Information Geometry of Polytopes

general theory and applications to stochastic matrices

Nihat Ay

Max Planck Institute for Mathematics in the Sciences
Leipzig, Germany

Santa Fe Institute, New Mexico, USA

7. April 2010

Notation

Simplex of probability measures

$$\mathbb{R}^{\mathscr{X}} \;:=\; \text{real valued functions on a finite set } \mathscr{X}$$

$$\delta_x \in \left(\mathbb{R}^{\mathscr{X}}\right)^*, \quad x \in \mathscr{X}, \qquad \text{(dual of the canonical basis)}$$

$$\mathcal{P} \;:=\; \left\{ \sum_x p_x\, \delta_x \in \left(\mathbb{R}^{\mathscr{X}}\right)^* \;:\; p_x > 0, \quad \sum_x p_x = 1 \right\}$$

Quotient vector space

$$\mathcal{C} \;:=\; \text{subspace of constant functions}$$

$$\mathbb{R}^{\mathscr{X}}/\mathcal{C} \;=\; \left\{ f + \mathcal{C} \;:\; f \in \mathbb{R}^{\mathscr{X}} \right\}$$

$\mathcal{P}$ as a Manifold

Tangent and cotangent space of $\mathcal{P}$

- Tangent space of $\mathcal{P}$:

$$\mathcal{T} \;:=\; \left\{\varphi \in \left(\mathbb{R}^{\mathscr{X}}\right)^* \;:\; \varphi(f) = 0 \text{ for all } f \in \mathcal{C}\right\}$$

- Cotangent space of $\mathcal{P}$:

$$\mathbb{R}^{\mathscr{X}}/\mathcal{C} \;\cong\; \mathcal{T}^*$$

$$\mathbb{R}^{\mathscr{X}}/\mathcal{C} \;\to\; \mathcal{T}^*, \qquad f + \mathcal{C} \;\mapsto\; \left(\varphi \;\mapsto\; \varphi(f)\right)$$

Remark

$$\mathcal{U} \;\subseteq\; \mathcal{V}, \qquad \mathcal{U}^0 := \{\varphi \in \mathcal{V}^* \;:\; \varphi(f) = 0 \text{ for all } f \in \mathcal{U}\}$$

$$\Rightarrow \qquad \mathcal{V}^*/\mathcal{U}^0 \;\to\; \mathcal{U}^*, \quad \varphi + \mathcal{U}^0 \;\mapsto\; \varphi|_{\mathcal{U}}, \quad \text{natural isomorphism}$$

$\mathcal{P}$ as a Riemannian Manifold

- Scalar product on $\mathbb{R}^{\mathscr{X}}/\mathcal{C}$:

$$\langle f + \mathcal{C}, g + \mathcal{C} \rangle_p \; := \; \mathrm{cov}(f, g) \; = \; \langle f \cdot g \rangle_p - \langle f \rangle_p \cdot \langle g \rangle_p$$

- Identification of $\mathbb{R}^{\mathscr{X}}/\mathcal{C}$ and $\mathcal{T}$ with respect to $\langle \cdot, \cdot \rangle_p$:

$$\phi_p : \; \mathbb{R}^{\mathscr{X}}/\mathcal{C} \; \to \; \mathcal{T}, \qquad f + \mathcal{C} \; \mapsto \; \sum_x p_x \left(f_x - \langle f \rangle_p \right) \delta_x$$

- Natural bundle isomorphism:

$$T^*\mathcal{P} = \mathcal{P} \times (\mathbb{R}^{\mathscr{X}}/\mathcal{C}) \; \stackrel{\phi}{\longrightarrow} \; \mathcal{P} \times \mathcal{T} = T\mathcal{P}$$

- Fisher metric, Shahshahani inner product:

$$\langle \cdot, \cdot \rangle_p : \mathcal{T} \times \mathcal{T} \to \mathbb{R}, \qquad \langle A, B \rangle_p \; := \; \sum_x \frac{1}{p_x} A_x B_x$$

