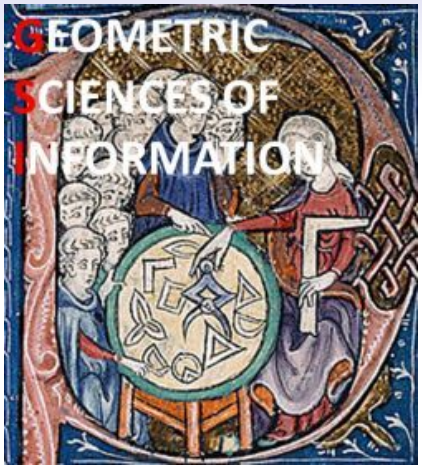


Nonparametric Information Geometry

http://www.giannidiorestino.it/GSI2013-talk.pdf

Giovanni Pistone

de Castro Statistics Initiative Collegio Carlo Alberto Moncalieri, Italy

August 30, 2013

Abstract

The differential-geometric structure of the set of positive densities on a given measure space has raised the interest of many mathematicians after the discovery by C.R. Rao of the geometric meaning of the Fisher information. Most of the research is focused on parametric statistical models. In series of papers by author and coworkers a particular version of the nonparametric case has been discussed. It consists of a minimalistic structure modeled according the theory of exponential families: given a reference density other densities are represented by the centered log likelihood which is an element of an Orlicz space. This mappings give a system of charts of a Banach manifold. It has been observed that, while the construction is natural, the practical applicability is limited by the technical difficulty to deal with such a class of Banach spaces. It has been suggested recently to replace the exponential function with other functions with similar behavior but polynomial growth at infinity in order to obtain more tractable Banach spaces, e.g. Hilbert spaces. We give first a review of our theory with special emphasis on the specific issues of the infinite dimensional setting. In a second part we discuss two specific topics, differential equations and the metric connection. The position of this line of research with respect to other approaches is briefly discussed.

References in

- GP, GSI2013 Proceedings. A few typos corrected in arXiv:1306.0480;
- GP, arXiv:1308.5312

- If μ_1, μ_2 are equivalent measures on the same sample space, a statistical model has two representations

$$L_1(x;\theta)\mu_1(dx) = L_2(x;\theta)\mu_2(dx).$$

- Fisher's score is a valid option

$$s(x;\theta) = \frac{d}{d\theta} \ln L_i(x;\theta), \quad i = 1, 2,$$

and $\mathrm{E}_\theta[s_\theta] = 0$.

- Each density q equivalent to p is of the form

$$q(x) = \frac{\mathrm{e}^{v(x)}p(x)}{\mathrm{E}_p[\mathrm{e}^v]} = \exp\left(v(x) - \ln\left(\mathrm{E}_p[\mathrm{e}^v]\right)\right)p(x),$$

where v is a random variable such that $\mathrm{E}_p[\mathrm{e}^v] < +\infty$.

- To avoid borderline cases, we actually require

$$\mathrm{E}_p\left[\mathrm{e}^{\theta v}\right] < +\infty, \quad \theta \in I \text{ open} \supset [0, 1].$$

- Finally, we require $\mathrm{E}_p[v] = 0$.

Plan

Part I Exponential manifold

Part II Vector bundles

Part III Deformed exponential

Part I Exponential manifold

Sets of densities

Definition

$\mathcal{P}^1$ is the set of real random variables f such that $\int f \, d\mu = 1$, $\mathcal{P}_{\geq}$ the convex set of probability densities, $\mathcal{P}_{>}$ the convex set of strictly positive probability densities:

$$\mathcal{P}_{>} \subset \mathcal{P}_{\geq} \subset \mathcal{P}^1$$

- We define the (differential) geometry of these spaces in a way which is meant to be a non-parametric generalization of Information Geometry
- We try to avoid the use of explicit parameterization of the statistical models and therefore we use a parameter free presentation of differential geometry.
- We construct a manifold modeled on an Orlicz space.
- We look for applications to applications intrisically non parametric, i.e. Statistical Physics, Information Theory, Optimization, Filtering.

