Entropy and Information

Matilde Marcolli

Ma148: Geometry of Information
Caltech, Fall 2021

Shannon Entropy/Information

- bit memory storage unit = switch with two on/off positions = digit 0 or 1
- A set of switches with $N = 2^{\#A}$ positions
- possible states: write a number $m = \sum_{k=0}^{\#A-1} s_k \, 2^k$ in binary notation $s_k \in \{0, 1\}$
- need $\#A = \frac{\log N}{\log 2}$ bits to select one particular possible configuration
- $b = \log N$ measured in $\log 2$ units is the bit number
- if have probability p_i of an event i in a set $i \in \{1, \ldots, R\}$ such as a frequency of occurrence

$$p_i = \frac{N_i}{N}, \qquad N = \sum_{i=1}^{R} N_i$$

- number of bits required to identify a particular configuration α among all possible is $\log N$
- to select an α either select among all or first select which set of N_i elements it belongs to and then among these so $b_i + \log N_i = \log N$ hence $b_i = -\log p_i$
- Shannon information measure: the average of the b_i with respect to the probabilities p_i

$$\mathcal{I}(P) = \sum_{i=1}^{R} p_i \log p_i$$

- Shannon Entropy: $S(P) = -\mathcal{I}(P)$ ("negative information", in fact positive $S(P) \geq 0$)
- measure of knowledge of the observed about what event to expect knowing $P = (p_i)$ (least knowledge at the uniform distribution, most knowledge at the delta measures δ_i)
- if the events i are dynamical microstates of a physical system then it is the entropy in the thermodynamic sense

Khinchin Axioms and Shannon Entropy $\mathcal{I}_R(p_1, \ldots, p_R)$

- Khinchin Axioms

1. continuous function of $P = (p_1, \ldots, p_R)$
2. minimum at the uniform distribution (max for entropy):
$$\mathcal{I}_R(\frac{1}{R}, \ldots, \frac{1}{R}) \leq \mathcal{I}_R(P)$$
3. extendability: $\mathcal{I}_R(p_1, \ldots, p_R) = \mathcal{I}_{R+1}(p_1, \ldots, p_R, 0)$
4. extensivity (implies additivity on independent subsystems)
$$\mathcal{I}(P) = \mathcal{I}(P') + \sum_i p'_i \mathcal{I}(Q|i)$$
for a composite system $P = (p_{ij})$ with $p_{ij} = Q(j|i)\, p'_i$ with conditional probabilities $Q(j|i)$ of j given i with conditional information
$$\mathcal{I}(Q|i) = \sum_j Q(j|i) \log Q(j|i)$$

Note: case of independent subsystems $p_{ij} = p'_i p''_j$ gives $\mathcal{I}(P) = \mathcal{I}(P') + \mathcal{I}(P'')$

Axiomatic characterization of the Shannon Entropy

• family of functionals $\mathcal{I} = \{\mathcal{I}_R\}$ satisfying Khinchin axioms agree with the Shannon information up to a positive constant

$$\mathcal{I}(P) = C \cdot \sum_i p_i \log p_i, \qquad \text{for some } C > 0$$

- at the uniform distribution: $p_{ij} = Q(j|i)\, p'_i$ with $p_{ij} = 1/N$ and $N = R \cdot r$ with $p'_i = 1/R$ and $Q(j|i) = 1/r$ obtain for $f(R) := \mathcal{I}_R(\frac{1}{R}, \ldots, \frac{1}{R})$ a function with $f(Rr) = f(R) + f(r)$ and continuous

$$f(R) = -C \cdot \log(R) \quad \text{for some } C \in \mathbb{R}^*$$

- also have $f(R) \geq f(R+1)$ by second and third axioms, so $C > 0$
- then from uniform to non-uniform: take p_{ij} and $Q(j|i)$ still uniform but p'_i arbitrary $f(N) = \mathcal{I}(P') + \sum_i p'_i f(N_i)$

$$\mathcal{I}(P') = -\sum_i p'_i (f(N_i) - f(N)) = C \sum_i p'_i \log p'_i$$

Rényi Entropy

• weaken the requirement of extensivity (non-extensive entropies) and replace only with additivity on statistically independent subsystems

$$p_{ij} = p'_i p''_j \;\Rightarrow\; \mathcal{I}(P) = \mathcal{I}(P') + \mathcal{I}(P'')$$

• then other solutions (not proportional to Shannon entropy): Rényi information

$$\mathcal{I}_\beta(P) = \frac{1}{\beta - 1} \log(\sum_{i=1}^R p_i^\beta)$$

Shannon Entropy as limit of Rényi Entropy

- $\mathcal{I}_\beta(P)$ defined for $\beta \in \mathbb{R}_+$ with $\beta \neq 1$
- limit when $\beta \to 1$: expand in $\epsilon = \beta - 1$

$$\sum_i p_i^{1+\epsilon} = \sum_i p_i \exp(\epsilon \log p_i) \sim \sum_i p_i (1 + \epsilon \log p_i)$$

$$= 1 + \epsilon \sum_i p_i \log p_i$$

so limit of the Rényi Entropy

$$\lim_{\epsilon \to 0} \mathcal{I}_{1+\epsilon}(P) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \log(1 + \epsilon \sum_i p_i \log p_i)$$

$$= \sum_i p_i \log p_i = \mathcal{I}(P)$$

Kullback–Leibler Divergence (Relative Entropy)

- given known probability distribution $P = (p_i)$ modified by some process to a new $Q = (q_i)$ with $q_i > 0$
- want to evaluate the information transfer of this process: $b_i(P) - b_i(Q) = \log(p_i/q_i)$
- estimate the mean value (in the known distribution)

$$KL(P|Q) := \sum_i p_i \log(p_i/q_i)$$

- non-negative because

$$\log x \geq 1 - \frac{1}{x} \;\Rightarrow\; \sum_i p_i \log\frac{p_i}{q_i} \geq \sum_i p_i(1 - \frac{p_i}{q_i}) = 0$$

- minimum value at 0 for $P = Q$ (again because $\log x > 1 - x^{-1}$ except at at $x = 1$ where equal)
- if uniform distribution $q_i = 1/R$ then $K(P|Q) = \mathcal{I}(P) + \log R$

Properties of Rényi Entropy

- monotonically increasing function: $\mathcal{I}_\beta(P) \leq \mathcal{I}_{\beta'}(P)$ when $\beta < \beta'$ for any P (so upper and lower bounds for Shannon entropy for $\beta > 1$ and $\beta < 1$)
- check monotonicity:

$$\frac{\partial \mathcal{I}_\beta(P)}{\partial \beta} = \frac{1}{(1-\beta)^2} \sum_i \mathbb{P}_i \log(\frac{\mathbb{P}_i}{p_i})$$

where escort probabilities

$$\mathbb{P}_i = \frac{p_i^\beta}{\sum_j p_j^\beta}$$

Kullback–Leibler Divergence is non-negative so monotonicity

- also another estimate for $\beta' > 0$ and $\beta\beta' > 0$

$$\frac{\beta - 1}{\beta}\mathcal{I}_\beta(P) \geq \frac{\beta' - 1}{\beta'}\mathcal{I}_{\beta'}(P)$$

- function x^σ convex for $\sigma > 1$ and concave for $0 < \sigma < 1$ so

$$(\sum_i a_i^\sigma) \geq \sum_i a_i^\sigma, \quad \forall \sigma > 1$$

$$(\sum_i a_i^\sigma) \leq \sum_i a_i^\sigma, \quad \forall 0 < \sigma < 1$$

take $a_i = p_i^\beta$ and $\sigma = \beta'/\beta$

$$(\sum_i p_i^\beta)^{\beta'/\beta} \geq \sum_i p_i^{\beta'} \quad \text{for } \beta' > \beta > 0$$

$$(\sum_i p_i^\beta)^{\beta'/\beta} \leq \sum_i p_i^{\beta'} \quad \text{for } \beta < \beta' < 0$$

- then taking $1/\beta'$ power (and then log)

$$(\sum_i p_i^\beta)^{1/\beta} \geq (\sum_i p_i^{\beta'})^{1/\beta'}$$

- monotonicity in β of

$$\Psi(\beta) := (1-\beta)\mathcal{I}_\beta = -\log \sum_i p_i^\beta$$

$$\Psi(\beta) \leq \Psi(\beta') \quad \text{for} \quad \beta' > \beta$$

because $p_i^\beta \geq p_i^{\beta'}$ and $-\log \sum_i p_i^\beta \leq -\log \sum_i p_i^{\beta'}$

- also have concavity in β

$$\frac{\partial^2 \Psi}{\partial \beta^2} \leq 0$$

Escort probabilities and statistical mechanics

- if write $p_i = \exp(-b_i)$ with $\sum_i p_i = 1$ (see later box-counting)
- then associated escort distribution

$$\mathbb{P}_i = \frac{p_i^\beta}{\sum_i p_i^\beta}$$

for $\beta \to \infty$ largest p_i dominates, for $\beta \to -\infty$ smallest

- analogy with statistical mechanics $\mathbb{P}_i = \exp(\Psi - \beta b_i)$ with $\Psi(\beta) = -\log Z(\beta)$ with partition function

$$Z(\beta) := \sum_i \exp(-\beta b_i) = \sum_i p_i^\beta$$

- Helmholtz free energy

$$F(\beta) := -\frac{1}{\beta} \log Z(\beta) = \frac{1}{\beta} \Psi(\beta)$$

- directly related to Rényi information

$$\mathcal{I}_\beta(P) = \frac{1}{\beta - 1} \log \sum_i p_i^\beta = -\frac{1}{\beta - 1} \Psi(\beta)$$

- probabilities p_i of microstates of a physical system
- M_i value at state i of a random variable M: expectation value

$$\langle M \rangle_P = \sum_i M_i \, p_i$$

- max-ent principle: look for p_i's that maximize entropy
- "unbiased guess" in information theory: minimize information
- generalized canonical distribution: p_i such that

$$\delta \mathcal{I}(P) = \sum_i (1 + \log p_i) \delta p_i = 0$$

with $\sum_i M_i^\sigma \, \delta p_i = 0$ (all observables M^σ) and $\sum_i \delta p_i = 0$

- multiply these constraints by an arbitrary factor β_σ (Lagrange multipliers)

$$\sum_i (\log p_i - \Psi + \sum_\sigma \beta_\sigma M_i^\sigma) \delta p_i = 0$$

- interpret then as probabilities

$$\mathbb{P}_i = \exp(\Psi - \sum_\sigma \beta_\sigma M^\sigma)$$

by imposing normalization condition $\sum_i \mathbb{P}_i = 1$

- normalization condition gives

$$\Psi = -\log Z(\beta) \quad \text{for} \quad Z(\beta) = \sum_i \exp(-\sum_\sigma \beta_\sigma M_i^\sigma)$$

- Example: Gibbs distribution mean energy $M = E = (E_i)$ of a system in thermodynamic equilibrium

$$\mathbb{P}_i = \exp(\beta(F - E_i)) \quad \text{with} \quad F = \frac{1}{\beta}\Psi(\beta)$$

Helmholtz free energy at inverse temperature $\beta = 1/T$

$$Z(\beta) = \exp(-\beta F) = \sum_i \exp(-\beta E_i)$$

sum of microstates of the system with energies E_i

- entropy in the thermodynamic sense for such a system is

$$S = \beta(E - F)$$

- Shannon entropy agrees with (expectation value of) thermodynamic entropy

$$-\sum_i \mathbb{P}_i \log \mathbb{P}_i = -\sum_i e^{\beta(F-E_i)} \beta(F - E_i) = \langle S \rangle$$

Box-counting and Rényi entropy

- bounded set $E \subset \mathbb{R}^N$, say $E \subset [0,1]^N$
- probability measure μ on $[0,1]^N$ with support on E
- divide $[0,1]^N$ in boxes of equal size: cubes of side ϵ
- count number r of boxes that meet E in a subset of positive μ-measure

$$r \leq R \sim \epsilon^{-N}$$

  total number of boxes in $[0,1]^N$

- $p_i = p_i(\epsilon)$ probability assigned to the i-th box B_i

$$p_i = \mu(E \cap B_i)$$

  crowding index

$$\alpha_i(\epsilon) = \frac{\log p_i(\epsilon)}{\log \epsilon}$$

- it is also function of x point where the box is centered $\alpha(x, \epsilon)$

- pointwise dimension $\alpha(x) = \lim_{\epsilon \to 0} \alpha(x, \epsilon)$ if limit exists (local scaling exponent)
- in terms of "bits numbers" $p_i = \exp(-b_i)$

$$b_i = -\alpha_i(\epsilon) \log \epsilon$$

- escort distribution

$$\mathbb{P}_i = \exp(\Psi - \beta b_i)$$

$$\Psi(\beta) = -\log \sum_i \exp(-\beta b_i) = -(\beta - 1)\mathcal{I}_\beta(P)$$