Vector Fields and Differential Equations

Replicator equations

Consider a map $f : \mathcal{P} \to \mathbb{R}^{\mathscr{X}}$ and the induced section

$$\mathcal{P} \to \mathbb{R}^{\mathscr{X}}/\mathcal{C}, \qquad p \mapsto f(p) + \mathcal{C}.$$

in $T^*\mathcal{P}$. This can be identified with the vector field

$$\mathcal{P} \to \mathcal{T}, \qquad p \mapsto \sum_x p_x \left(f_x(p) - \langle f(p) \rangle_p \right) \delta_x$$

and the corresponding differential equations

$$\dot{p}_x = p_x \Big(f_x(p) - \langle f(p) \rangle_p \Big), \qquad x \in \mathscr{X}.$$

These equations are known as *replicator equations*.

Observation

Replicator equations are nothing but ordinary differential equations expressed in terms of a section in the cotangent bundle $T^*\mathcal{P}$ instead of a section in the tangent bundle $T\mathcal{P}$.

Gradient fields

Consider a function $F : \mathcal{P} \to \mathbb{R}$. Then the gradient of F with respect to the Fisher metric is given by

$$(\mathrm{grad}_p F)_x \;=\; p_x \left(\partial_x F(p) - \sum_{x'} p_{x'} \, \partial_{x'} F(p) \right)$$

- J. Hofbauer & K. Sigmund. *Evolutionary Games and Population Dynamics.* Cambridge University Press 2002.
- N. Ay & I. Erb. *On a Notion of Linear Replicator Equations.* Journal of Dynamics and Differential Equations 2005.

Example: Darwinian selection, "survival of the fittest"

Consider a function $f : \mathscr{X} \to \mathbb{R}$ (f_x is called *fitness of species x*) and

$$\dot{p}_x \;=\; p_x\Big(f_x - \langle f\rangle_p\Big) \;=\; \mathrm{grad}_p\langle f\rangle, \qquad p(0) \;=\; p.$$

Solution curves satisfy the Price-equation and Fisher's fundamental theorem of natural selection:

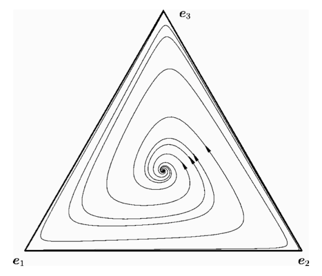
$$\frac{d}{dt}\langle g\rangle_{p(t)} \;=\; \mathrm{cov}(f,g) \qquad\qquad \frac{d}{dt}\langle f\rangle_{p(t)} \;=\; \mathrm{var}(f)$$

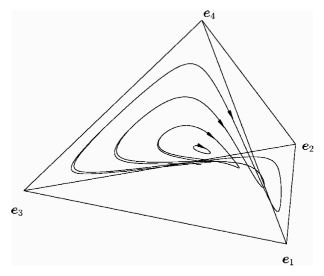
Solution curves: $t \;\mapsto\; p(t) = \frac{p\,e^{tf}}{\left\langle e^{tf}\right\rangle_p}$

$$\lim_{t\to-\infty} p(t) \;=\; \begin{cases} \frac{p_x}{\sum_{x\in\mathrm{argmin}(f)} p_x}, & \text{falls } x \in \mathrm{argmin}(f) \\ 0, & \text{sonst} \end{cases}$$

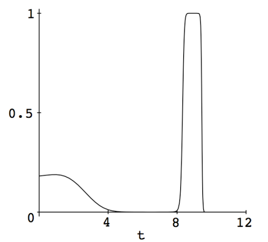
$$\lim_{t\to\infty} p(t) \;=\; \begin{cases} \frac{p_x}{\sum_{x\in\mathrm{argmax}(f)} p_x}, & \text{falls } x \in \mathrm{argmax}(f) \\ 0, & \text{sonst} \end{cases}$$