Banach manifold

Definition

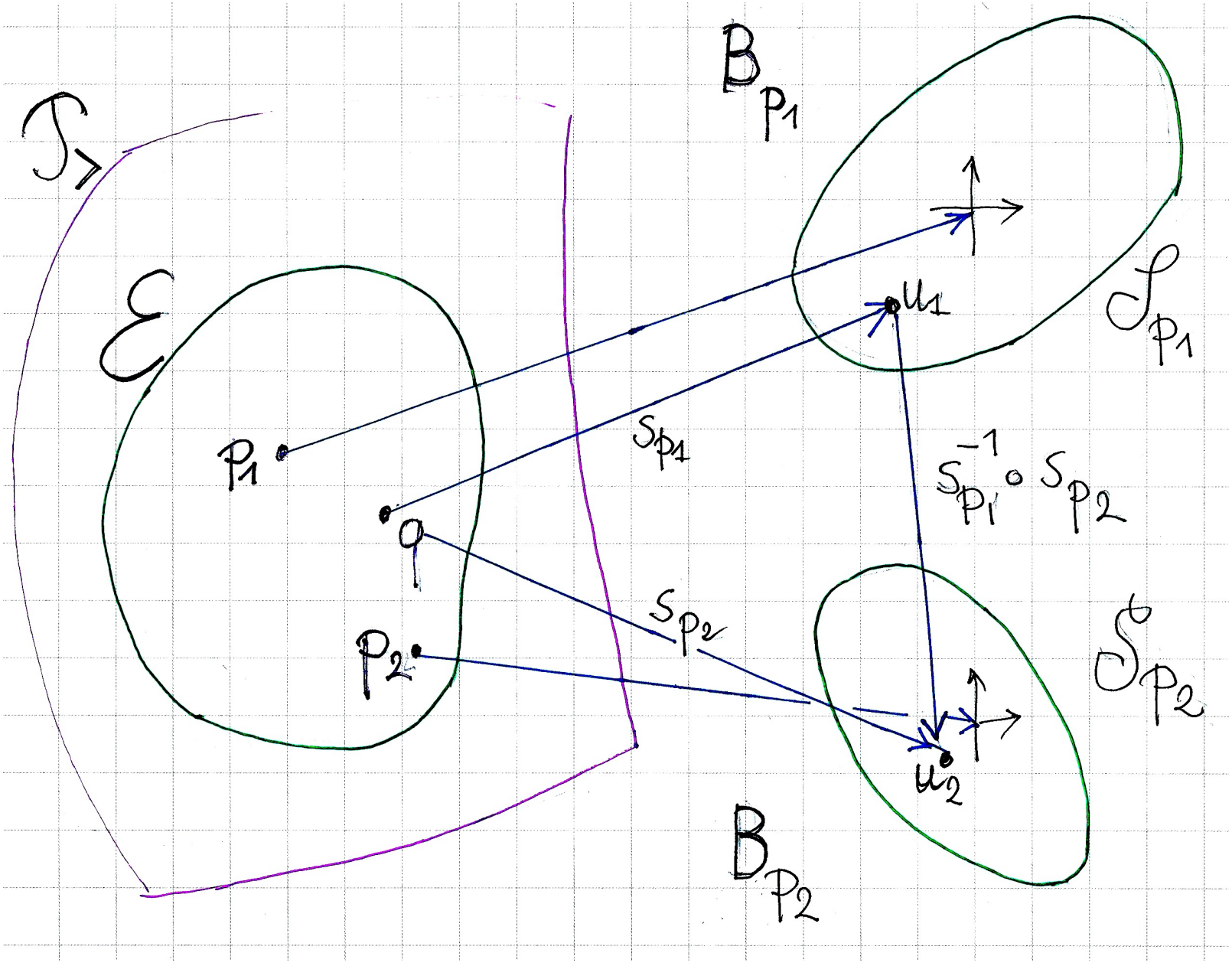
1. Let $\mathcal{P}$ be a set, $\mathcal{E} \subset \mathcal{P}$ a subset, B a Banach space. A 1-to-1 mapping $s\colon\ \mathcal{E} \to B$ is a chart if the image $s(\mathcal{E}) = \mathcal{S} \subset B$ is open.
2. Two charts $s_1\colon\ \mathcal{E}_1 \to B_1$, $s_2\colon\ \mathcal{E}_2 \to B_2$, are both defined on $\mathcal{E}_1 \cap \mathcal{E}_2$ and are compatible if $s_1(\mathcal{E}_1 \cap \mathcal{E}_2)$ is an open subset of B_1 and the change of chart mapping

$$s_2 \circ s_1^{-1}\colon\ s_1(\mathcal{E}_1 \cap \mathcal{E}_2) \xrightarrow{s_1^{-1}} \mathcal{E}_1 \cap \mathcal{E}_2 \xrightarrow{s_2} s_2(\mathcal{E}_1 \cap \mathcal{E}_2)$$

   is smooth.
3. An atlas is a set of compatible charts.

- Condition 2 implies that the model spaces B_1 and B_2 are isomorphic.
- In our case: $\mathcal{P} = \mathcal{P}_>$, the atlas has a chart s_p for each $p \in \mathcal{P}_>$ such that $s_p(p) = 0$ and two domains $\mathcal{E}_{p_1}$ and $\mathcal{E}_{p_2}$ are either equal or disjoint.

Charts on $\mathcal{P}_>$



Model space

Orlicz Φ-space

If $\phi(y) = \cosh y - 1$, the Orlicz Φ-space $L^{\Phi}(p)$ is the vector space of all random variables such that $\mathrm{E}_p[\Phi(\alpha u)]$ is finite for some $\alpha > 0$.