- and partition function

$$Z(\beta) = \sum_i p_i^\beta = \sum_i \exp(-\beta b_i)$$

$$\mathcal{I}_\beta(P) = \frac{1}{\beta - 1} \log Z(\beta) = \frac{1}{\beta - 1} \log \sum_i p_i^\beta$$

Rényi (box-counting) dimensions

- the partition function $Z(\beta)$ for $p_i = p_i(\epsilon)$ diverges for $\epsilon \to 0$
- but it satisfies a power law with exponent that gives an associated dimension
- Rényi dimension

$$D(\beta) = \lim_{\epsilon \to 0} \frac{\mathcal{I}_\beta(P_\epsilon)}{\log \epsilon} = \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \frac{1}{\beta - 1} \log \sum_i p_i(\epsilon)^\beta$$

$$Z(\beta) \sim_{\epsilon \to 0} \epsilon^{(\beta-1)D(\beta)}$$

Meaning of Rényi Dimensions

- at $\beta = 0$ have $\mathcal{I}_0(P) = -\log r(\epsilon)$ with $r(\epsilon) =$ min number of boxes of size ϵ covering set E so $D(0)$ is box-counting dimension (with grid)

$$D(0) = -\lim_{\epsilon \to 0} \frac{\log r(\epsilon)}{\log \epsilon}$$

- Shannon entropy dimension: at $\beta = 1$ limit of Rényi entropies is Shannon entropy $\mathrm{Sh}(P) = \mathcal{I}(P) = -\sum_i p_i \log p_i$

$$D(1) = \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \sum_{i=1}^{r(\epsilon)} p_i(\epsilon) \log p_i(\epsilon) = -\lim_{\epsilon \to 0} \frac{\mathrm{Sh}(P_\epsilon)}{\log \epsilon}$$

- $D(2)$ is called correlation dimension: it estimates effects of propagation of errors in iterates of a chaotic dynamical system; shown by Yorke, Grebogi, Ott that for certain classes of chaotic dynamical systems average period length $\sim \Delta^{-D(2)/2}$ (where Δ is a measure of precision)
- limit $\beta \to \infty$ of $D(\beta)$ measures scaling properties of region of E where measure μ most concentrated
- limit $\beta \to -\infty$ of $D(\beta)$ regions where least concentrated

• Note: these Rényi dimensions $D(\beta) = D_\mu(\beta)$ depend also on the measure μ used to compute $p_i = \mu(E \cap B_i)$ for the boxes B_i

Properties of Rényi Dimensions

- positivity $D(\beta) \geq 0$
- monotonicity $D(\beta') \leq D(\beta)$ for $\beta' \geq \beta$
- other relation: for $\beta' \geq \beta$ and $\beta\beta' > 0$

$$\frac{\beta' - 1}{\beta'} D(\beta') \geq \frac{\beta - 1}{\beta} D(\beta)$$

- limiting cases

$$D(\beta) \leq \frac{\beta}{\beta - 1} D(\infty) \quad \text{for} \quad \beta > 1$$

$$D(\beta) \geq \frac{\beta}{\beta - 1} D(-\infty) \quad \text{for} \quad \beta < 0$$

All of these properties follow from the corresponding properties of the Rényi entropy

Thermodynamic relations when box size $\epsilon \to 0$

- take $V = -\log \epsilon$ so $V \to \infty$
- dynamically homogeneous system if for large V quantities like entropy S or observables M^σ become proportional to V
- especially so that for β fixed and $V \to \infty$ ratios S/V or M^σ/V remain finite
- continuum limit: formally replace summations by integrals

$$\Psi = -\log \int_{\alpha_{\min}}^{\alpha_{\max}} \exp(-\beta\alpha V)\,\gamma(\alpha)\,d\alpha$$

- density of states $\gamma(\alpha)\,d\alpha$ umber of boxes with crowding index between α and $\alpha + d\alpha$
- expect asymptotic scaling behavior $\gamma(\alpha) = \epsilon^{-f(\alpha)}$ for some function $f(\alpha)$
- if $\gamma(\alpha) \sim \epsilon^{-f(\alpha)}$

$$\Psi = -\log \int_{\alpha_{\min}}^{\alpha_{\max}} \exp((f(\alpha) - \beta\alpha)V)\,d\alpha$$

- Saddle point approximation method
  - if integrand has only one maximum in interval then as $V \to \infty$ integral concentrated near the maximum
  - in general: want to evaluate

$$\mathcal{I} = \int \exp(F(x)V) dx$$

for $V \to \infty$, with some smooth function $F(x)$ with single max at $x = x_0$ (e.g. $F(x) = -(x - x_0)^2$)

  - with $F'(x_0) = 0$ and $F''(x_0) < 0$

$$\mathcal{I} \sim \int \exp((F(x_0) + \frac{1}{2}(x - x_0)^2 F''(x_0))V)\, dx$$

$$= \left(\frac{2\pi}{V\, F''(x_0)}\right)^{1/2} \exp(F(x_0)V)$$

  - so have $-\log \mathcal{I} \sim -F(x_0)V$

- Entropy Density
  - take $F(\alpha) = f(\alpha) - \beta\alpha$
  - $b := \alpha V$ mean value of bit number $\sum_i b_i p_i$ with $b_i = -\alpha_i \log \epsilon$
  - with saddle point approximation

$$\Psi \sim (\beta\alpha - f(\alpha))V = \beta b - S$$

  - α mean crowding index is like a mean energy density so $\Psi = \beta F = \beta E - S = \beta\alpha V - S$
  - so function $f(\alpha)$ is entropy density

$$f(\alpha) = \lim_{V \to \infty} \frac{S}{V}$$

  - interpret $f(\alpha)$ as an estimate of the fractal dimension of a set of boxes of average pointwise dimension α
  - $f(\alpha) =$ spectrum of local dimensions (multifractal)

- Legendre transform
  - density $\tau(\beta)$

$$\tau(\beta) = \lim_{V \to \infty} \frac{\Psi}{V}$$

  - by previous relation of Ψ to Rényi entropy: function of Rényi dimension

$$\tau(\beta) = (\beta - 1)\, D(\beta)$$

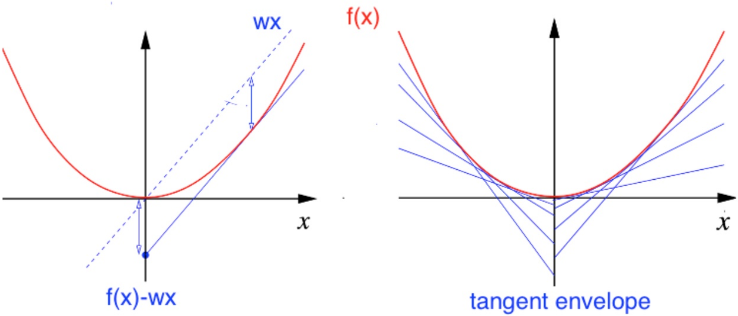
  - Legedre transforms

$$S(b) = \beta b - \Psi(\beta) \quad \text{with} \quad \frac{d\Psi}{d\beta} = b, \quad \frac{dS}{d\beta} = \beta$$

$$f(\alpha) = \beta\alpha - \tau(\beta) \quad \text{with} \quad \frac{d\tau}{d\beta} = \alpha, \quad \frac{df}{d\alpha} = \beta$$

  - convex differentiable function $F(x)$ Legendre transform

$$F^*(w) := \sup_x (wx - F(x))$$



value of Legendre transform $F^*(w)$ is the negative of the y-intercept of the tangent line to the graph of F that has slope w

- take $\alpha(\beta)$ to be the value α where $\beta\alpha - f(\alpha)$ takes minimum
- from $\tau(\beta) = (\beta - 1)D(\beta)$ and Legendre transform get

$$\alpha(\beta) = D(\beta) + (\beta - 1)D'(\beta)$$

$$f(\alpha(\beta)) = D(\beta) + \beta(\beta - 1)D'(\beta)$$

- for $\beta = 0$ and $\beta = 1$ this gives

$$f(\alpha(0)) = D(0) = \alpha(0) + D'(0)$$

with $D(0)$ box-counting dimension

$$f(\alpha(1)) = D(1) = \alpha(1)$$

entropy dimension

Tsallis Entropy

- Tsallis deformation of the Shannon entropy

$$S_q(P) = \frac{1}{q-1}\left(1 - \sum_i p_i^q\right)$$

- $q \to 1$ limit recovers the Shannon entropy

$$\lim_{q\to 1} S_q(P) = S(P) = -\sum_i p_i \log p_i$$

- For Shannon entropy have

$$S(P) = -\lim_{x\to 1} \frac{d}{dx} \sum_i p_i^x$$

- Tsallis entropy same property with respect to q-derivative

$$S_q(P) = -\lim_{x\to 1} D_q \sum_i p_i^x$$

- q-derivative

$$D_q f(x) = \frac{f(qx) - f(x)}{qx - x}$$

q-analogs and Tsallis entropy

- q-derivative $D_q x^n = \frac{1-q^n}{1-q} x^{n-1} = [n]_q x^{n-1}$

$$[n]_q = \frac{1-q^n}{1-q}$$

  q-analogs of the integers: $\lim_{q\to 1} [n]_q = n$

- Tsallis entropy $S_q(P)$ is a q-analog of Shannon entropy
- non-extensive thermodynamics: X, Y independent $\mathbb{P}(X,Y) = \mathbb{P}(X)\mathbb{P}(Y)$

$$S_q(X,Y) = S_q(X) + S_q(Y) + (1-q) S_q(X) S_q(Y)$$

  lack of linearity over independent systems measured by $1-q$

Tsallis deformation of KL divergence

- one-parameter deformation of the Kullback–Leibler divergence

$$\mathrm{KL}_\alpha(P||Q) = \frac{1}{1-\alpha} \sum_i P_i \left((\frac{P_i}{Q_i})^{1-\alpha} - 1 \right).$$

- recovers KL divergence in the limit $\alpha \to 1$

$$\mathrm{KL}_\alpha(P||Q) \to_{\alpha\to 1} \mathrm{KL}(P||Q) = \sum_i P_i \log(\frac{P_i}{Q_i})$$

q-analogs and geometry of the Tsallis entropy

- J.P. Vigneaux, *Information theory with finite vector spaces*, arXiv:1807.05152
- combinatorial meaning of the Shannon entropy: asymptotics of multinomial coefficients

$$\lim_{n\to\infty} \log \binom{n}{k_1,\ldots,k_N} = -\sum_{i=1}^N p_i \log p_i \qquad (p_i = k_i/n)$$

$$\sum_{k_1+\cdots+k_N=n} \binom{n}{k_1,\ldots,k_N} u_1^{k_1}\cdots u_N^{k_N} = (u_1+\cdots+u_N)^n$$

with $\binom{n}{k_1,\ldots,k_N} = \frac{n!}{k_1!\cdots k_N!}$

meaning of Shannon entropy and multinomial coefficients

- sequence of length n with symbols in an alphabet $\mathfrak{A} = \{z_1, \dots, z_N\}$ with probabilities $P = (P_z)$
- sequences generated by memoryless Bernoulli process with probabilities P
- cardinality of set of sequences of a certain type in P (eg ratio of zeros and ones)

$$\binom{n}{P(z_1)n, \dots, P(z_N)n} \sim \exp(nS(P))$$

  $P(z_i)$ is fraction of z_i entries in length n string, $P(z_i)n = k_i$ number of z_i entries in message
- each with probability approximately

$$\prod_{z \in \mathfrak{A}} P(z)^{nP(z)} \sim \exp(-nS(P))$$

- Shannon's principle: "it is possible for most purposes to treat the long sequences as though there were just 2^{nS} of them, each with a probability 2^{-nS}"

q-analog of multinomial coefficients

- q-analog of the integers $[n]_q = \frac{1-q^n}{1-q}$
- q-factorial $[n]_q! = [n]_q[n-1]_q \cdots [1]_q$
- q-multinomial coefficients

$$\left[\begin{array}{c} n \\ k_1, \ldots, k_N \end{array} \right]_q := \frac{[n]_q!}{[k_1]_q! \cdots [k_N]_q!}, \quad \sum_{i=1}^{N} k_i = n$$

- when $q = p^r$ some prime p these count points over field $\mathbb{F}_q$

$$[n]_q = \#\mathbb{P}^1(\mathbb{F}_q), \quad \left[\begin{array}{c} n \\ k_1, \ldots, k_N \end{array} \right]_q = \#\mathcal{F}_{\underline{k},n}(\mathbb{F}_q)$$