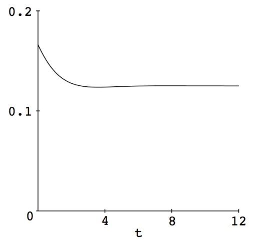
Hypercycle dynamics (Eigen und Schuster) with three and four species

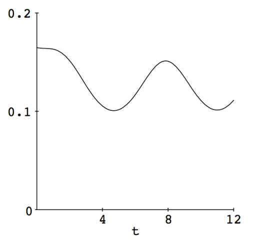




Hypercycle dynamics with eight species for different initial conditions







Natural Connections

Mixture Connection

$$\overset{m}{\Pi}_{p,q}\colon\ T_p\mathcal{P} \longrightarrow T_q\mathcal{P}, \qquad (p, A) \longmapsto (q, A)$$

Exponential Connection

$$\overset{e}{\Pi}_{p,q}\colon\ T_p^*\mathcal{P} \longrightarrow T_q^*\mathcal{P}, \qquad (p, f + \mathcal{C}) \longmapsto (q, f + \mathcal{C})$$

$$\overset{e}{\Pi}_{p,q}\colon\ T_p\mathcal{P} \longrightarrow T_q\mathcal{P}, \qquad (p, A) \longmapsto (q, (\phi_q \circ {\phi_p}^{-1})(A))$$

Duality

$$\left\langle \overset{e}{\Pi}_{p,q} A, \overset{m}{\Pi}_{p,q} B \right\rangle_q = \langle A, B \rangle_p$$

Differential version

$$\overset{m,e}{\nabla}_A B\Big|_p \;:=\; \lim_{t\to 0}\frac{1}{t}\left(\overset{m,e}{\Pi}_{\gamma(t),p}\left(B_{\gamma(t)}\right) - B_p\right) \;\in\; T_p\mathcal{P}.$$

Geodesics satisfying $\overset{m,e}{\gamma}(0) = p$ and $\overset{m,e}{\gamma}(1) = q$

1. *m*-geodesics:

$$\overset{m}{\gamma}_{p,q}\colon\; [0,1] \to \mathcal{P}, \qquad t \mapsto (1-t)\,p + t\,q$$

2. *e*-geodesics:

$$\overset{e}{\gamma}_{p,q}\colon\; [0,1] \to \mathcal{P}, \qquad t \mapsto \frac{p^{1-t}q^{t}}{\sum_x p_x^{1-t}q_x^{t}}$$

Corresponding exponential maps

1. *m*-exponential map:

$$\overset{m}{\exp}: \ \left\{(p, q-p) \in T\mathcal{P} \ : \ p, q \in \mathcal{P}\right\} \ \to \ \mathcal{P}, \qquad (p, A) \ \mapsto \ p + A,$$

2. *e*-exponential map:

$$\overset{e}{\exp}: \ T\mathcal{P} \ \to \ \mathcal{P}, \qquad (p, A) \ \mapsto \ \frac{p\, e^{\frac{A}{p}}}{\sum_x p_x\, e^{\frac{A_x}{p_x}}}$$

Relative entropy

$$D(p \,\|\, q) \ = \ \begin{cases} \sum_x p_x \ln \frac{p_x}{q_x}, & \text{if } \operatorname{supp}(p) \subseteq \operatorname{supp}(q) \\ +\infty \quad , & \text{otherwise} \end{cases}$$

$$\overset{m}{\exp}_q{}^{-1}(p) \ = \ -\operatorname{grad}_q D(p \,\|\, \cdot) \qquad \overset{e}{\exp}_q{}^{-1}(p) \ = \ -\operatorname{grad}_q D(\cdot \,\|\, p)$$

Motivation of the previous relations

- Squared distance from q:

$$D_q : \mathbb{R}^n \to \mathbb{R}, \qquad p \mapsto \frac{1}{2}\|q-p\|^2 = \frac{1}{2}\sum_{i=1}^{n}(q_i - p_i)^2$$