Properties of the Φ-space

1. $u \in L^{\Phi}(p)$ if, and only if, the moment generating function $\alpha \mapsto \mathrm{E}_p[\mathrm{e}^{\alpha u}]$ is finite in a neighborhood of 0.
2. The set $S_{\leq 1} = \left\{ u \in L^{\Phi}(p) \middle| \mathrm{E}_p[\Phi(u)] \leq 1 \right\}$ is the closed unit ball of a Banach space with norm

$$\|u\|_p = \inf \left\{ \rho > 0 \middle| \mathrm{E}_p \left[\Phi\left(\frac{u}{\rho} \right) \right] \leq 1 \right\}.$$

3. $\|u\|_p = 1$ if either $\mathrm{E}_p[\Phi(u)] = 1$ or $\mathrm{E}_p[\Phi(u)] < 1$ and $\mathrm{E}_p\left[\Phi\left(\frac{u}{\rho}\right)\right] = \infty$ for $\rho < 1$. If $\|u\|_p > 1$ then $\|u\|_p \leq \mathrm{E}_p[\Phi(u)]$. In particular, $\lim_{\|u\|_p \to \infty} \mathrm{E}_p[\Phi(u)] = \infty$.

Example: boolean state space

- In the case of a finite state space, the moment generating function is finite everywhere, but its computation can be challenging.
- Boolean case: $\Omega = \{+1, -1\}^n$, uniform density $p(x) = 2^{-n}$, $x \in \Omega$. A generic real function on Ω has the form $u(x) = \sum_{\alpha \in L} \hat{u}(\alpha) x^\alpha$, with $L = \{0, 1\}^n$, $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$, $\hat{u}(\alpha) = 2^{-n} \sum_{x \in \Omega} u(x) x^\alpha$.
- The moment generating function of u under the uniform density p is

$$\mathrm{E}_p \left[\mathrm{e}^{tu} \right] = \sum_{B \in \mathcal{B}(\hat{u})} \prod_{\alpha \in B^c} \cosh(t\hat{u}(\alpha)) \prod_{\alpha \in B} \sinh(t\hat{u}(\alpha)),$$

  where $\mathcal{B}(\hat{u})$ are those $B \subset \operatorname{Supp} \hat{u}$ such that $\sum_{\alpha \in B} \alpha = 0 \mod 2$.

-

$$\mathrm{E}_p \left[\Phi(tu) \right] = \sum_{B \in \mathcal{B}_0(\hat{u})} \prod_{\alpha \in B^c} \cosh(t\hat{u}(\alpha)) \prod_{\alpha \in B} \sinh(t\hat{u}(\alpha)) - 1,$$

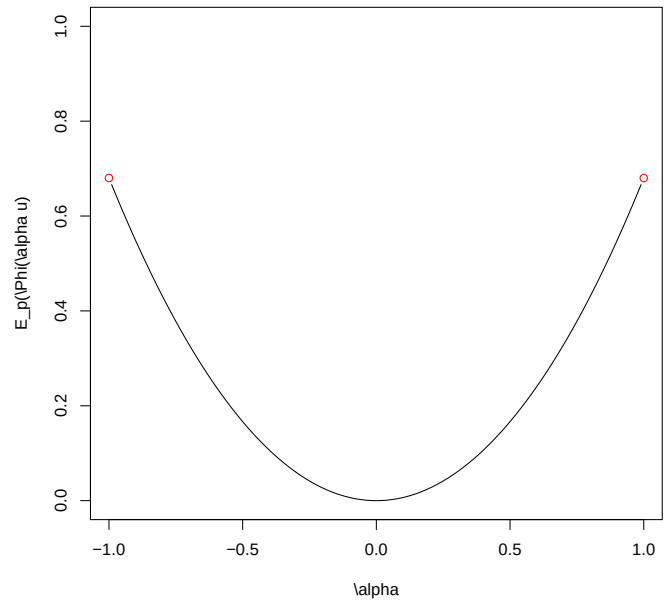
  where $\mathcal{B}_0(\hat{u})$ are those $B \subset \operatorname{Supp} \hat{u}$ such that $\sum_{\alpha \in B} \alpha = 0 \mod 2$ and $\sum_{\alpha \in \operatorname{Supp} \hat{u}} \alpha = 0$.

Example : the sphere is not smooth in general

- $p(x) \propto (a+x)^{-\frac{3}{2}} \mathrm{e}^{-x}$, $x, a > 0$.
- For the random variable $u(x) = x$, the function

$$\mathrm{E}_p\left[\Phi(\alpha u)\right] = \frac{1}{\mathrm{e}^a \, \Gamma\left(-\frac{1}{2}, a\right)} \int_0^\infty (a+x)^{-\frac{3}{2}} \frac{\mathrm{e}^{-(1-\alpha)x} + \mathrm{e}^{-(1+\alpha)x}}{2} \, dx - 1$$

is convex lower semi-continuous on $\alpha \in \mathbb{R}$, finite for $\alpha \in [-1, 1]$, infinite otherwise, hence not smooth.