  $\mathcal{F}_{\underline{k},n}$ variety of flags $V_1 \subset V_2 \subset \cdots \subset V_N = \mathbb{F}_q^n$ with $\dim V_\ell = \sum_{i=1}^{\ell} k_i$, flags of type $\underline{k} = (k_1, \ldots, k_N)$
- q-binomial coefficient

$$\left[\begin{array}{c} n \\ k \end{array} \right]_q := \left[\begin{array}{c} n \\ k, n-k \end{array} \right]_q = \#\{V \subset \mathbb{F}_q^n, \ \dim V = k\}$$

q-analog of Bernoulli generated sequences

- statistical model where length n message is a flag of vector spaces $V_1 \subset V_2 \subset \cdots \subset V_N = \mathbb{F}_q^n$ with $\dim V_\ell \setminus V_{\ell-1} = k_\ell$
- choice of a flag in $\mathbb{F}_q^n$ replaces "a configuration of n-particles"
- for configuration of particles total energy depends on type $\underline{k}$

$$\langle E \rangle = \text{mean internal energy } = \sum_{i=1}^{N} \frac{k_i}{n} E_i$$

with E_i energy associated to spin state $z_i \in \mathfrak{A}$

- for a flag $V_1 \subset V_2 \subset \cdots \subset V_N = \mathbb{F}_q^n$ energy

$$\langle E \rangle = \sum_{i=1}^{N} \frac{k_i}{n} E_i = \sum_{i=1}^{N} \frac{\dim(V_i)}{n} \tilde{E}_i$$

with $\dim V_k = \sum_{i=1}^{k} k_i$ and $\tilde{E}_i$ satisfying identity (V_i-energy)

max-entropy principle

- equations

$$\langle E \rangle = \sum_{i=1}^{N} \frac{k_i}{n} E_i \quad \text{and} \quad \sum_{i=1}^{N} k_i = n$$

do not determine uniquely $\underline{k} = (k_1, \ldots, k_N)$

- max-entropy: among all solutions $\underline{k}$ of the equations select the one that corresponds to the largest number of configurations of the system
- here it means maximizing the q-deformed multinomial coefficient

$$\begin{bmatrix} n \\ k_1, \ldots, k_N \end{bmatrix}_q$$

Limiting behavior of q-multinomial coefficients

- Pochhammer symbol

$$(a;x)_n := \prod_{k=0}^{n-1}(1-ax^k), \quad (a;x)_0 = 1$$

- q-Gamma function $\Gamma_q(n+1) = [n]_q!$

$$\Gamma_q(x) = (q^{-1};q^{-1})_\infty \, q^{\binom{x}{2}} \, (q-1)^{1-x} \sum_{n=0}^{\infty} \frac{q^{-nx}}{(q^{-1};q^{-1})_n}$$

- q-multinomial coefficients

$$\left[\begin{array}{c} n \\ k_1,\ldots,k_N \end{array} \right]_q = \frac{\Gamma_q(n+1)}{\Gamma_q(k_1+1)\cdots\Gamma_q(k_N+1)}$$

- quadratic Tsallis entropy $Ts_2(p_1,\ldots,p_N) = 1 - \sum_{i=1}^N p_i^2$

$$\left[\begin{array}{c} n \\ k_1,\ldots,k_N \end{array} \right]_q = (q^{-1};q^{-1})_\infty^{1-N} q^{n^2 Ts_2(\frac{k_1}{n},\ldots,\frac{k_N}{n})/2} \frac{\prod_{i=1}^N (q^{-(k_i+1)};q^{-1})_\infty}{(q^{-(n+1)};q^{-1})_\infty}$$

binary Bernoulli process and q-analog ($\mathfrak{A} = \{0, 1\}$)

- binary string produced by a Bernoulli process $(p, 1-p)$
- Y_n sum of the first n outputs
- probability of $Y_n = k$ is $\binom{n}{k} p^k (1-p)^{n-k}$ (sequences with k ones each with probablity $p^k(1-p)^{n-k}$)
- q-binomial formula

$$(x+y)_q^n := (x+y)(x+qy)\cdots(x+q^{n-1}y) = \sum_{k=0}^{n} \left[\begin{array}{c} n \\ k \end{array} \right]_q q^{\binom{k}{2}} y^k x^{n-k}$$

- get probability distribution

$$\mathrm{Bin}_q(k|n,x,y) := \left[\begin{array}{c} n \\ k \end{array} \right]_q \frac{q^{\binom{k}{2}} y^k x^{n-k}}{(x+y)_q^n}$$

$$\mathrm{Bin}_q(k|n,\theta) := \left[\begin{array}{c} n \\ k \end{array} \right]_q \frac{q^{\binom{k}{2}} \theta^k}{(-\theta;q)_n}, \qquad \text{for } \theta = y/x \geq 0$$

- variable Y_n with this distribution can be written as sum of n independent variables X_i taking values in $\{0,1\}$ with probabilities $\frac{x}{x+yq^{i-1}}$ and $\frac{yq^{i-1}}{x+yq^{i-1}}$

vector space valued stochastic processes

- Grassmannian $Gr(k,n)$ of k-dim subspaces in $\mathbb{F}_q^n$ and $Gr(n) = \cup_k Gr(k,n)$ total Grassmannian
- fixed embeddings $\mathbb{F}_q^n \hookrightarrow \mathbb{F}_q^{n+1}$ relate $Gr(n)$ & $Gr(n+1)$
- $V_0 = \{0\}$ trivial vector space, V_{n+1} random variable with values in $Gr(n+1)$
- for $W \in Gr(n)$ (not in $Gr(n-1)$) dilation

$$\mathrm{Dil}_{n+1}(W) = \{V \in Gr(n+1) \,|\, W \subset V,\ V \not\subset Gr(n),\ \dim V - \dim W = 1\}$$

- probability distribution

$$P(V_{n+1} = V | V_n = W, X_{n+1} = 0) = \delta_{V,W}$$

$$P(V_{n+1} = V | V_n = W, X_{n+1} = 1) = \frac{\chi_{\mathrm{Dil}_{n+1}(W)}(V)}{\#\mathrm{Dil}_{n+1}(W)}$$

normalized characteristic function of set $\mathrm{Dil}_{n+1}(W)$

- from this distribution get for $\dim V = k$

$$P(V_n = V) = \frac{\theta^k q^{k(k-1)/2}}{(-\theta; q)_n}$$

$$P(\dim V_n = k) = \begin{bmatrix} n \\ k \end{bmatrix}_q \frac{\theta^k q^{k(k-1)/2}}{(-\theta; q)_n}$$

- show inductively for $V \in Gr(n+1)$ and $V \not\subset \mathbb{F}_q^n$

$$P(V_{n+1} = V) = \sum_{W \in Gr(n)} P(V_{n+1} | V_n, X_{n+1}) P(Y_n = W) P(X_{n+1} = 1)$$

$$= \sum_{W \in Gr(k-1,n), W \subset V} \frac{1}{\#\mathrm{Dil}_{n+1}(W)} \frac{\theta^{k-1} q^{\binom{k-1}{2}}}{(-\theta; q)_n} \frac{\theta q^n}{1 + \theta q^n}$$

$$= \frac{\theta^k q^{\binom{k-1}{2}} q^n}{\#\mathrm{Dil}_{n+1}(V \cap \mathbb{F}_q^n)(-\theta; q)_n}$$

where last uses $W \subset V$ is in $V \cap \mathbb{F}_q^n$ and same dim so $W = V \cap \mathbb{F}_q^n$

asymptotics

- fixed d and $n \to \infty$

$$P(V_n \in Gr(n-d,n)) \sim \frac{q^{-\frac{1}{2}(d-(\frac{1}{2}-\log_q \theta))^2+\frac{1}{2}(\frac{1}{2}-\log_q \theta)^2}(q^{-(d+1)};q^{-1})_\infty}{(q^{-1};q^{-1})_\infty(-\theta^{-1};q^{-1})_\infty}$$

- and sum over all $d \geq 0$ of rhs equal to 1 (asymptotic probability distribution)
- analogous processes for multinomial case with alphabet $\#\mathfrak{A} = N > 2$
- Question: are there other combinatorial quantities generalizing q-multinomial coefficients with asymptotics

$$\sim \exp(Ts_\alpha(p_1,\ldots,p_n)n^\alpha + o(n^\alpha))$$

for $\alpha \neq 1,2$?
- Fontené-Ward generalized multinomial coefficients
- J.P. Vigneaux, *A homological characterization of generalized multinomial coefficients related to the entropic chain rule*, arXiv:2003.02021

Summary of q-deformed information (binary case $\mathfrak{A} = \{0,1\}$)

Concept	**Shannon case**	**q-case**
Message at time n (n-message)	Word $w \in \{0,1\}^n$	Vector subspace $v \subset F_q^n$
Type	Number of ones	Dimension
Number of n-messages of type k	$\binom{n}{k}$	$\begin{bmatrix} n \\ k \end{bmatrix}_q$
Probability of a n-message of type k	$\xi^k (1-\xi)^{n-k}$	$\dfrac{\theta^k q^{k(k-1)/2}}{(-\theta;q)_n}$

What is the "field with one element"? (Manin, Soulé, etc.)

Finite geometries $\quad (q = p^k,\ p \text{ prime})$

$$\#\mathbb{P}^{n-1}(\mathbb{F}_q) = \frac{\#(\mathbb{A}^n(\mathbb{F}_q) \smallsetminus \{0\})}{\#\mathbb{G}_m(\mathbb{F}_q)} = \frac{q^n - 1}{q - 1} = [n]_q$$

$$\#\mathrm{Gr}(n,j)(\mathbb{F}_q) = \#\{\mathbb{P}^j(\mathbb{F}_q) \subset \mathbb{P}^n(\mathbb{F}_q)\}$$

$$= \frac{[n]_q!}{[j]_q![n-j]_q!} = \binom{n}{j}_q$$

$$[n]_q! = [n]_q[n-1]_q \cdots [1]_q, \quad [0]_q! = 1$$

The origin of $\mathbb{F}_1$-geometry: Jacques Tits observed if take $q = 1$

$$\mathbb{P}^{n-1}(\mathbb{F}_1) := \text{ finite set of cardinality } n$$

$$\mathrm{Gr}(n,j)(\mathbb{F}_1) := \text{ set of subsets of cardinality } j$$

Is there an algebraic geometry over $\mathbb{F}_1$?

Extensions $\mathbb{F}_{1^n}$ (Kapranov-Smirnov)
Monoid $\{0\} \cup \mu_n$ (n-th roots of unity)
- Vector space over $\mathbb{F}_{1^n}$: pointed set (V, v) with free action of μ_n on $V \smallsetminus \{v\}$
- Linear maps: permutations compatible with the action

$$\mathbb{F}_{1^n} \otimes_{\mathbb{F}_1} \mathbb{Z} := \mathbb{Z}[t, t^{-1}]/(t^n - 1)$$

Counting of points: for geometries X over $\mathbb{Z}$, reductions mod p

$$N_q(X) = \#X(\mathbb{F}_q), \qquad q = p^r$$

Polynomially countable if $N_q(X) = P_X(q)$ polynomial in q.
Counting of "points over the field with one element and its extensions"

$$P_X(m+1) = \#X(\mathbb{F}_{1^m})$$

General question: can reformulate combinatorial interpretation of Shannon and Tsallis entropies in terms of $\mathbb{F}_1$-geometry?

Shannon and Rényi Entropy and Functional Equation

- Rényi entropy $\mathcal{R}_\beta(P) = -\mathcal{I}_\beta(P) = \frac{1}{1-\beta}\log(\sum_i p_i^\beta)$
- $\lim_{\beta\to 1}\mathcal{R}_\beta(P) = S(P) = -\mathcal{I}(P)$ Shannon entropy
- Functional equation of Shannon entropy (extensivity)

$$H(x) + (1-x)H(\frac{y}{1-x}) - H(y) - (1-y)H(\frac{x}{1-y}) = 0$$

- equivalently for $ab = x$ and $y = 1-a$

$$S(ab) + (1-ab)S(\frac{a(1-b)}{1-ab}) = S(a) + aS(b)$$

- More general functional equation

$$H(x) + (1-x)^\beta H(\frac{y}{1-x}) - H(y) - (1-y)^\beta H(\frac{x}{1-y}) = 0$$

- Gy.Maksa, *The general solution of a functional equation related to the mixed theory of information*, Aequationes Mathematicae, Vol. 22 (1981), 90–96

Functional equations and polylogarithms over finite fields

- P. Elbaz-Vincent, H. Gangl, *On poly(ana)logs. I*. Compositio Math. 130 (2002), no. 2, 161–210.
- M.Kontsevich, *The* $1\frac{1}{2}$*-logarithm*, Appendix to previous paper, Compos. Math. 130 (2002) N.2, 211– 214.
- P. Elbaz-Vincent, H. Gangl, *Finite polylogarithms, their multiple analogues and the Shannon entropy*, Geometric science of information, 277–285, Lecture Notes in Comp. Sci., 9389, Springer, 2015.