- The differential $d_p D_q \in (\mathbb{R}^n)^*$:

$$d_p D_q : \mathbb{R}^n \to \mathbb{R}, \qquad v \mapsto \frac{\partial D_q}{\partial v}(p)$$

- Identification of the differential with a vector in $\mathbb{R}^n$:

$$-\mathrm{grad}_p\, D_q = -\mathrm{grad}\big(d_p D_q\big) = q - p = \overrightarrow{p\,q}$$

$\mathcal{P}$ as an Affine Space

Note that $\overset{e}{\exp}$ is not an affine action:

$$\overset{e}{\exp}\left(\overset{e}{\exp}(p, A), B\right) \neq \overset{e}{\exp}(p, A + B)$$

Consider instead the composition

$$\mathrm{t}: \; \mathcal{P} \times (\mathbb{R}^{\mathscr{X}}/\mathcal{C}) = T^*\mathcal{P} \;\; \xrightarrow{\phi} \;\; T\mathcal{P} \xrightarrow{\overset{e}{\exp}} \mathcal{P},$$

$$(p, f + \mathcal{C}) \;\mapsto\; \frac{p\, e^f}{\langle e^f \rangle_p}$$

This is an affine action with difference vector

$$\mathrm{vec}: \mathcal{P} \times \mathcal{P} \to \mathbb{R}^{\mathscr{X}}/\mathcal{C}, \qquad (p, q) \mapsto \mathrm{vec}(p, q) = \ln\left(\frac{q}{p}\right) + \mathcal{C}$$

and $\mathcal{P}$ is an affine space over $\mathbb{R}^{\mathscr{X}}/\mathcal{C}$.

Definition: Exponential families

An affine subspace $\mathcal{E}$ of $\mathcal{P}$ with respect to t is called *exponential family*. Given a probability measure q and a linear subspace $\mathcal{L}$ of $\mathbb{R}^{\mathscr{X}}$, the following submanifold of $\mathcal{P}$ is an exponential family:

$$\mathcal{E}(q, \mathcal{L}) \quad := \quad \left\{ \frac{q\, e^f}{\langle e^f \rangle_q} \; : \; f \in \mathcal{L} \right\}$$

Clearly, all exponential families are of this structure. We always assume $\mathcal{C} \subseteq \mathcal{L}$ and thereby ensure uniqueness of $\mathcal{L}$. Furthermore, with this assumption we have $\dim(\mathcal{E}) = \dim(\mathcal{L}) - 1$.

Remark

An exponential family can be identified with a polytope via the expectation value map.

Extension to Polytopes
work in progress with Johannes Rauh

Consider a polytope $\bar{\mathcal{C}} \subset \mathbb{R}^d$, and its set $\mathrm{ext}(\bar{\mathcal{C}})$ of extreme points.

$$\bar{\mathcal{C}} \;=\; \left\{ \sum_{x \in \mathrm{ext}(\bar{\mathcal{C}})} p(x)\, x \;\in\; \mathbb{R}^d \;:\; p \in \bar{\mathcal{P}}(\mathrm{ext}(\bar{\mathcal{C}})) \right\}$$

$$\varphi : \bar{\mathcal{P}}(\mathrm{ext}(\bar{\mathcal{C}})) \;\to\; \bar{\mathcal{C}}, \qquad p \;\mapsto\; \varphi(p) \;:=\; \sum_{x \in \mathrm{ext}(\bar{\mathcal{C}})} p(x)\, x$$

The set $\varphi^{-1}(\{\kappa\})$ is a convex set. Choose that member p_κ that has maximal entropy. This corresponds to the canonical representation.