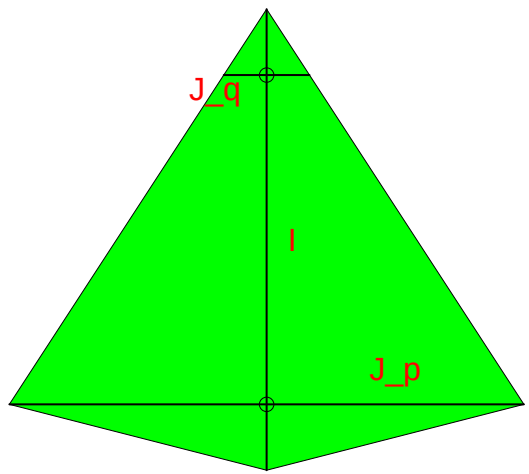
Isomorphism of L^Φ spaces

Theorem

$L^\Phi(p) = L^\Phi(q)$ as Banach spaces if $\int p^{1-\theta} q^\theta \, d\mu$ is finite on an open neighborhood I of $[0,1]$. It is an equivalence relation $p \smile q$ and we denote by $\mathcal{E}(p)$ the class containing p. The two spaces have equivalent norms

Proof.

Assume $u \in L^\Phi(p)$ and consider the convex function $C : (s,\theta) \mapsto \int \mathrm{e}^{su} p^{1-\theta} q^\theta \, d\mu$. The restriction $s \mapsto C(s,0) = \int \mathrm{e}^{su} p \, d\mu$ is finite on an open neighborhood J_p of 0; the restriction $\theta \mapsto C(0,\theta) = \int p^{1-\theta} q^\theta \, d\mu$ is finite on the open set $I \supset [0,1]$. hence, there exists an open interval $J_q \ni 0$ where $s \mapsto C(s,1) = \int \mathrm{e}^{su} q \, d\mu$ is finite. □



e-charts

Definition (e-chart)

For each $p \in \mathcal{P}_>$, consider the chart $s_p \colon \mathcal{E}(p) \to L_0^\Phi(p)$ by

$$\boxed{q \mapsto s_p(q) = \log\left(\frac{q}{p}\right) + D(p\|q) = \log\left(\frac{q}{p}\right) - \mathrm{E}_p\left[\log\left(\frac{q}{p}\right)\right]}$$

For $u \in L_0^\Phi(p)$ let $K_p(u) = \ln \mathrm{E}_p\left[\mathrm{e}^u\right]$ the cumulant generating function of u and let $\mathcal{S}_p$ the interior of the proper domain. Define

$$\boxed{e_p \colon \mathcal{S}_p \ni u \mapsto \mathrm{e}^{u - K_p(u)} \cdot p}$$

$e_p \circ s_p$ is the identity on $\mathcal{E}(p)$ and $s_p \circ e_p$ is the identity on $\mathcal{S}_p$.

Theorem (Exponential manifold)

$\{s_p \colon \mathcal{E}(p) | p \in \mathcal{P}_>\}$ *is an affine atlas on* $\mathcal{P}_>$.

Cumulant functional

- The divergence $q \mapsto D(p\|q)$ is represented in the chart centered at p by $K_p(u) = \log \mathrm{E}_p\left[\mathrm{e}^u\right]$, where $q = \mathrm{e}^{u-K_p(u)} \cdot p$, $u \in B_p = L_0^\Phi(p)$.
- $K_p : B_p \to \mathbb{R}_{\geq} \cup \{+\infty\}$ is convex and its proper domain $\mathrm{Dom}\,(K_p)$ contains the open unit ball of T_p.
- K_p is infinitely Gâteaux-differentiable on the interior $\mathcal{S}_p$ of its proper domain and analytic on the unit ball of B_p.
- For all $v, v_1, v_2, v_3 \in B_p$ the first derivatives are:

$$\begin{aligned} \mathrm{d}\, K_p u v &= \mathrm{E}_q\left[v\right] \\ \mathrm{d}^2\, K_p u(v_1, v_2) &= \mathrm{Cov}_q\left(v_1, v_2\right) \\ \mathrm{d}^3\, K_p u(v_1, v_2, v_3) &= \mathrm{Cov}_q(v_1, v_2, v_3) \end{aligned}$$

Change of coordinate

The following statements are equivalent:

1. $q \in \mathcal{E}(p)$;
2. $p \smile q$;
3. $\mathcal{E}(p) = \mathcal{E}(q)$;
4. $\ln\left(\frac{q}{p}\right) \in L^{\Phi}(p) \cap L^{\Phi}(q)$.

1. If $p, q \in \mathcal{E}(p) = \mathcal{E}(q)$, the change of coordinate

$$s_q \circ e_p(u) = u - \mathrm{E}_q[u] + \ln\frac{p}{q} - \mathrm{E}_q\left[\ln\frac{p}{q}\right]$$

   is the restriction of an affine continuous mapping.

2. $u \mapsto u - \mathrm{E}_q[u]$ is an affine transport from $B_p = L_0^{\Phi}(p)$ unto $B_q = L_0^{\Phi}(q)$.