Finite Logarithm

- finite logarithm: finite field $\mathbb{F}_q$, char p

$$\mathcal{L}_1^{(p)}(x) = \sum_{k=1}^{p-1} \frac{x^k}{k}$$

compare with usual $-\log(1-x) = \sum_{m\geq 1} \frac{x^m}{m}$

- Kontsevich observed: the finite logarithm is a solution to the general functional equation for $\beta = p$

$$\mathcal{L}_1^{(p)}(a) - \mathcal{L}_1^{(p)}(b) + a^p \mathcal{L}_1^{(p)}(\frac{b}{a}) + (1-a)^p \mathcal{L}_1^{(p)}(\frac{1-b}{1-a}) = 0$$

Functional equation

- the functional equation

$$\mathcal{L}_1^{(p)}(a) - \mathcal{L}_1^{(p)}(b) + a^p \mathcal{L}_1^{(p)}(\frac{b}{a}) + (1-a)^p \mathcal{L}_1^{(p)}(\frac{1-b}{1-a}) = 0$$

is a specialization to $(\infty, 0, 1, a, b)$ of the 5-terms relation

$$\sum_{i=1}^{5} (-1)^i \delta(x_1, \ldots, \hat{x}_i, \ldots, x_5)\, \mathcal{L}_1^{(p)}(cr(x_1, \ldots, \hat{x}_i, \ldots, x_5)),$$

$$cr(a,b,c,d) := \frac{a-c}{a-d}\frac{b-d}{b-c}, \quad \delta(a,b,c,d) = (a-d)(b-c)$$

- 5-terms relation is in fact equivalent to functional equation

Cohomological interpretation of the functional equation

- function $\varphi : \mathbb{F}_p \times \mathbb{F}_p \to \mathbb{F}_p$, zero if $x + y = 0$ and

$$\varphi(x, y) = (x + y)H(\frac{x}{x + y}), \quad \text{if } x + y \neq 0$$

  is a 2-cocycle

$$\varphi(b, c) - \varphi(a + b, c) + \varphi(a, b + c) - \varphi(b, c) = 0$$

- to see use $H(x) = H(1 - x)$ and set $X = \frac{x}{x+y+z}$ and $Y = \frac{y}{x+y+z}$ in functional equation
- $\varphi = d\eta$ coboundary if $\varphi(x, y) = -\eta(x + y) + \eta(x) + \eta(y)$

- φ is homogeneous so if coboundary $\varphi(\lambda x, \lambda y) = -\lambda\eta(x+y) + \lambda\eta(x) + \lambda\eta(y)$
- obtain additive morphism $\psi_\lambda(x) = \eta(\lambda x) - \lambda\eta(x)$, determined by $\psi_\lambda(1)$
- check that satisfies $\mu\psi_\lambda(1) = \psi_\lambda(\mu)$ and $\psi_{\lambda\mu}(1) = \psi_\lambda(\mu) + \lambda\psi_\mu(1)$ so

$$\psi_{\lambda^m}(1) = m\lambda^{m-1}\psi_\lambda(1)$$

- $\mathbb{F}_p^*$ generated by a primitive root ω with $\omega^{p-1} = 1$ and

$$0 = \psi_1(1) = (p-1)\omega^{p-2}\psi_\omega(1) \quad \Rightarrow \quad \psi_\omega(1) = 0$$

- this gives $\eta(\lambda x) = \lambda\eta(x)$ then η additive map so $d\eta = 0$ so $\varphi \neq 0$ cannot be a coboundary
- Kontsevich: solutions of general functional equation give non-zero 2-cocycles in $H^2(\mathbb{Z}/p\mathbb{Z}, \mathbb{Z}/p\mathbb{Z}) \simeq \mathbb{Z}/p\mathbb{Z}$
- so functional equation has 1-dim space of solutions, determines $\mathcal{L}_1^{(p)}$ up to a constant factor

Finite Polylogarithms

- Finite polylogs:

$$\mathcal{L}_n^{(p)}(x) = \sum_{k=1}^{p-1} \frac{x^k}{k^n}$$

Properties of Finite Polylogs

- satisfy differential relation

$$d\mathcal{L}_n^{(p)}(x) = \mathcal{L}_{n-1}^{(p)}\, d\log(x)$$

  and periodicity relation (Frobenius action $x \mapsto x^p$)

$$\mathcal{L}_{n+p-1}^{(p)} = \mathcal{L}_n^{(p)}$$

- inversion relation

$$\mathcal{L}_n^{(p)}(x) = (-1)^n x^p \mathcal{L}_n^{(p)}(1/x)$$

- if the field $\mathbb{F}_q$ contains m-th roots of 1 also relation (duplication for $m = 2$)

$$\mathcal{L}_n^{(p)}(x^m) = m^{n-1} \sum_{\zeta^m=1} \frac{1 - x^{pm}}{1 - \zeta^p x^p} \mathcal{L}_n^{(p)}(\zeta x)$$

Functional equations for $\mathcal{L}_2^{(p)}$ for $p > 3$

- for $\mathcal{L}_1^{(p)}$ functional equation together with inversion and duplication relation identify $\mathcal{L}_1^{(p)}$ uniquely
- functional equation for $\mathcal{L}_2^{(p)}$ (3-term relation)

$$x^p F(1 - \frac{1}{x}) - F(x) + F(1 - x) = 0$$

- this equation has a space of solutions of dimension at least $1 + \frac{p-1}{3}$

$$\tau_{i,p}(x) = x^i (1 - x)^i (x^{p-3i} + (-1)^i), \quad i = 0, \ldots, (p-1)/3$$

  give independent solutions
- additional equations for $\mathcal{L}_2^{(p)}$ (duplication)

$$2(1 + x^p)F(x) + 2(1 - x^p)H(-x) - F(x^2) = 0$$

- with functional equation above these characterize $\mathcal{L}_2^{(p)}$ up to a constant factor

Information Loss: categorical formulation (Baez–Fritz–Leinster)

- revisiting Khinchin axioms characterizing Shannon entropy in categorical terms
- category FinProb of probabilities (X, P) and morphisms $f : X \to Y$ measure preserving functions

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

- information loss $F : Mor_{\texttt{FinProb}} \to \mathbb{R}_+$
- axioms of information loss
  1. *functoriality*: $F(f \circ g) = F(f) + F(g)$ on composable morphisms
  2. *convex linearity*: $F(\lambda f \oplus (1-\lambda)g) = \lambda F(f) + (1-\lambda)F(g)$ where $\lambda f \oplus (1-\lambda)g$ induced on $(X \sqcup Y, \lambda P \oplus (1-\lambda)Q)$
  3. *continuity*: $F(f)$ continuous function of f

- If F satisfies axioms above then $F(f) = c\,(S(P) - S(Q))$ for some $c \geq 0$ and with $S(P) = -\sum_i p_i \log p_i$ the Shannon entropy
- first note that $S(P) - S(Q)$ satisfies axioms: key fact

$$S(P) - S(Q) = -\sum_i p_i \log p_i + \sum_j q_j \log q_j = \sum_{i \in X} p_i \log \frac{q_{f(i)}}{p_i}$$

  a conditional entropy
- to show that any F with info-loss axioms is proportional to $S(P) - S(Q)$ use Faddeev reformulation of Khinchin axioms

Faddeev formulation of Khinchin axioms

- $\mathcal{I}$ mapping probability measures on finite sets to $\mathbb{R}_+$ satisfying
  1. $\mathcal{I}$ invariant under bijections
  2. $\mathcal{I}$ continuous
  3. for $P = (p_1, \ldots, p_n)$ and $0 \leq t \leq 1$

$$\mathcal{I}(tp_1, (1-t)p_1, p_2, \ldots, p_n) = \mathcal{I}(p_1, \ldots, p_n) + p_1 \mathcal{I}(t, 1-t)$$

- then $\mathcal{I}$ must be a constant non-negative multiple of the Shannon entropy $S(P)$
- key is equivalence between last condition and extensivity of the Shannon entropy

$$S(P') = S(P) + \sum_i P_i S(Q|i) \qquad \text{for } P' = (P_i Q(j|i))$$

Baez-Fritz-Leinster information loss characterization

- unique morphism $1_P : (X, P) \to (\{x\}, 1)$ in $\mathtt{FinProb}$ (losing all information about (X, P) by collapsing it to a single point)
- $1_P = 1_Q \circ f$ for all $f : (X, P) \to (Y, Q)$
- $F(1_P) = F(1_Q) + F(f)$ so $F(f) = F(1_P) - F(1_Q)$
- set $\mathcal{I}(P) = F(1_P)$ and show this entropy function is indeed Shannon entropy by showing it satisfies Faddeev characterization

Characterization of Tsallis entropy (Baez-Fritz-Leinster)

- Tsallis information loss $F_\alpha : Mor_{\mathtt{FinProb}} \to \mathbb{R}_+$
- axioms of information loss
  1. *functoriality*: $F_\alpha(f \circ g) = F_\alpha(f) + F_\alpha(g)$ on composable morphisms
  2. *convex linearity*: $F_\alpha(\lambda f \oplus (1-\lambda) g) = \lambda^\alpha F(f) + (1-\lambda)^\alpha F(g)$ where $\lambda f \oplus (1-\lambda) g$ induced on $(X \sqcup Y, \lambda P \oplus (1-\lambda) Q)$
  3. *continuity*: $F(f)$ continuous function of f
- then $F_\alpha(f) = c\,(Ts_\alpha(P) - Ts_\alpha(Q))$ Tsallis entropy
- similar argument but replacing extensivilty property of Shannon entropy with nonextensive version of Tsallis
- version of Faddeev characterization for $Ts_\alpha(P)$

$$\mathcal{I}_\alpha(tp_1, (1-t)p_1, p_2, \ldots, p_n) = \mathcal{I}(p_1, \ldots, p_n) + p_1^\alpha \mathcal{I}(t, 1-t)$$

Kullback–Leibler divergence and Fisher–Rao metric

- Kullback–Leibler divergence $KL(P|Q) = \sum_i p_i \log(p_i/q_i)$ is not a metric
- ...but up to first order approximation it defines a metric

$$KL(P|P+dP) = \sum_i p_i \log(\frac{p_i}{p_i + dp_i})$$

- expansion

$$\sum_i p_i \log(\frac{p_i}{p_i + \epsilon q_i} = -\sum_i p_i \log(1 + \epsilon \frac{p_i}{q_i})$$

$$= -\sum_i p_i \epsilon \frac{p_i}{q_i} + \frac{1}{2} \sum_i p_i (\epsilon \frac{p_i}{q_i})^2 + o(\epsilon^2)$$

with $\sum_i q_i = 0$ (since $p_i + \epsilon q_i$ probability) so first term quadratic

- Fisher-Rao information metric

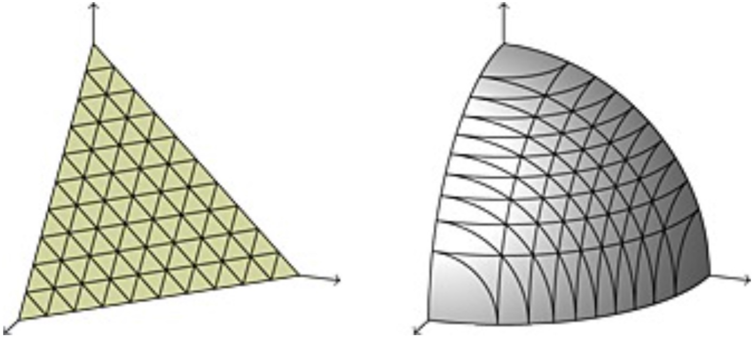
$$ds^2 = \sum_i \frac{dp_i^2}{p_i}$$

Properties of the Fisher–Rao metric

- with change of coordinates $X_i = \sqrt{p_i}$ with $dX_i = \frac{dp_i}{2\sqrt{p_i}}$ becomes Euclidean metric

$$ds^2 = \sum_i dX_i^2$$

but restricted to locus $\sum_i X_i^2 = \sum_i p_i = 1$, i.e. metric induced by ambient Euclidean space on the unit sphere