Information geometry of polytopes, the main idea

The image of the map $\kappa \mapsto p_\kappa$ is the closure of an exponential family $\mathcal{E} = \mathcal{E}(\mathcal{C})$. Take the push-forward of the induced geometry $(\mathcal{E}, g|_{\mathcal{E}}, \overset{m}{\nabla}|_{\mathcal{E}}, \overset{e}{\nabla}|_{\mathcal{E}})$. In particular,

$$D(\kappa \,\|\, \sigma) \;:=\; D(p_\kappa \,\|\, p_\sigma).$$

Application of this idea to the setting of Markov kernels

Polytope of Markov kernels

$$\bar{\mathcal{C}} \;:=\; \bar{\mathcal{C}}(\mathscr{X};\mathscr{Y}) \;=\; \left\{\kappa \in \mathbb{R}^{\mathscr{X}\times\mathscr{Y}} \;:\; \kappa(x;y) \geq 0, \quad \sum_y \kappa(x;y) = 1\right\}$$

$$f \in \mathscr{Y}^{\mathscr{X}}, \quad \kappa_f(x;y) \;:=\; \delta_{f(x)}(y), \qquad \mathrm{ext}(\bar{\mathcal{C}}) = \left\{\kappa_f \;:\; f \in \mathscr{Y}^{\mathscr{X}}\right\}$$

$$\varphi : \bar{\mathcal{P}}(\mathscr{Y}^{\mathscr{X}}) \;\to\; \bar{\mathcal{C}}, \qquad p \;\mapsto\; \varphi(p) \;:=\; \sum_f p(f)\,\kappa_f$$

Corresponding exponential family

$$\mathcal{E} = \mathcal{E}(\mathcal{C}) = \frac{e^{\sum_{x,y} \theta_{x,y} \kappa_{\cdot}(x;y)}}{\sum_{f \in \mathscr{Y}^{\mathscr{X}}} e^{\sum_{x,y} \theta_{x,y} \kappa_f(x;y)}}$$

Proposition

The restriction $\varphi|_{\bar{\mathcal{E}}}$ *has the inverse*

$$\varphi^{-1} : \bar{\mathcal{C}} \to \bar{\mathcal{E}}, \qquad \kappa \mapsto p_\kappa := \bigotimes_{x \in \mathscr{X}} \kappa(x;\cdot).$$

Example: Structural equation model and Markov kernels

Structural equation model

Consider $\mathscr{Z}$ and $F : \mathscr{X} \times \mathscr{Z} \to \mathscr{Y}$

$$y \;=\; F(x,z), \quad z \text{ disturbance with distribution } p$$

$$\leadsto \qquad \kappa(x;y) \;:=\; \sum_{z \in \mathscr{Z}} p(z)\, \delta_{F(x,z)}(y) \;=\; \sum_{z \in \mathscr{Z}} p(z)\, \delta_{F_z(x)}(y)$$

$$\Rightarrow \qquad \kappa(x;\cdot) \;\in\; \bar{\mathcal{P}}(\mathscr{Y}) \quad \text{for all } x$$

Canonical representation

Given κ, choose $\mathscr{Z} := \mathscr{Y}^{\mathscr{X}}$,

$$F : \mathscr{X} \times \mathscr{Y}^{\mathscr{X}} \;\to\; \mathscr{Y}, \qquad (x,f) \;\mapsto\; f(x), \qquad \text{and}$$

$$p_\kappa(f) \;:=\; \prod_{x \in \mathscr{X}} \kappa(x; f(x))$$

Fisher metric

$$g_\kappa(A, B) \;:=\; \sum_{x,y} \frac{1}{\kappa(x;y)} A(x;y)\, B(x;y)$$

m-Geodesics

$$\gamma^{(m)}(t) \;:=\; (1-t)\,\kappa(x;y) + t\,\sigma(x;y)$$

e-Geodesics

$$\gamma^{(e)}(t) \;:=\; \frac{\kappa(x;y)^{1-t}\sigma(x;y)^t}{\sum_{y'} \kappa(x;y')^{1-t}\sigma(x;y')^t}$$