Summary

$$p \smile q \implies \begin{array}{ccccccc} \mathcal{E}(p) & \xrightarrow{s_p} & \mathcal{S}_p & \xrightarrow{I} & B_p & \xrightarrow{I} & L^\Phi(p) \\ \| & & \downarrow{\scriptstyle s_q \circ s_p^{-1}} & & \downarrow{\scriptstyle d(s_q \circ s_p^{-1})} & & \| \\ \mathcal{E}(q) & \xrightarrow[s_q]{} & \mathcal{S}_q & \xrightarrow[I]{} & B_q & \xrightarrow[I]{} & L^\Phi(q) \end{array}$$

- If $p \smile q$, then $\mathcal{E}(p) = \mathcal{E}(q)$ and $L^\Phi(p) = L^\Phi(q)$.
- $B_p = L_0^\Phi(p)$, $B_q = L_0^\Phi(q)$
- $\mathcal{S}_p \neq \mathcal{S}_q$ and $s_q \circ s_p^{-1} \colon \mathcal{S}_p \to \mathcal{S}_q$ is affine

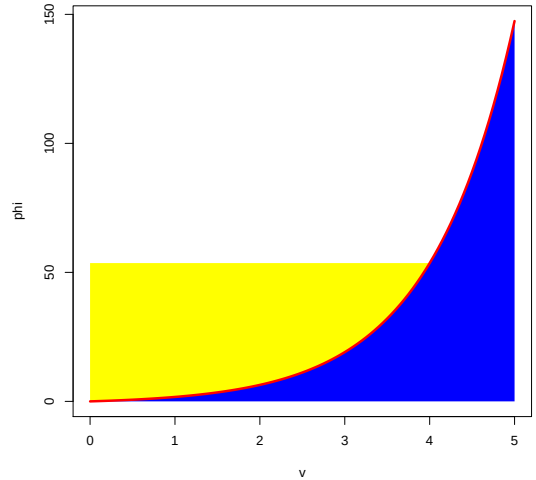
$$s_q \circ s_p^{-1}(u) = u - \mathrm{E}_q[u] + \ln\left(\frac{p}{q}\right) - \mathrm{E}_q\left[\ln\left(\frac{p}{q}\right)\right]$$

- The tangent application is $d(s_q \circ s_p^{-1})(v) = v - \mathrm{E}_q[v]$ (does not depend on p)

Duality

Young pair (N–function)

- $\phi^{-1} = \phi_*$,
- $\Phi(x) = \int_0^{|x|} \phi(u)\, du$
- $\Phi_*(y) = \int_0^{|y|} \phi_*(v)\, dv$
- $|xy| \leq \Phi(x) + \Phi_*(y)$



$\phi_*(u)$	$\phi(v)$	$\Phi_*(x)$	$\Phi(y)$
$\ln(1+u)$	$e^v - 1$	$(1+\lvert x\rvert)\ln(1+\lvert x\rvert) - \lvert x\rvert$	$e^{\lvert y\rvert} - 1 - \lvert y\rvert$
$\sinh^{-1} u$	$\sinh v$	$\lvert x\rvert \sinh^{-1}\lvert x\rvert - \sqrt{1+x^2} + 1$	$\cosh y - 1$

- $L^{\Phi_*}(p) \times L^{\Phi}(p) \ni (v,u) \mapsto \langle u, v\rangle_p = \mathsf{E}_p[uv]$
- $\left|\langle u, v\rangle_p\right| \leq 2\,\|u\|_{\Phi_*,p}\,\|v\|_{\Phi,p}$
- $(L^{\Phi_*}(p))' = L^{\Phi}(p)$ because $\Phi_*(ax) \leq a^2\Phi_*(x)$ if $a > 1$ (Δ_2).

m-charts

For each $p \in \mathcal{P}_>$, consider a second type of chart on $f \in \mathcal{P}^1$:

$$\boxed{\eta_p : f \to \eta_p(f) = \frac{f}{p} - 1}$$

Definition (Mixture manifold)

The chart is defined for all $f \in \mathcal{P}^1$ such that $f/p - 1$ belongs to ${}^*B_p = L_0^{\Phi_+}(p)$. The atlas $(\eta_p : {}^*\mathcal{E}(p))$, $p \in \mathcal{P}_>$ defines a manifold on $\mathcal{P}^1$.

If the sample space is not finite, such a map does not define charts on $\mathcal{P}_>$, nor on $\mathcal{P}_\geq$.

Example: $\mathsf{N}(\mu, \Sigma)$, $\det \Sigma \neq 0$ I

$$\mathcal{G} = \left\{ (2\pi)^{-\frac{n}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) \right) \middle| \mu \in \mathbb{R}^n, \Sigma \in \mathsf{Sym}_+^n \right\}.$$

$$\begin{aligned}
\ln\left(\frac{f(x)}{f_0(x)}\right) &= -\frac{1}{2}\ln(\det\Sigma) - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) + \frac{1}{2}x^T x \\
&= \frac{1}{2}x^T(I-\Sigma^{-1})x + \mu^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu - \frac{1}{2}\ln(\det\Sigma) \\
\mathsf{E}_{f_0}\left[\ln\left(\frac{f}{f_0}\right)\right] &= \frac{1}{2}(n - \operatorname{Tr}\Sigma^{-1}) - \frac{1}{2}\mu^T\Sigma^{-1}\mu - \frac{1}{2}\ln(\det\Sigma) \\
u(x) &= \ln\left(\frac{f(x)}{f_0(x)}\right) - \mathsf{E}_{f_0}\left[\ln\left(\frac{f}{f_0}\right)\right] \\
&= \frac{1}{2}x^T(I-\Sigma^{-1})x + \mu^T\Sigma^{-1}x - \frac{1}{2}(n - \operatorname{Tr}\Sigma^{-1}) \\
K_{f_0}(u) &= -\frac{1}{2}(n - \operatorname{Tr}\Sigma^{-1}) + \frac{1}{2}\mu^T\Sigma^{-1}\mu + \frac{1}{2}\ln(\det\Sigma)
\end{aligned}$$