- Hessian Riemannian metrics: with a convex potential Φ

$$g_{ij} = \partial_i \partial_j \Phi$$

- totally symmetric rank 3 tensor

$$A_{ijk} = \partial_i \partial_j \partial_k \Phi$$

- Fisher-Rao metric tensor $ds^2 = \sum_{ij} g_{ij}(P) dp_i dp_j$ is Hessian of the Shannon entropy

$$g_{ij}(P) = -\frac{1}{4} \partial_i \partial_j S(P) = \frac{1}{4} \partial_i \partial_j \sum_k p_k \log p_k$$

as $-\partial_j S(P) = 1 + \log p_j$ and $\partial_j(1 + \log p_j) = \delta_{ij} \, p_i^{-1}$

- if T stochastic matrix $T \geq 0$ and $\sum_i T_{ij} = 1$

$$\sum_i \frac{(T\, dp)_i^2}{p_i} \leq \sum_i \frac{dp_i^2}{p_i}$$

- Levi-Civita connection (Christoffel symbols)

$$\Gamma^{\rho}_{\nu\sigma} = \frac{1}{2} g^{\rho\mu} (\partial_\sigma g_{\mu\nu} + \partial_\nu g_{\mu\sigma} - \partial_\mu g_{\nu\sigma})$$

convention of summation over repeated indices for tensor calculus

- Riemannian curvature $R^{\rho}{}_{\sigma\mu\nu}$

$$R^{\rho}{}_{\sigma\mu\nu} = \partial_\mu \Gamma^{\rho}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma}$$

- for Hessian metrics

$$\Gamma_{ijk} = \frac{1}{2} \partial_i \partial_j \partial_k \Phi$$

$$R_{ijkl} = \frac{1}{2} (S_{jikl} - S_{ijkl})$$

$$S_{ijkl} = \frac{1}{2} \frac{\partial^4 \Phi}{\partial_i \partial_j \partial_k \partial_l} - \frac{1}{2} g^{rs} \frac{\partial^3 \Phi}{\partial_i \partial_k \partial_r} \frac{\partial^3 \Phi}{\partial_j \partial_l \partial_s}$$

Conjugate connections manifolds

- conjugate connections manifold (M, g, ∇, ∇^*)

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla^*_X Z\rangle$$

  for all X, Y, Z smooth vector fields and $\langle , \rangle$ pairing via g
- given (M, g, ∇) unique dual structure (M, g, ∇^*) and $(\nabla^*)^* = \nabla$
- parallel transport along the dual connections preserves the metric

$$\langle X, Y\rangle_{\gamma(0)} = \langle \Pi^{\nabla}_{\gamma}(X), \Pi^{\nabla^*}_{\gamma}(Y)\rangle_{\gamma(t)}$$

- average $\bar\nabla = \frac{1}{2}(\nabla + \nabla^*)$ is self dual hence it is the Levi-Civita connection of g characterized by

$$X\langle Y, Z\rangle = \langle \bar\nabla_X Y, Z\rangle + \langle Y, \bar\nabla_X Z\rangle$$

$$\partial_k g_{ij} = \langle \bar\nabla_{\partial_k} \partial_i, \partial_j\rangle + \langle \partial_i, \bar\nabla_{\partial_k} \partial_j\rangle$$

$$\bar\Gamma^k_{ij} = \sum_l \frac{1}{2} g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})$$

Statistical manifolds

- statistical manifold (M, g, A) with Amari–Chentsov tensor

$$A(X,Y,Z) = \langle \nabla_X Y - \nabla^*_X Y, Z \rangle$$

$$A_{ijk} = \Gamma^k_{ij} - \Gamma^{*k}_{ij} \qquad A_{ijk} = A(\partial_i, \partial_j, \partial_k) = \langle \nabla_{\partial_i}\partial_j - \nabla^*_{\partial_i}\partial_j, \partial_k \rangle$$

- totally symmetric cubic tensor

useful fact: if a torsion-free affine connection ∇ has constant curvature κ then its conjugate ∇^* has same constant curvature κ

$$R^\nabla(X,Y)Z = \kappa\,(g(Y,Z)X - g(X,Z)Y)$$

$$R^\nabla(X,Y)Z := \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$$

$$R^\nabla(\partial_j, \partial_k)\partial_i = \sum_\ell R^\ell_{jki}\,\partial_\ell$$

(for details of proof: O.Calin, C.Udriste, *Geometric Modeling in Probability and Statistics*, Springer, 2014 [Proposition 8.1.4]

α-families and deformed connections

- statistical manifold (M, g, A)
- ∇^{LC} Levi-Civita connection of the metric g
- one-parameter families of connections

$$\Gamma^{\alpha}_{ijk} = \Gamma^{LC}_{ijk} - \frac{\alpha}{2} A_{ijk}, \qquad \Gamma^{-\alpha}_{ijk} = \Gamma^{LC}_{ijk} + \frac{\alpha}{2} A_{ijk}$$

- gives a conjugate connections manifold $(M, g, \nabla^{-\alpha}, \nabla^{\alpha} = (\nabla^{-\alpha})^*)$
- starting from conjugate connections manifold (M, g, ∇, ∇^*): α-deformations

$$\Gamma^{\alpha}_{ijk} = \frac{1+\alpha}{2} \Gamma_{ijk} + \frac{1-\alpha}{2} \Gamma^*_{ijk}$$

dual flat structures

- (M, g, A) is α-flat if ∇^{α} is flat
- $R^{\alpha} = -R^{-\alpha}$ so also $\nabla^{-\alpha}$ (dual) flat
- $\alpha = \pm 1$: ∇-flat iff ∇^*-flat

Kullback–Leibler divergence and thermodynamics

- for probability distribution $P_n = \frac{e^{-\beta\lambda_n}}{Z(\beta)}$ with partition function $Z(\beta) = \mathrm{Tr}(e^{-\beta H})$ and $Spec(H) = \{\lambda_n\}$ Shannon entropy is thermodynamic entropy

$$S = \left(1 - \beta \frac{\partial}{\partial\beta}\right) \log Z(\beta)$$

$$S = -\sum_n P_n \log P_n = \sum_n P_n \log Z(\beta) + \beta \sum_n P_n \lambda_n$$

  with $\sum_n P_n \lambda_n = \frac{\partial}{\partial\beta} \log Z(\beta)$

- free energy $F = -\log Z(\beta)$

- $Q_x = \frac{e^{-\beta H_x}}{Z(\beta)}$ with $Z(\beta) = \sum_x e^{-\beta H_x}$ partition function, and $P =$ given probability distribution
- Gibbs free energy given by

$$G(P) = -\log Z(\beta) + \sum_x P_x \log \frac{P_x}{Q_x},$$

- Kullback–Leibler divergence

$$KL(P|Q) = G(P) + \log Z(\beta)$$

- free energy is minimization of Gibbs energy over configuration space: since $KL(P|Q) \geq 0$

$$\min_P G(P) = -\log Z(\beta)$$

- mean field theory when computation of free energy not directly accessible, consider a trial Hamiltonian $\tilde{H}$ with probability distribution $P_x = \tilde{Z}(\beta)^{-1} e^{-\beta \tilde{H}_x}$ and $\tilde{Z}(\beta) := \sum_x e^{-\beta \tilde{H}_x}$
- Helmholtz free energy

$$-\sum_x P_x \log P_x = \log \tilde{Z}(\beta) + \beta \langle \tilde{H} \rangle = (1 - \beta \frac{\partial}{\partial \beta}) \log \tilde{Z}(\beta)$$

$$\sum_x P_x \log \frac{P_x}{Q_x} = \log \frac{Z(\beta)}{\tilde{Z}(\beta)} + \beta \langle H - \tilde{H} \rangle.$$

- mean field theory assumption $\langle H \rangle = \langle \tilde{H} \rangle$ (averages in the probability P_x) then get

$$\sum_x P_x \log \frac{P_x}{Q_x} = -\log \tilde{Z}(\beta) + \beta \langle \tilde{H} \rangle + \log Z(\beta) - \beta \langle H \rangle = \log \frac{Z(\beta)}{\tilde{Z}(\beta)}$$

- 1–parameter family of *commuting* Hamiltonians $H(\epsilon)$ analytic in ϵ with

$$H(\epsilon) = \tilde{H} + \epsilon \frac{\partial \tilde{H}}{\partial \epsilon}|_{\epsilon=0} + O(\epsilon^2)$$

- then have

$$\sum_x P_x H_x(\epsilon) \sim \sum_x P_x \tilde{H}_x + \epsilon \sum_x P_x \frac{\partial H_x(\epsilon)}{\partial \epsilon}|_{\epsilon=0}.$$

- generalized force corresponding to variable ϵ

$$L_x = -\frac{\partial H_x(\epsilon)}{\partial \epsilon}|_{\epsilon=0}$$

$$\langle L \rangle = \sum_x P_x L_x = \frac{1}{\beta} \frac{\partial}{\partial \epsilon} \log Z_\epsilon(\beta)|_{\epsilon=0},$$

where $Z_\epsilon(\beta) = \sum_x e^{-\beta H(\epsilon)}$

- for $P_x(\epsilon) = Z_\epsilon(\beta)^{-1} e^{-\beta H(\epsilon)}$ have

$$\log P_x(\epsilon) = -\log Z_\epsilon(\beta) - \beta(\tilde{H}_x + \epsilon L_x + O(\epsilon^2))$$

- Kullback–Leibler divergence $\sum_x P_x \log \frac{P_x}{P_x(\epsilon)} =$

$$\sum_x P_x \log P_x + \log Z_\epsilon(\beta) + \beta \sum_x P_x \tilde{H}_x + \epsilon\beta \sum_x P_x L_x + O(\epsilon^2) =$$

$$-(1-\beta\frac{\partial}{\partial\beta}) \log \tilde{Z}(\beta) + \log Z_\epsilon(\beta) + \beta \sum_x P_x \tilde{H}_x + \epsilon \frac{\partial}{\partial\epsilon} \log Z_\epsilon(\beta)|_{\epsilon=0}$$

$$+O(\epsilon^2) = \log \frac{Z_\epsilon(\beta)}{\tilde{Z}(\beta)} + \epsilon \frac{\partial}{\partial\epsilon} \log Z_\epsilon(\beta)|_{\epsilon=0} + O(\epsilon^2)$$

is completely described in terms of partition functions (up to higher order)

Information Geometry

- S. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, vol. 28. Springer, 1985.
- S. Amari, *Information Geometry and Its Applications*, Springer, 2016.
- S. Amari, H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society, 2007
- S. Amari, A. Chichoki, *Information Geometry derived of divergence functions*, Bull. Polish Acad. Sci. Tech. Ser., Vol.58 (2010), No. 1, 183–195
- F. Nielsen, *An Elementary Introduction to Information Geometry*, Entropy, 2020, 22, 1100, 61 pages

Fisher–Rao metric and Information Geometry

- Probability distributions depend on a space of parameters $P = P(\gamma) = (P_x(\gamma))$ with $\gamma = (\gamma_1, \ldots, \gamma_r)$
- Fisher–Rao information metric given by

$$g_{ij}(\gamma) := \sum_x P_x(\gamma) \frac{\partial \log P_x(\gamma)}{\partial \gamma_i} \frac{\partial \log P_x(\gamma)}{\partial \gamma_j}.$$

- for commuting Hamiltonians $H(\gamma)$

$$P_x(\gamma) = \frac{e^{-\beta H_x(\gamma)}}{Z_\gamma(\beta)}, \quad Z_\gamma(\beta) = \sum_x e^{-\beta H_x(\gamma)},$$

- generalized forces

$$L_{x,i} = -\frac{\partial H_x(\gamma)}{\partial \gamma_i},$$

- then Fisher-Rao metric

$$g_{ij}(\gamma) = \frac{\partial \log Z_\gamma(\beta)}{\partial \gamma_i} \frac{\partial \log Z_\gamma(\beta)}{\partial \gamma_j} + \beta^2 \sum_x P_x(\gamma) L_{x,i} L_{x,j}$$

Hessian and KL-divergence

- Fisher–Rao metric is Hessian matrix of Kullback–Leibler divergence

$$g_{ij}(\gamma_0) = \frac{\partial^2}{\partial\gamma_i \partial\gamma_j} KL(P(\gamma)|P(\gamma_0))|_{\gamma=\gamma_0}$$

- equivalently

$$g_{ab} = \sum_n P_n \, \partial_a \log P_n \, \partial_b \log P_n = \sum_n \frac{\partial_a P_n \, \partial_b P_n}{P_n}$$

$$= -\sum_n P_n \, \partial_a \partial_b \log P_n = \partial_a \partial_b KL(P|Q)|_{P=Q}$$

Amari-Chentsov tensor

- statistical manifold (M, g, A) manifold with Riemannian metric and a totally symmetric 3-tensor A (Amari-Chentsov tensor)

$$A_{abc} = A(\partial_a, \partial_b, \partial_c) = \langle \nabla_a \partial_b - \nabla^*_a \partial_b, \partial_c \rangle$$