Relative entropy

$$D(\kappa \,\|\, \sigma) \;:=\; D(p_\kappa \,\|\, p_\sigma) \;=\; \sum_{x,y} \kappa(x;y) \ln \frac{\kappa(x;y)}{\sigma(x;y)}$$

Affine action

$f, g \in \mathbb{R}^{\mathscr{X} \times \mathscr{Y}}$:

$$f \sim g \quad :\Leftrightarrow \quad f - g \in \mathbb{R}^{\mathscr{X}}$$

$$\mathcal{V} := \mathbb{R}^{\mathscr{X} \times \mathscr{Y}} / \mathbb{R}^{\mathscr{X}}$$

$$\mathcal{C} \times \mathcal{V} \to \mathcal{C}, \qquad (\kappa, f + \mathbb{R}^{\mathscr{X}}) \mapsto \frac{\kappa\, e^{f}}{\sum_y \kappa(\cdot; y)\, e^{f(\cdot;y)}}$$

Proposition

The affine subspaces of $\mathcal{C}(\mathscr{X};\mathscr{Y})$ are the exponential families in $\mathcal{C}(\mathscr{X};\mathscr{Y})$. Given a subspace $\mathcal{L}$ of dimension $|\mathscr{X}| + d$ with $\mathbb{R}^{\mathscr{X}} \subseteq \mathcal{L} \subseteq \mathbb{R}^{\mathscr{X} \times \mathscr{Y}}$, and given a reference element κ, the following subfamily is a d-dimensional exponential family:

$$\mathcal{E}(\kappa, \mathcal{L}) := \left\{ \frac{\kappa\, e^{f}}{\sum_y \kappa(\cdot; y)\, e^{f(\cdot;y)}} \; : \; f \in \mathcal{L} \right\}$$

Maximization of Entropy Distance from Exponential Families

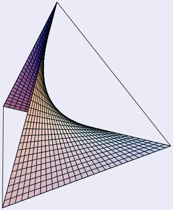
based on Ay 2002, Matúš & Ay 2003

Multi-Information as an Example

Family of product distributions

$$\otimes : \ \bar{\mathcal{P}}(\mathscr{X}_1) \times \cdots \times \bar{\mathcal{P}}(\mathscr{X}_N) \ \rightarrow \ \bar{\mathcal{P}}(\mathscr{X}_1 \times \cdots \times \mathscr{X}_N)$$

$$(p_1, \ldots, p_N) \ \mapsto \ p_1 \otimes \cdots \otimes p_N$$



Definition of multi-information

$$I : \ \bar{\mathcal{P}}(\mathscr{X}_1 \times \cdots \times \mathscr{X}_N) \rightarrow \mathbb{R}, \qquad p \mapsto I(p) := \inf_{q \in \mathrm{im}(\otimes)} D(p \,\|\, q)$$

General problem

Given an exponential family $\mathcal{E}$, maximize the function

$$D_{\mathcal{E}} : \bar{\mathcal{P}} \to \mathbb{R}, \qquad p \mapsto \inf_{q \in \mathcal{E}} D(p \,\|\, q).$$

Results

1. Support bound: Let p be a local maximizer of $D_{\mathcal{E}}$. Then

$$|\text{supp}(p)| \leq \dim(\mathcal{E}) + 1$$

   In particular, $H(p) \leq \ln(\dim(\mathcal{E}) + 1)$.

2. Extended exponential families: There exists an exponential family $\widetilde{\mathcal{E}}$ of dimension

$$\dim\left(\widetilde{\mathcal{E}}\right) \leq 3\dim(\mathcal{E}) + 2$$

   that contains $\mathcal{E}$ and all local maximizers of $D_{\mathcal{E}}$ in its closure.