Example: $\mathrm{N}(\mu, \Sigma)$, $\det \Sigma \neq 0$ II

$\mathcal{G}$ as a sub-manifold of $\mathcal{P}_>$

$$\mathcal{G} = \left\{ x \mapsto \mathrm{e}^{u(x)-K(u)} f_0(x) \middle| u \in \mathcal{H}_{1,2} \cap \mathcal{S}_{f_0} \right\}$$

- $\mathcal{H}_{1,2}$ is the Hemite space of total degree 1 and 2, that is the vector space generated by the Hermite polynomials

$$X_1, \ldots, X_n, (X_1^2 - 1), \ldots, (X_n^2 - 1), X_1 X_2, \ldots, X_{n-1} X_n$$

- If the matrix S, $S_{ii} = \beta_{ii} - \frac{1}{2}$, $S_{ij} = S_{ji} = \frac{\beta_{ij}}{2}$ is negative definite then $u \in \mathcal{S}_{f_0}$,

$$u(x) = \beta_1 x_1 + \cdots + \beta_n x_n + \beta_{11}(x_1^2 - 1) + \cdots + \beta_{nn}(x_n^2 - 1) + \beta_{12} x_1 x_2 + \cdots + \beta_{(n-1)n} x_{n-1} x_n$$

Part 2: Vector Bundles

Velocity

- Given a one dimensional statistical model $p_\theta \in \mathcal{E}(p)$, $\theta \in I$, I open interval, $0 \in I$, the representation in the chart centered at p of the e-manifold is $u_\theta = s_P(p_\theta)$,

$$p_\theta = e_p(u_\theta) = \mathrm{e}^{u_\theta - K_p(u_\theta)} \cdot p.$$

- As $\mathcal{E}(p) \subset {}^*\mathcal{E}(p)$, there is a representation in the m-manifold $\eta_\theta = \eta_p(p_\theta)$

$$\eta_\theta = \frac{p_\theta}{p} - 1 = \mathrm{e}^{u_\theta - K_p(u_\theta)} - 1$$

- To compute the velocity along a one-parameter statistical model in the s_p chart we use $\dot{u}_\theta$.
- To compute the velocity along a one-parameter statistical model in the η_p chart we use $\dot{\eta}_\theta$.

Relation between the two presentation

- We get in the first case

$$\dot{p}_\theta = p_\theta(\dot{u}_\theta - \mathsf{E}_\theta\left[\dot{u}_\theta\right])$$

so that

$$\dot{u}_\theta = \frac{\dot{p}_\theta}{p_\theta} - \mathsf{E}_p\left[\frac{\dot{p}_\theta}{p_\theta}\right] = {}^{\mathsf{e}}\mathbb{U}^{p}_{p(t)}\frac{\dot{p}_\theta}{p_\theta}$$

- In the second case we get

$$\dot{\eta}_p(\theta) = \frac{\dot{p}_\theta}{p} = \frac{p_\theta}{p}\frac{\dot{p}_\theta}{p_\theta} = {}^{\mathsf{m}}\mathbb{U}^{p_\theta}_{p}\frac{\dot{p}_\theta}{p_\theta}$$

Moving frame

- Both in the e-manifold and the m-manifold there is one chart centered at each density. The two representations $\dot{u}_\theta$ and $\dot{\eta}_p(\theta)$ are equal at $\theta = 0$ and are transported back to θ:

$$\frac{\dot{p}_\theta}{p_\theta} = \dot{u}_\theta - \mathrm{E}_\theta\left[\dot{u}_\theta\right] = \frac{p}{p_\theta}\dot{\eta}_\theta .$$

Fisher information

The random variable $\dot{p}_\theta / p_\theta$ is the Fisher score at θ of the one-parameter model p_θ. The Fisher information at θ is the L^2-norm of the velocity vector of the statistical model in the moving frame centered at θ. Moreover,

$$\mathrm{E}_\theta\left[\left(\frac{\dot{p}_\theta}{p_\theta}\right)^2\right] = \mathrm{E}_\theta\left[\left(\dot{u}_\theta - \mathrm{E}_\theta\left[\dot{u}_\theta\right]\right)\left(\dot{\eta}_\theta \frac{p}{p_\theta}\right)\right] = \mathrm{E}_p\left[\dot{u}_\theta \dot{\eta}_\theta\right].$$