Divergence functions and Bregman generators

- divergence function on manifold M: differentiable, non-negative real valued function $D(x|y)$, for $x, y \in M$, that vanishes only when $x = y$ and such that the Hessian in the x–coordinates evaluated at $y = x$ is positive definite
- divergence function determines a statistical manifold

$$g_{ab} = \partial_{x_a}\partial_{x_b} D(x|y)|_{y=x}$$

$$A_{abc} = (\partial_{x_a}\partial_{x_b}\partial_{y_c} - \partial_{x_c}\partial_{y_a}\partial_{y_b}) D(x|y)|_{y=x}$$

- this Amari-Chentsov tensor A_{abc} vanishes identically if divergence $D(x|y)$ is symmetric
- statistical manifold induced by Bregman generator if there is a potential Φ (locally)

$$D(x|y) = \Phi(x) - \Phi(y) - \langle \nabla\Phi(y), x - y\rangle$$

Statistical manifold of Shannon entropy

- space of probability distributions on a (finite) set with KL divergence and Fisher-Rao metric
- Amari-Chentsov 3-tensor given by

$$A_{abc} = \sum_i P_i \, \partial_a \log P_i \;\, \partial_b \log P_i \;\, \partial_c \log P_i = \sum_i \frac{\partial_a P_i \, \partial_b P_i \, \partial_c P_i}{P_i^2}$$

$$= (\partial_a \partial_b \partial_{c'} - \partial_c \partial_{a'} \partial_{b'}) KL(P|Q)|_{P=Q}$$

with a, b, c variation indices for P and a', b', c' for Q

- Bregman generator is the Shannon information

$$\Phi(P) = -S(P) = \sum_i P_i \log P_i$$

$$KL(P|Q) = \Phi(P) - \Phi(Q) - \langle \nabla \Phi(Q), P - Q \rangle$$

Bregman potential and dual coordinates

- as above divergence with Bregman potential

$$D(x|y) = \Phi(x) - \Phi(y) - \langle \nabla\Phi(y), x - y \rangle$$

- dual potential: Legendre transform

$$\Psi(\eta) = \sup_x \{ \langle x, \eta \rangle - \Phi(x) \}$$

- if Φ lower semicontinuous and convex then Legendre transform $\Psi = \Phi^\vee$ is involutive $(\Phi^\vee)^\vee = \Phi$
- in a dually flat manifold: dual affine coordinate systems $\eta = \nabla\Phi(x)$ and $x = \nabla\Psi(\eta)$

Linear case

- Special case: if dependence of P on parameters is linear

$$\partial_a \partial_b P = 0$$

- then the Amari-Chenstov tensor is the tensor of third derivatives of the Bregman potential

$$A_{abc} = \partial_a \, \partial_b \, \partial_c \, \Phi$$

- in case of Shannon entropy recover previous case of rank 3 tensor of Fisher-Rao metric

$$g_{ab} = \partial_a \partial_b \Phi, \quad A_{abc} = \partial_a \, \partial_b \, \partial_c \, \Phi$$

with potential the Shannon entropy

Divergence functions, flatness and decomposability

- given a divergence function $D(P|Q)$ additional requirements
  1. invariance under invertible transformations of variables
  2. decomposability: $D(P|Q) = \sum_i d(p_i, q_i)$ for some function d (e.g. $KL(P|Q) = -\sum_i p_i \log(q_i/p_i)$)
  3. flatness: Riemannian metric g (Hessian) and dual pair of connections ∇, ∇^* related by the metric, require these have vanishing curvature (dually flat structure)
- invariant + decomposable $\Leftrightarrow D(P|Q) = \sum_i p_i f(q_i/p_i)$ some differentiable convex function f
- only divergence satisfying all 3 properties is KL

Dual connections of a divergence function

- divergence $D(P|Q)$
- metric (pos def Hessian: quadratic term in expansion) $g^{(D)}$

$$D(P+\xi|P+\eta) \sim \frac{1}{2}\sum_{i,j} g^{(D)}_{ij}(P)\xi^i\eta^j + \text{ higher order terms}$$

- cubic term determines a connection

$$h^{(D)}_{ijk} = \partial_i g^{(D)}_{jk} + \Gamma^{(D)}_{jk,i}$$

- connection $\nabla^{(D)}$ with Christoffel symbols

$$\Gamma^{(D)}_{ij,k} = \Gamma^{(D)}_{ji,k}$$

- dual divergence $D^*(P|Q) := D(Q|P)$
- determines same metric $g^{(D^*)} = g^{(D)}$
- dual connection $\nabla^{(D^*)}$, dual to $\nabla^{(D)}$ under $g^{(D)}$
- duality condition for connections ∇, ∇^* under metric g: for any triple of vector fields V, W, Z

$$Z\, g(X,Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y)$$

Geodesics and Pythagorean relation

- given a triple $(g^{(D_f)}, \nabla^{(D_f)}, \nabla^{(D_f^*)})$ associated to a divergence (for some convex function f)

$$D_f(P|Q) = \sum_i P_i\, f(\frac{Q_i}{P_i})$$

- in the space of probabilities P have both $\nabla^{(D_f)}$-geodesics and $\nabla^{(D_f^*)}$-geodesics
- paths $\gamma(t)$ solutions of geodesic equation

$$\ddot{\gamma}(t)^k + \sum_{ij} \Gamma^k_{ij}(\gamma(t))\, \dot{\gamma}^i(t)\dot{\gamma}^j(t) = 0,$$

  with Γ^k_{ij} Christoffel symbols of corresponding connection
- P, Q, R three probability distributions: consider $\nabla^{(D)}$-geodesic from P to Q and $\nabla^{(D^*}$-geodesic from Q to R
- if these meet orthogonally at Q, then Pythagorean relation

$$D_f(P|R) = D_f(P|Q) + D_f(Q|R)$$

Dually flat structure and projection

- Pythagorean theorem: if $D(P|Q)$ defines a dually flat structure then

$$D(P|R) = D(P|Q) + D(Q|R)$$

  when P, Q, R form an orthogonal triangle, namely when geodesic paths PQ and QR orthogonal
- dual flat coordinate systems $x = (x^a)$ and $\eta = (\eta_a)$ related by Legendre transform
- take paths $\gamma(t) = (1-t)x(Q) + tx(R)$ and $\gamma^\vee(t) = (1-t)\eta(P) + t\eta(Q)$

$$\frac{d}{dt}\gamma = x(R) - x(Q), \quad \frac{d}{dt}\gamma^\vee = \eta(Q) - \eta(P)$$

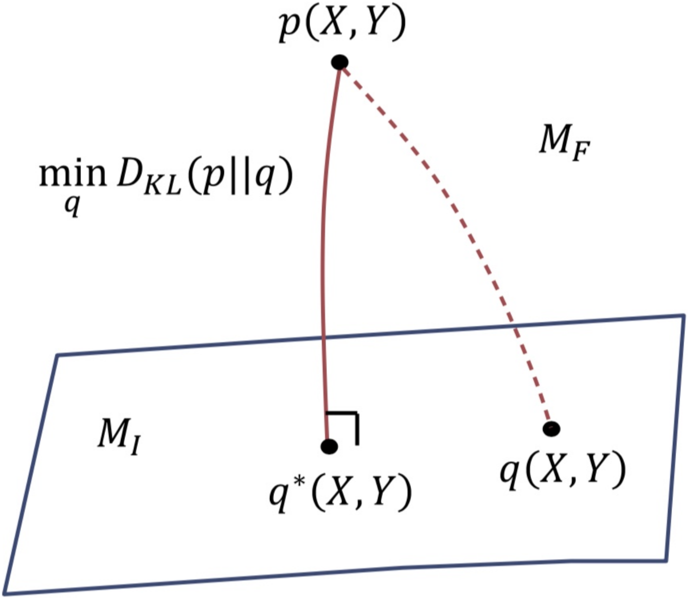
- the two paths are orthogonal in the metric

$$\langle \eta(Q) = \eta(P), x(R) - x(Q) \rangle = 0$$

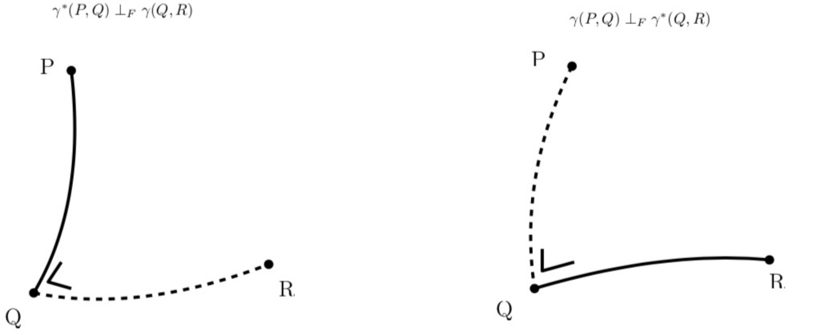
- this gives the Pythagorean relation above (Amari, 2016)

Projection theorem of Information Geometry (Amari)

- P and submanifold $P \notin \mathcal{M}$: $\nabla^{(D)}$-geodesic from P meets $\mathcal{M}$ orthogonally
$$Q^* = \mathrm{argmin}_{Q\in\mathcal{M}} D(P|Q)$$
- full space $\mathcal{M}_F$ of probabilities (depending on parameters), submanifold $\mathcal{M}_I$ satisfying given constraints
- given P minimization problem for KL divergence
$$KL(P|Q_{min}) = \min_{Q\in\mathcal{M}_I} KL(P|Q)$$
- argmin Q_{min} can be found by orthogonal projection of P onto $\mathcal{M}_I$
- orthogonal projection: dual geodesic (η-coords) connecting P and Q_{min} orthogonal to any tangent vector in $\mathcal{M}_I$ at Q_{min}
- if submanifold $\mathcal{M}_I$ itself flat, for any other point $Q \in \mathcal{M}_I$ and geodesics PQ_{min} and $Q_{min}Q \Rightarrow$ orthogonal triangle so
$$KL(P|Q) = KL(P|Q_{min}) + KL(Q_{min}, Q)$$
with $Q = Q_{min}$ minimizing lhs



$p(X,Y)$
$\min_q D_{KL}(p||q)$
M_F
M_I
$q^*(X,Y)$
$q(X,Y)$



$D(P:R) = D(P:Q) + D(Q:R)$

$D^*(P:R) = D^*(P:Q) + D^*(Q:R)$

dual Pythagorean theorems in a dually flat space

Frobenius Manifolds and Information Geometry

- Yu.I. Manin, *Frobenius Manifolds, Quantum Cohomology, and Moduli Spaces*, Colloquium Publications, Vol. 47, American Mathematical Society, 1999.
- C. Hertling, Yu.I. Manin, *Weak Frobenius manifolds*, Int. Math. Res. Notices 6 (1999), 277–286
- C. Hertling, M. Marcolli (Eds.), *Frobenius manifolds. Quantum cohomology and singularities*, Aspects of Mathematics, E36, Vieweg, 2004.
- N. Combe, Yu.I. Manin, *F-manifolds and geometry of information*, Bull. Lond. Math. Soc. 52 (2020), 777–792
- N. Combe, Ph. Combe, H. Nencka, *Frobenius Statistical Manifolds and Geometric Invariants*, Geometric Science of Information 2021, Lecture Notes in Computer Science, Vol.12829, pp. 565–573, Springer, 2021.
- N. Combe, Yu.I. Manin, M. Marcolli, *Geometry of Information: classical and quantum aspects*, arXiv:2107.08006

Frobenius Manifolds

- Frobenius manifold (M, g, Φ) a manifold M with *flat* metric g and potential Φ so that (in local affine coordinates) tensor $A_{abc} = \partial_a \partial_b \partial_c \Phi$ defines *associative*, commutative multiplication with unit
$$\partial_a \circ \partial_b = \sum_c A_{ab}{}^c \partial_c$$
equivalently $g(\partial_a \circ \partial_b, \partial_c) = A_{abc}$
- associativity condition for multiplication: WDVV (Witten–Dijkgraaf–Verlinde–Verlinde) nonlinear differential equations for potential Φ
$$A_{bce} g^{ef} A_{fad} = A_{bae} g^{ef} A_{fcd}, \quad \text{with } A_{abc} = \partial_a \partial_b \partial_c \Phi$$
- first structure connection (λ parameter)
$$\nabla_{\lambda, \partial_a} \partial_b = \lambda \sum_c A_{ab}{}^c \partial_c = \lambda \partial_a \circ \partial_b$$
- associativity of product and existence of potential equivalent to connection ∇_λ being flat