Applied to multi-information

1. Let p be a local maximizer of the multi-information. Then

$$|\mathrm{supp}(p)| \;\leq\; \sum_{i=1}^{N}(|\mathscr{X}_i| - 1) + 1 \quad (= \; N + 1 \;\ll\; 2^N \text{ for binary nodes})$$

In particular

$$H(p) \;\leq\; \ln\left(\sum_{i=1}^{N}(|\mathscr{X}_i| - 1) + 1\right).$$

2. There exists an exponential family of dimension at most

$$3\sum_{i=1}^{N}(|\mathscr{X}_i| - 1) \;+\; 2 \qquad (3\,N \;+\; 2 \;\ll\; 2^N - 1 \text{ for binary nodes})$$

that contains all local maximizers of the multi-information in its closure.

Multi-Information for Kernels

Factorized kernels

Consider a set $[N]$ of nodes with input symbols $\mathscr{X}_i$ and output symbols $\mathscr{Y}_i$ and denote with $\mathscr{X}$ and $\mathscr{Y}$ the products.

$$\otimes : \quad \prod_{i \in [N]} \bar{\mathcal{C}}(\mathscr{X}_i; \mathscr{Y}_i) \;\hookrightarrow\; \bar{\mathcal{C}}(\mathscr{X}; \mathscr{Y})$$

$$(\kappa_i)_i \;\mapsto\; (\otimes_i \kappa_i)(x; y) := \kappa_1(x_1; y_1) \cdots \kappa_N(x_N; y_N)$$

This is the closure of an exponential family in $\mathcal{C}(\mathscr{X}; \mathscr{Y})$ of dimension

$$\sum_{i=1}^{N} |\mathscr{X}_i|(|\mathscr{Y}_i| - 1)$$

Multi-information

$$I(\kappa) \;:=\; \inf_{\sigma \in \mathrm{im}(\otimes)} D(\kappa \,\|\, \sigma)$$

Local maximizers

Let κ be a local maximizer of I. Then the dimension of the face $F(\kappa)$ of $\bar{\mathcal{C}}(\mathscr{X};\mathscr{Y})$ in which κ is contained is upper bounded by

$$d \;:=\; \sum_{i=1}^{N} |\mathscr{X}_i|(|\mathscr{Y}_i| - 1) \qquad (d = 2\,N \text{ for binary nodes})$$

The theorem of Carathéodory implies there are at most $d+1$ functions so that κ can be written as their convex combination:

$$\kappa \;=\; \sum_{f} p(f)\,\kappa_f$$

This implies

$$|\mathrm{supp}\,\kappa(x;\cdot)| \;\leq\; d+1$$

and therefore $H(Y\,|\,X) \leq \ln(d+1)$ for any distribution of X.

Low-dimesional exponential families

There exists an exponential family $\mathcal{E}$ of dimension at most $3\,d + 2$ such that all maximizers of the multi-information are contained in the set

$$\left\{ \sum_{f \in \mathscr{Y}^{\mathscr{X}}} p(f)\,\kappa_f \;:\; p \in \bar{\mathcal{E}} \right\} \;\subseteq\; \bar{\mathcal{C}}(\mathscr{X};\mathscr{Y})$$

A second approach based on Stephan Weis' dissertation, University of Erlangen Nuremberg 2009

A different way to see the previous approach

The previous geometry can also be obtained by the affine embedding

$$\bar{\mathcal{C}}(\mathscr{X};\mathscr{Y}) \hookrightarrow \bar{\mathcal{P}}(\mathscr{X} \times \mathscr{Y}), \qquad \kappa \mapsto \frac{1}{|\mathscr{X}|}\,\kappa$$

The more general affine embedding:

$$\kappa \mapsto p \otimes \kappa, \qquad (p \otimes \kappa)(x;y) := p(x)\,\kappa(x;y).$$

It does change the geometry but does not change the results on maximizers of multi-information.