Transport

- At each point $p \in \mathcal{P}_>$ of the statistical manifold is attached the vector space B_p. Similarly, ${}^*B_p$.
- The derivatives of the change of reference maps provide two isomorphic mappings

$$\text{e-transport} \quad {}^{\mathrm{e}}\mathbb{U}_p^q \colon B_p \ni u \mapsto u - \mathrm{E}_q[u] \in B_q$$

$$\text{m-transport} \quad {}^{\mathrm{m}}\mathbb{U}_p^q \colon {}^*B_p \ni v \mapsto \frac{p}{q} v \in {}^*B_q$$

- The two transport are adjoint to each other. If $u \in B_p$, $v \in {}^*B_q$.

$$\mathrm{E}_q\left[\left({}^{\mathrm{e}}\mathbb{U}_p^q u\right) v\right] = \mathrm{E}_q\left[(u - \mathrm{E}_q[u])v\right] = \mathrm{E}_p\left[u\left(\frac{q}{p}v\right)\right] = \mathrm{E}_p\left[u\left({}^{\mathrm{m}}\mathbb{U}_q^p v\right)\right]$$

- If $u \in \mathcal{S}_p$, $q = e_p(u) \in \mathcal{E}(p)$, $v, w \in B_p$,

$$d^2 K_p(u)(v,w) = \mathrm{Cov}_q(v,w) = \mathrm{E}_q\left[\left({}^{\mathrm{e}}\mathbb{U}_p^q u\right)\left({}^{\mathrm{e}}\mathbb{U}_p^q v\right)\right] = \mathrm{E}_p\left[\left({}^{\mathrm{m}}\mathbb{U}_q^p\, {}^{\mathrm{e}}\mathbb{U}_p^q v\right) w\right]$$

Tangent bundle

Definition

1. For each maximal exponential model $\mathcal{E}$ the tangent bundle $\mathrm{T}\,\mathcal{E}$ is the set
$$\mathrm{T}\,\mathcal{E} = \{(p, u) | p \in \mathcal{E}, u \in B_p\}$$
endowed with the atlas of charts
$$s_p \colon \mathrm{T}\,\mathcal{E} \ni (q, v) \mapsto (s_p(q), {}^{\mathrm{e}}\mathbb{U}^p_q v) \in \mathcal{S}_p \times B_p \subset B_p \times B_p$$
2. $T\mathcal{P}_> = \cup_{\mathcal{E}} \mathrm{T}\,\mathcal{E}$

Velocity field

1. If $p(t)$, $t \in I$ is a smooth curve in $\mathcal{E}$, then $(p(t), \delta p(t))$, $t \in I$ is a curve in $T\,\mathcal{E}$, $\delta p(t) = \dot{p}(t)/p(t)$.
2. If $E \colon \mathcal{E} \to \mathbb{R}$ and $E_p = E \circ e_p \colon \mathcal{S}_p \to \mathbb{R}$ is C^1, $p \in \mathcal{E}$, then
$$\frac{d}{dt} E(p(t)) = \frac{d}{dt} E_p(u(t)) = dE_p(u(t))\, {}^{\mathrm{e}}\mathbb{U}^p_{p(t)} \delta p(t)$$

e-covariant-derivative

Definition (Exponential covariant derivative)

1. Let $E\colon \mathcal{E} \to \mathbb{R}$ be smooth. For each vector field G of $\mathrm{T}\,\mathcal{E}$ the covariant derivative $D_G E\colon \mathcal{E} \to \mathbb{R}$ is defined by $D_G E(p) = dE_p(0)G(p)$.
2. If the linear functional $dE_p \in \mathcal{L}(B_p)$ is representable in ${}^*B_p$, the representative is the (natural) gradient ∇E and $D_G E(p) = \mathrm{E}_p\left[(\nabla E(p))\, G(p)\right]$
3. Let F be a vector field in $\mathrm{T}\,\mathcal{E}$. In the chart centered at p we have $F_p(u) = {}^{\mathrm{e}}\mathbb{U}^p_{e_p(u)} F \circ e_p(u)$. It is a mapping $F_p\colon \mathcal{S} \to B_p$ we assume differentiable. The covariant derivative $D_G F$ is the vector field of $\mathrm{T}\,\mathcal{E}$ defined by $D_G F(p) = dF_p(0)G(p)$.

Example: KL divergence

- For each $q \in \mathcal{E}(p_0)$, $E\colon\ \mathcal{E}(p_0) \ni q \mapsto \mathrm{D}(q\|p_0) = \mathrm{E}_q\left[\ln\left(\frac{q}{p_0}\right)\right]$.
- In the chart centered at any $p \in \mathcal{E}(p_0)$, $q = \mathrm{e}^{u - K_p(u)} \cdot p$, $p_0 = \mathrm{e}^{u_0 - K_p(u_0)} \cdot p$,

  $$E_p(u) = \mathrm{E}_q\left[u - K_p(u) - u_0 + K_p(u_0)\right] = dK_p(u)(u-u_0) - K_p(u) + K_p(u_0)$$

- $dE_p(0)v = d^2K_p(0)(-u_0, v) + \cancel{dK_p(0)v} - \cancel{K_p(0)v} = -E_p(u_0 v)$
-

$$D_G E(p) = \mathrm{E}_p\left[\left(\ln\left(\frac{p}{p_0}\right) - \mathrm{E}_p\left[\ln\left(\frac{p}{p_0}\right)\right]\right) G(p)\right]$$
$$= \mathrm{E}_p\left[\left(\ln\left(\frac{p}{p_0}\right) - E(p)\right) G(p)\right]$$