F-manifolds

- or "weak Frobenius manifold", introduced by Hertling–Manin
- F-manifold $(M, \circ, e)$ is a manifold with a commutative and associative multiplication $\circ$ on the tangent bundle TM with a unit vector field e
- F-manifold is a Frobenius manifold if $\circ$ induced by a flat metric g and a potential Φ
- for both F-manifolds and Frobenius can also include Euler vector field $E = \sum_a x_a \partial_a$
- difficulty of upgrading F-manifolds to Frobenius manifolds is *flatness* of the metric

Frobenius and F-manifolds in algebraic geometry

- notion of Frobenius manifold first introduced by Dubrovin in the mathematical formulation of TQFT
- B. Dubrovin, *Geometry of 2D topological field theories*, Integrable systems and quantum groups, Lecture Notes in Mathematics 1620, 120–348, Springer 1993.
- applications in singularity theory: Saito's Frobenius structure on moduli (unfolding) spaces of germs of isolated singularities of hypersurfaces
- Gromov–Witten invariants and quantum cohomology (Kontsevich–Manin, Barannikov–Kontsevich)

Frobenius manifolds and Gromov-Witten invariants

- $M = H^*(X, \mathbb{Z})$ with (X, ω) compact symplectic manifold
- Gromov-Witten invariants $\mathcal{I}^X_{g,n}(\gamma_{a_1}, \ldots, \gamma_{a_n})$, with $\gamma_{a_i} \in H^{d_i}(X, \mathbb{C})$, counts genus g pseudoholomorphic curves in X homological constraints imposed at n points of the curve
- γ_a homogeneous basis of $H^*(X, \mathbb{C})$ and t^a dual basis

$$g = \frac{1}{2} \sum_{a,b} \eta_{ab} dt^a dt^b$$

$$\eta_{ab} = \int \gamma_a \cup \gamma_b$$

metric from intersection product

- Frobenius manifold potential

$$\Phi = \sum_{n \geq 3} \frac{1}{n!} \sum_{a_1, \ldots, a_n} t^{a_1} \cdots t^{a_n} \mathcal{I}^X_{g,n}(\gamma_{a_1}, \ldots, \gamma_{a_n})$$

- $e = \partial / \partial t_0$

Cones and characteristic functions (Combe–Manin)

- X finite set, $\mathbb{R}^X$ real vector space spanned by X, probability simplex Δ_X (extremal points basis of $\mathbb{R}^X$)
- union of all oriented half-lines in $\mathbb{R}^X$ starting at $\underline{0}$: open convex cone
- more general convex cones: R fin dim real vector space and $V \subset R$ subset closed under addition and multiplication by positive reals, Δ_V simplex in V
- require that closure of V does not contain any real linear subspace of positive dimension
- characteristic function of convex cone V with dual $W \subset R^\vee$

$$V \ni x \mapsto \varphi_V(x) = \int_W e^{-\langle x, x' \rangle} dvol_W(x')$$

  with translation invariant volume form of $R^\vee$
- metric on V (hence on Δ_V) given by

$$g_{ij} = \frac{\partial^2}{\partial x^i \partial x^j} \log \varphi_V \qquad \Gamma^i_{jk} = \sum_l \frac{1}{2} g^{il} \partial_i \partial_j \partial_l \log \varphi_V$$

F-manifolds: flat structure and vector potential

- flat structure: torsionless flat connection ∇ and $\mathcal{T}_M^\nabla \subset \mathcal{T}_M$ with $\mathcal{T}_M^\nabla = \mathrm{Ker}\nabla$ flat vector fields
- flat F-manifold $(M, \circ, e, \nabla)$ flat connection with $\nabla e = 0$ and $\nabla + \alpha\circ$ flat for all $\alpha \in \mathbb{C}$
- then there is a vector potential $F = (F^i)$ with

$$\partial_j \circ \partial_k = c^i_{jk} \partial_i, \qquad c^i_{jk} = \partial_j \partial_k F^i$$

- equivalently for any $X, Y \in \mathcal{T}_M^\nabla$ and F vector potential

$$X \circ Y = [X, [Y, F]]$$

- associativity of $\circ$ quadratic differential constraint on F "oriented associativity equations"
- in Frobenius case vector potential comes from derivatives of scalar potential and metric
- see Yu.I.Manin, *F-manifolds with flat structure and Dubrovin's duality*, Advances in Mathematics 198 (2005) 5–26.

F-manifold structure on cones and statistical manifolds

- Δ_V with metric g with potential $\log \varphi_V$
- the WDVV equations for $A_{abc} = \partial_a \partial_b \partial_c \log \varphi_V$

$$A_{bce} g^{ef} A_{fad} = A_{bae} g^{ef} A_{fcd}$$

are trivially satisfied for this choice of potential $\log \varphi$

$$\varphi(X) = \int e^{-\langle X, Y \rangle} dv(Y) = \prod_{i=1}^n \int e^{-X_i Y_i} dY_i = \prod_{i=1}^n \varphi_i(X_i)$$

- some notation: for $\dim V = n$ and $I \subset \{1, \ldots n\}$

$$\varphi_I = \prod_{i \in I} \varphi_i, \quad \varphi_{I^c} = \prod_{i \notin I} \varphi_i$$

$$\psi_i = \int Y_i e^{-X_i Y_i} dY_i = -\partial_i \phi_i, \quad \psi_I = \prod_{i \in I} \psi_i$$

$$\psi_{i,k_i} = \int Y_i^{k_i} e^{-X_i Y_i} dY_i, \quad \psi_{I,\underline{k}} = \prod_{i \in I} \psi_{i,k_i}$$

- then have

$$\partial_a \log \varphi = \frac{\partial_a \varphi}{\varphi} = \frac{-\psi_a \, \varphi_{a^c}}{\varphi} = -\frac{\psi_a}{\varphi_a}$$

- metric

$$g_{ab} = \partial_a \partial_b \log \varphi = \delta_{ab} \partial_a \frac{-\psi_a}{\varphi_a} = \delta_{ab} \left(\frac{\psi_{a,2}}{\varphi_a} - \frac{\psi_a^2}{\varphi_a^2} \right)$$

positivity $\psi_{a,2} \varphi_a \geq \psi_a^2$ by Cauchy-Schwartz

$$\left(\int Y^2 e^{-XY} dY \right) \left(\int e^{-XY} dY \right) \geq \left(\int Y e^{-XY} dY \right)^2$$

- A_{abc} similarly just

$$A_{iii} = -\frac{\psi_{i,3}}{\varphi_i} + 3 \frac{\psi_i \psi_{i,2}}{\varphi_i^2} - 2 \frac{\psi_i^3}{\varphi_i^3}$$

and both sides of WDVV are $A_{aaa}^2 g^{aa}$ so F-manifold, not flat so not Frobenius

More general statistical manifolds and WDVV equation

- statistical manifold (M, g, A) is a Frobenius manifold if the Amari–Chentsov tensor satisfies

$$A_{bce} g^{ef} A_{fad} = A_{bae} g^{ef} A_{fcd}$$

- equivalent to equation for Bregman potential Φ

$$\begin{aligned}
&\langle \partial_e \nabla\Phi(P), \partial_a \partial_b P\rangle g^{ef} \langle \partial_f \nabla\Phi(P), \partial_c \partial_d P\rangle + \\
&\langle \partial_e \nabla\Phi(P), \partial_a \partial_b P\rangle g^{ef} \langle \partial_c \partial_d \nabla\Phi(P), \partial_f P\rangle + \\
&\langle \partial_a \partial_b \nabla\Phi(P), \partial_e P\rangle g^{ef} \langle \partial_f \nabla\Phi(P), \partial_c \partial_d P\rangle + \\
&\langle \partial_a \partial_b \nabla\Phi(P), \partial_e P\rangle g^{ef} \langle \partial_c \partial_d \nabla\Phi(P), \partial_f P\rangle = \\
&\langle \partial_e \nabla\Phi(P), \partial_a \partial_c P\rangle g^{ef} \langle \partial_f \nabla\Phi(P), \partial_b \partial_d P\rangle + \\
&\langle \partial_e \nabla\Phi(P), \partial_a \partial_c P\rangle g^{ef} \langle \partial_b \partial_d \nabla\Phi(P), \partial_f P\rangle + \\
&\langle \partial_a \partial_c \nabla\Phi(P), \partial_e P\rangle g^{ef} \langle \partial_f \nabla\Phi(P), \partial_b \partial_d P\rangle + \\
&\langle \partial_a \partial_c \nabla\Phi(P), \partial_e P\rangle g^{ef} \langle \partial_b \partial_d \nabla\Phi(P), \partial_f P\rangle
\end{aligned}$$

- give usual WDVV equation for Φ in the linear case where $A_{abc} = \partial_a \partial_b \partial_c \Phi$

Frobenius manifold structures?

- can use flat families ∇^{α} to improve to Frobenius? (Combe-Combe-Nencka)
- is there a deformation Φ_{α} of potential $\log \varphi_V$ that still satisfies WDVV but nontrivially?
- proposed version of "statistical Gromov–Witten invariants" (Combe-Combe-Nencka) related to higher mutual informations
- can these provide a Φ_{α} as in GW case with flat Frobenius structure?
- F-manifold structures on cones (and probability spaces) similar setting to F-manifold and Frobenius manifold structures for singularities and unfolding of singularities (Saito, Hertling, etc)

Hochschild cohomology

- A associative algebra over a field $\mathbb{K}$ (say $\mathbb{C}$)
- M an A-bimodule (ie two commuting actions $a(mb) = (am)b$)
- $C^0(A, M) = M$ and $C^n(A, M) = \mathrm{Hom}(A^{\otimes n}, M)$ (tensor over $\mathbb{K}$)
- Hochschild coboundary $\delta : C^n(A, M) \to C^{n+1}(A, M)$
  - $n = 0$ then $(\delta m)(a) = am - ma$ difference between left and right action
  - $n > 0$ then

$$(\delta f)(a_0, \ldots, a_n) = a_0 f(a_1, \ldots, a_n)$$

$$+\sum_{i=1}^{n} (-1)^i f(a_1, \ldots, a_i a_{i+1}, \ldots, a_n)$$

$$+(-1)^n f(a_0, \ldots, a_{n-1}) a_n$$

- satisfies $\delta^2 = 0$ so Hochschild cohomology $HH^*(A, M) = H^*(C^*(A, M), \delta) = \mathrm{Ker}(\delta)/\mathrm{Im}(\delta)$

- note how it generalizes case of groups: for an abelian group and $f : G^{\otimes n} \to \mathbb{Z}$ (trivial action on $\mathbb{Z}$)

$$(\delta f)(a_0, \ldots, a_n) = f(a_1, \ldots, a_n) +$$

$$\sum_{i=1}^{n-1} (-1)^i f(a_0, \ldots, a_i + a_{i+1}, \ldots, a_n) + (-1)^n f(a_0, \ldots, a_{n-1})$$

- special case $M = A^* = \mathrm{Hom}(A, \mathbb{K})$, then $\mathrm{Hom}(A^{\otimes n}, A^*) = \mathrm{Hom}(A^{\otimes(n+1)}, \mathbb{K})$ with $f(a_1, \ldots, a_n)(a_0) =: \varphi(a_0, a_1, \ldots, a_n)$ and $\delta f = b\varphi$ with

$$(b\varphi)(a_0, \ldots, a_{n+1}) = \sum_{i=0}^{n} (-1)^i \varphi(a_0, \ldots, a_i a_{i+1}, \ldots, a_n)$$

$$+(-1)^{n+1} \varphi(a_{n+1} a_0, a_1, \ldots, a_n)$$

- Example: $HH^0(A, M) = \{m \in M \,|\, am = ma, \forall a \in A\}$ in case of $M = A^*$ *traces*
  $HH^0(A, A^*) = \{\tau : A \to \mathbb{K} \,|\, \tau(ab) = \tau(ba), \forall ab \in A\}$
- Example: M-valued derivations modulo inner derivations (coboundaries)

$$HH^1(A, M) = \mathrm{Ker}(\delta)/\mathrm{Im}(\delta)$$

$$\mathrm{Ker}(\delta) = \{f : A \to M \,|\, f(ab) = af(b) + f(a)b, \forall a, b \in A\}$$

$$\mathrm{Im}(\delta) = \{f : A \to M \,|\, f(a) = [m, a] = ma - am\}$$

Mutual Information and Hochschild cohomology

- P. Baudot, D. Bennequin, *The homological nature of entropy*, Entropy 17 (2015) no. 5, 3253– 3318.
- mutual information $\mathcal{I}(X,Y) = S(X) + S(Y) - S(X,Y)$ with Shannon entropy

$$S(X) = -\sum_i \mathbb{P}(X = x_i) \log \mathbb{P}(X = x_i)$$

- for extensivity property use notation $S(X,Y) = S(X) + X \cdot S(Y)$ (think of as coboundary)
- more generally, random variables X_i, probability $\mathbb{P}$, and some entropy functional $F(X_1, \ldots, X_N; \mathbb{P})$

