b'b}} (X_b, x_b) \vee (X'_{b'}, x'_{b'}), \quad (X'_{b'}, x'_{b'}) \xrightarrow{\mathcal{F}'_{bb'}} (X_b, x_b) \vee (X'_{b'}, x'_{b'})$$

respectively chosen with probability $\lambda'_{b'}$ and λ_b

Coproduct in probabilistic categories \mathcal{PC}

- similarly if \mathcal{C} has coproduct then probabilistic category \mathcal{PC} also has coproduct

$$\Lambda C \amalg \Lambda' C' = \sum_{i,j} \lambda_i \lambda'_j C_i \amalg_{\mathcal{C}} C'_j$$

- it satisfies universal property: same argument as for \mathcal{PS}_*

Statistical independence: product or coproduct?

- **categorical product**: an object $X_1 \times X_2$ such that, for all morphisms $f_1 : X \rightarrow X_1$ and $f_2 : X \rightarrow X_2$ $\exists!$ morphism $h : X \rightarrow X_1 \times X_2$ with commuting

$$\begin{array}{ccccc} & & X & & \\ & \swarrow f_1 & \downarrow h & \searrow f_2 & \\ X_1 & \xleftarrow{\pi_1} & X_1 \times X_2 & \xrightarrow{\pi_2} & X_2 \end{array}$$

- category \mathcal{FP} does *not* have a universal categorical product
- ...*but* **tensor category with projections**: (\mathcal{C}, \otimes) with two natural transformations $\pi_i : \otimes \rightarrow \Pi_i$ with $\Pi_i(X_1, X_2) = X_i$ and for any morphisms $f_i : Y_i \rightarrow X_i$ commuting

$$\begin{array}{ccccc} Y_1 & \xleftarrow{\pi_{Y_1}} & Y_1 \otimes Y_2 & \xrightarrow{\pi_{Y_2}} & Y_2 \\ \downarrow f_1 & & \downarrow f_1 \otimes f_2 & & \downarrow f_2 \\ X_1 & \xleftarrow{\pi_{X_1}} & X_1 \otimes X_2 & \xrightarrow{\pi_{X_2}} & X_2 \end{array}$$

- in tensor category with projections two morphisms $f_i : X \rightarrow X_i$ are *independent* if there exists a morphism $h : X \rightarrow X_1 \otimes X_2$ such that product diagram commutes
- in \mathcal{FP} this notion of independence agrees with statistical independence

$$(X_1, P_1) \otimes (X_2, P_2) = (X_1 \times X_2, P_1 P_2)$$

- in previous construction of coproduct shown that product of statistically independent measures can also be interpreted as *coproduct* in the category \mathcal{FP}