Stationarity and the Kirchhoff polytope

Consider

$$\otimes : \bar{\mathcal{P}}(\mathscr{X}) \times \bar{\mathcal{C}}(\mathscr{X};\mathscr{X}) \rightarrow \bar{\mathcal{P}}(\mathscr{X} \times \mathscr{X}),$$

$$(p,\kappa) \mapsto (p \otimes \kappa)(x;x') := p(x)\,\kappa(x;x')$$

Remark: This map is surjective but at the boundary not injective.

Assume stationarity:

$$\bar{\mathcal{S}}(\mathscr{X}) := \left\{ p \in \bar{\mathcal{P}}(\mathscr{X} \times \mathscr{X}) \,:\, \sum_{x'} p(x,x') = \sum_{x'} p(x',x) \right\}$$

We have the following correspondence:

$$\mathcal{S}(\mathscr{X}) \longleftrightarrow \mathcal{C}(\mathscr{X};\mathscr{X})$$

Multi-information

Consider N nodes with state sets $\mathscr{X}_i$ and the following map:

$$\otimes : \;\; \bar{\mathcal{S}}(\mathscr{X}_1) \times \cdots \times \bar{\mathcal{S}}(\mathscr{X}_N) \;\; \to \;\; \bar{\mathcal{S}}(\mathscr{X}_1 \times \cdots \times \mathscr{X}_N)$$

Note that $\mathrm{im}(\otimes)$ is not an exponential family. It is the intersection of an exponential family with a convex set.

$$I(p) \;:=\; \inf_{q \in \mathrm{im}(\otimes)} D(p \,\|\, q)$$

Extreme points of $\bar{\mathcal{S}}(\mathscr{X})$

Consider a set $\emptyset \neq \mathscr{U} \subseteq \mathscr{X}$ and a cyclic permutation $\pi : \mathscr{U} \to \mathscr{U}$.

$$c(\mathscr{U}, \pi)(x, x') := \frac{1}{|\mathscr{U}|} \cdot \begin{cases} 1, & \text{if } \pi(x) = x' \\ 0, & \text{otherwise} \end{cases}$$

The number of extreme points

$$|\mathrm{ext}(\bar{\mathcal{S}}(\mathscr{X}))| = \sum_{k=1}^{|\mathscr{X}|} \binom{|\mathscr{X}|}{k} (k-1)! \leq |\mathscr{X}|^{|\mathscr{X}|} = |\mathrm{ext}(\bar{\mathcal{C}}(\mathscr{X};\mathscr{X}))|$$

$\lvert\mathscr{X}\rvert$	dim	$\lvert\mathrm{ext}(\bar{\mathcal{S}}(\mathscr{X}))\rvert$	$\lvert\mathrm{ext}(\bar{\mathcal{C}}(\mathscr{X};\mathscr{X}))\rvert$
2	2	3	4
3	6	8	27
4	12	24	256
5	20	89	3125

Local maximizers

Each local maximizer p of I satisfies

$$\dim\big(F(p)\big) \;\leq\; \sum_{i=1}^{N} (|\mathscr{X}_i|^2 - |\mathscr{X}_i|) \qquad (=\; 2\,N \text{ for binary nodes})$$

With the Carathéodory theorem there are at most $\sum_{i=1}^{N}(|\mathscr{X}_i|^2 - |\mathscr{X}_i|) + 1$ cycles so that p can be represented as convex combination of them. This implies that for all x with $\sum_{x'} p(x,x') > 0$:

$$|\mathrm{supp} p(\cdot\,|\,x)| \;\leq\; \sum_{i=1}^{N} (|\mathscr{X}_i|^2 - |\mathscr{X}_i|) + 1,$$

which implies

$$H(X'\,|\,X) \;\leq\; \ln\left(\sum_{i=1}^{N} (|\mathscr{X}_i|^2 - |\mathscr{X}_i|) + 1\right)$$

Next IGAIA

Information Geometry and its Applications III

August 2 - 6, 2010 MPI for Mathematics in the Sciences, Leipzig, Germany

http://www.mis.mpg.de/calendar/conferences/2010/infgeo.html