- define (left) action $X_0 \cdot F$ (and trivial right action)

$$X_0 \cdot F(X_1, \ldots, X_N; \mathbb{P}) := \sum_i \mathbb{P}(X_0 = x_i) \, F(X_1, \ldots, X_N | X_0 = x_i)$$

- then Hochschild coboundary

$$(\delta F)(X_0, \ldots, X_N; \mathbb{P}) = X_0 \cdot F(X_1 \ldots, X_N; \mathbb{P})$$

$$+ \sum_{i=1}^{N-1} F(X_1 \ldots, X_i X_{i+1}, \ldots, X_N; \mathbb{P}) + (-1)^N F(X_0, \ldots, X_N; \mathbb{P})$$

- also consider version where also left action trivial and corresponding $\tilde{\delta}$ Hochschild coboundary as above with first term just $F(X_1 \ldots, X_N; \mathbb{P})$

Shannon higher mutual informations

- for $J \subset \{1, \dots, N\}$ join X_J of the X_i random variables with $i \in J$ (composite system)

$$\mathcal{I}_N(X_1 \dots, X_N; \mathbb{P}) := \sum_{k=1}^{N} (-1)^{k-1} \sum_{\#J=k} S(X_J; \mathbb{P})$$

- then $\mathcal{I}_{2m} = \tilde{\delta}\delta \cdots \delta\tilde{\delta} S$ (with $(m-1)$ δ's and m $\tilde{\delta}$'s) and $\mathcal{I}_{2m+1} = -\delta\tilde{\delta}\delta \cdots \delta\tilde{\delta} S$ (with m δ's and m $\tilde{\delta}$'s)
- $\mathcal{I}_{2m}$ is a $\tilde{\delta}$-cocycle (coboundary) and $\mathcal{I}_{2m+1}$ is a δ-cocycle (coboundary)

More on cohomological information theory

- J.P. Vigneaux, *Generalized information structures and their cohomology*, arXiv:1709.07807
- J.P. Vigneaux, *A homological characterization of generalized multinomial coefficients related to the entropic chain rule*, arXiv:2003.02021
- J.P. Vigneaux, *Topology of statistical systems. A cohomological approach to information theory*, PhD Thesis, Institut de mathématiques de Jussieu, Université de Paris Diderot, 2019

Vigneaux's categorical formalism of information structures

- finite information structure: (S, M) pair of a thin category S (observables) and a functor $M : S \to \mathcal{F}$ to category of finite probability spaces
- category S: objects $X \in \mathrm{Obj}(S)$ random variables values in a finite probability space; a morphism $\pi : X \to Y$ if the random variable Y is coarser than X (values of Y determined by values of X)
- if there are morphisms $X \to Y$ and $X \to Z$ then $YZ = Y \wedge Z$ (random variable given by joint measurement of Y and Z) also an object of S.
- category S has a terminal object $\mathbf{1}$, random variable with value set $\{\star\}$ a singleton

Category of finite information structures

- functor $M : S \to \mathcal{F}$ maps a random variable X to the finite probability space given by its range of values M_X
- morphisms $\pi : X \to Y$ map to surjections $M(\pi) : M_X \to M_Y$
- value set $M_{X \wedge Y}$ is a subset of $M_X \times M_Y$
- category $\mathcal{IS}$ of finite information structures
    - objects pairs (S, M) as above
    - morphisms $\varphi : (S, M) \to (S', M,')$ pairs $\varphi = (\varphi_0, \varphi^{\#})$ of a functor $\phi_0 : S \to S'$ and a natural transformation $\phi^{\#} : M \to M' \circ \phi_0$ with properties:
    - $\phi_0(\mathbf{1}) = \mathbf{1}$
    - $\phi_0(X \wedge Y) = \phi_0(X) \wedge \phi_0(Y)$ whenever $X \wedge Y$ is an object in S
    - for all X the morphism $\phi^{\#}_X : M_X \to M'_{\phi_0(X)}$ is a surjection

products and coproducts

- category $\mathcal{IS}$ has finite products $(S \times S', M \times M')$ with objects pairs (X, X') of random variables with value set $M_X \times M'_{X'}$
- $\mathcal{IS}$ also has finite coproducts $(S \vee S', M \vee M')$ with objects $\mathrm{Obj}(S \vee S') = \mathrm{Obj}(S) \vee \mathrm{Obj}(S') = \mathrm{Obj}(S) \sqcup \mathrm{Obj}(S') / \mathbf{1}_S \sim \mathbf{1}_{S'}$ and value set M_X or $M'_{X'}$, if $X \in \mathrm{Obj}(S)$ or $X' \in \mathrm{Obj}(S')$

Probability functors

- *probability functor* $\mathcal{Q} : (S, M) \to \Delta$
- object X mapped to a simplicial set $\mathcal{Q}_X$ of probabilities on the set M_X
- $\mathcal{Q}_X$ is a *subset* of the simplex $\Pi(M_X)$ of all probability distributions on M_X
- morphisms $\pi : X \to Y$ mapped to morphism $\pi_* : \mathcal{Q}_X \to \mathcal{Q}_Y$ with
$$\pi_*(P)(y) = \sum_{x \in \pi^{-1}(y)} P(x)$$
- For each $X \in \mathrm{Obj}(S)$ there is a *semigroup*
$$\mathcal{S}_X = \{ Y \in \mathrm{Obj}(S) \,|\, \exists \pi : X \to Y \}$$
with product $Y \wedge Z$
- *semigroup algebra* $\mathcal{A}_X := \mathbb{R}[\mathcal{S}_X]$

Functor of measurable functions

- *contravariant functors* $\mathcal{F}(\mathcal{Q}) : (S, M) \to \mathrm{Vect}$
- assign to objects $X \in \mathrm{Obj}(S)$ and probabilities $P_X \in \mathcal{Q}_X$ the vector space of real valued (measurable) functions on (M_X, P_X)
- assigns to a morphism $\pi : X \to Y$ the map $\mathcal{F}(\mathcal{Q})(\pi) : f \mapsto f \circ \pi_*$
- action σ_α of the semigroup $\mathcal{S}_X$ on $\mathcal{F}(\mathcal{Q}_X)$ by

$$\sigma_\alpha(Y) : f \mapsto Y(f)(P_X) = \sum_{y \in E_Y \,:\, Y_* P_X(y) \neq 0} (Y_* P_X(y))^\alpha \, f(P_X|_{\pi^{-1}(y)})$$

  for $Y \in \mathcal{S}_X$ and for some arbitrary $\alpha > 0$
- $\mathcal{A}_X$-*module* structure $\mathcal{F}_\alpha(\mathcal{Q}_X)$ on $\mathcal{F}(\mathcal{Q}_X)$, determined by the semigroup action σ_α

Modules over sheaves of algebras

- category $\mathcal{A}$-Mod of modules over the sheaf of algebras $X \mapsto \mathcal{A}_X$
- $\mathcal{A}$-Mod is an *abelian category*
- sequence $\mathcal{B}_n(X)$ of free $\mathcal{A}_X$-modules generated by symbols $[X_1 \,|\, \dots \,|\, X_n]$ with $\{X_1, \dots, X_n\} \subset \mathcal{S}_X$
- with boundary maps $\partial_n : \mathcal{B}_n \to \mathcal{B}_{n-1}$ of Hochschild form

$$\begin{aligned} \partial_n[X_1 \,|\, \dots \,|X_n] = \; & X_1 \,[X_2 \,|\, \dots \,|\, X_n] \\ + \; & \sum_{k=1}^{n-1} (-1)^k [X_1 \,|\, \dots \,|\, X_k X_{k+1} \,|\, \dots \,|\, X_n] \\ + \; & (-1)^n [X_1 \,|\, \dots \,|X_{n-1}]. \end{aligned}$$

- modules $\mathcal{B}_n(X)$ give a projective bar resolution of the trivial $\mathcal{A}_X$-module

Functorial Hochschild cochain complex

- functor $C^\bullet(\mathcal{F}_\alpha(\mathcal{Q})) : (S, M) \to \mathrm{Ch}(\mathbb{R})$ to category of *cochain complexes*
- objects $X \in \mathrm{Obj}(S)$ mapped cochain complexes $(C^\bullet(\mathcal{F}_\alpha(\mathcal{Q}_X)), \delta)$

$$C^\bullet(\mathcal{F}_\alpha(\mathcal{Q}_X))^n = \mathrm{Hom}_{\mathcal{A}_X}(\mathcal{B}_n(X), \mathcal{F}_\alpha(\mathcal{Q}_X))$$

  natural transformations of functors $\mathcal{B}_n \to \mathcal{F}_\alpha(\mathcal{Q})$ compatible with $\mathcal{A}$-action
- coboundary δ given by Hochschild coboundary

$$\begin{aligned}\delta(f)[X_1 \,|\, \dots \,|\, X_{n+1}] = \;& X_1(f)[X_2 \,|\, \dots \,|\, X_{n+1}] \\ + \;& \sum_{k=1}^{n} (-1)^k f[X_1 \,|\, \dots \,|\, X_k X_{k+1} \,|\, \dots \,|\, X_{n+1}] \\ + \;& (-1)^{n+1} f[X_1 \,|\, \dots |X_n].\end{aligned}$$

Hochschild cohomology and entropy functionals

- complex $C^\bullet((S,M),\mathcal{F}_\alpha(\mathcal{Q})) := (C^\bullet(\mathcal{F}_\alpha(\mathcal{Q}_X)),\delta)$ with cohomology
$$H^\bullet((S,M),\mathcal{F}_\alpha(\mathcal{Q}))$$
- zeroth cohomology is $\mathbb{R}$ when $\alpha = 1$ and zero otherwise
- first cohomology: any non-trivial 1-cocycle is locally a multiple of the Tsallis entropy
$$S_\alpha[X](P) = \frac{1}{\alpha - 1}\left(1 - \sum_{x\in M_X} P(x)^\alpha\right),$$
for $\alpha \neq 1$ or of the Shannon entropy for $\alpha = 1$
- higher cohomologies represent all possible higher mutual information functionals

KL divergence

- information structures (S, M) and (S', M') and a joint random variable (X, Y) with values in a finite set $M_{XY} \subset M_X \times M'_Y$ with $X \in \mathrm{Obj}(S)$ and $Y \in \mathrm{Obj}(S')$
- pair of probability functors $\mathcal{Q} : (S, M) \times (S', M') \to \Delta$ and $\mathcal{Q}' : (S, M) \times (S', M') \to \Delta$,
- simplicial sets $\mathcal{Q}_{(X,Y)}$ and $\mathcal{Q}'_{(X,Y)}$ are subsimplicial sets of the full simplex $\Pi(M_{XY})$
- contravariant functor $\mathcal{F}^{(2)}(\mathcal{Q}, \mathcal{Q}') : (S, M) \times (S', M') \to \mathrm{Vect}$
- maps $(X, Y) \mapsto \mathcal{F}^{(2)}(X, Y)$ vector space of real valued (measurable) functions on simplicial set of probabilities $\mathcal{Q}_{(X,Y)} \times \mathcal{Q}'_{(X,Y)}$

- $X \in \mathrm{Obj}(S)$, $Y \in \mathrm{Obj}(S')$, the semigroup $\mathcal{S}_{(X,Y)}$ acts on $\mathcal{F}^{(2)}(X,Y)$ by

$$((X',Y')\cdot f)(P,Q) =$$
$$\sum_{(x',y')\in M_{X'Y'}} P(x',y')^{\alpha} Q(x',y')^{1-\alpha} f((P,Q)|_{(X',Y')=(x',y')})$$

  $(X',Y') \in \mathcal{S}_X$ and $(P,Q) \in \mathcal{Q}_{(X,Y)} \times \mathcal{Q}'_{(X,Y)}$ with $\{(X',Y')=(x',y')\} = \pi^{-1}(x',y')$ under surjection $\pi : M_{(X',Y')} \to M_{(X,Y)}$ determined by morphism $\pi : (X',Y') \to (X,Y)$
- $\mathcal{F}^{(2)}_{\alpha}(\mathcal{Q},\mathcal{Q}')$ denotes $\mathcal{F}^{(2)}(\mathcal{Q},\mathcal{Q}')$ with $\mathcal{A}$-module structure
- Kullback–Leibler divergence (Tsallis α-deformation) is a 1-cocycle in resulting chain complex $(C^{\bullet}(\mathcal{F}^{(2)}_{\alpha}(\mathcal{Q},\mathcal{Q}')),\delta)$