- $\nabla E(p) = \ln\left(\frac{p}{p_0}\right) - E(p) \in B_p$

Pre-tangent bundle

Definition

1. For each maximal exponential model $\mathcal{E}$ the pretangent bundle ${}^*\mathrm{T}\,\mathcal{E}$ is the set

$$ {}^*\mathrm{T}\,\mathcal{E} = \{(p, v) | p \in \mathcal{E}, v \in {}^*B_p\} $$

endowed with the atlas of charts

$$ s_p \colon {}^*\mathrm{T}\,\mathcal{E} \ni (q, v) \mapsto (s_p(q), {}^{\mathrm{m}}\mathbb{U}_q^p v) \in \mathcal{S}_p \times {}^*B_p \subset B_p \times {}^*B_p $$

2. ${}^*\mathrm{T}\,\mathcal{P}_> = \cup_{\mathcal{E}}\, {}^*\mathrm{T}\,\mathcal{E}$

m-covariant-derivative

Definition (Mixture covariant derivative)

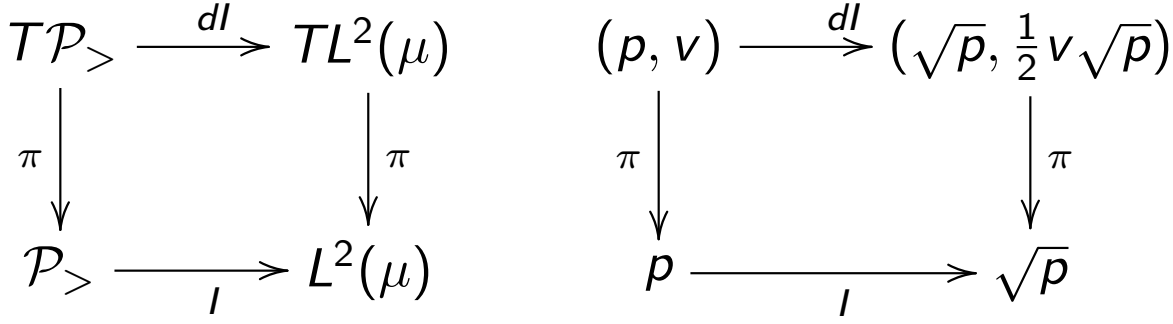
Let F be a vector field in ${}^*\mathrm{T}\,\mathcal{E}$. In the chart centered at p we have $F_p(u) = {}^{\mathrm{m}}\mathbb{U}^p_{e_p(u)} F \circ e_p(u)$. It is a mapping $F_p : \mathcal{S}_{\rightarrow} {}^*B_p$ that we assume to be differentiable. Let G be a vector field in $\mathrm{T}\,\mathcal{E}$. The covariant derivative $D_G F$ is the vector field of ${}^*\mathrm{T}\,\mathcal{E}$ defined by $D_G F(p) = dF_p(0) G(p)$.

Theorem

If G, H are smooth vector fields in $\mathrm{T}\,\mathcal{E}$ and F is a smooth vector field in ${}^*\mathrm{T}\,\mathcal{E}$, then $\langle G, F \rangle (p) = \mathrm{E}_p [G(p) F(p)]$ is a smooth real function on $\mathcal{E}$ and

$$D_H \langle G, F \rangle = \langle D_H G, F \rangle + \langle G, D_H F \rangle$$

Embedding of $\mathcal{P}_>$ into $L^2(\mu)$



- For each $p \in \mathcal{P}_>$, $\int (\sqrt{p})^2 \, d\mu = 1$.
- The embedding $I : p \mapsto \sqrt{p}$ is represented in the chart centered at p by

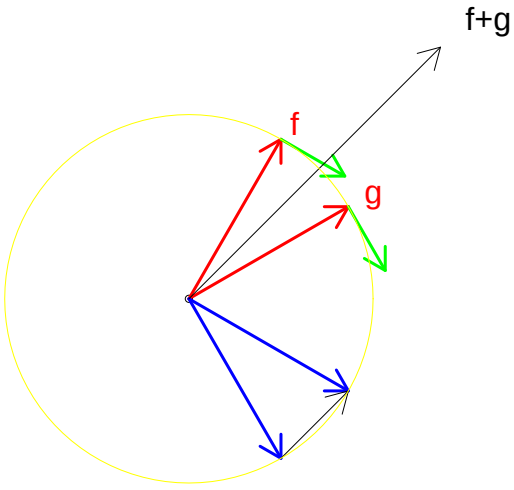
$$I_p : u \mapsto \mathrm{e}^{u - K_p(u)} \; p \mapsto \mathrm{e}^{\frac{1}{2}u - \frac{1}{2}K_p(u)} \; \sqrt{p}$$

- The mapping $u \mapsto \mathrm{e}^{\frac{1}{2}u - \frac{1}{2}K_p(u)} \in L^2(p)$ is analytic around 0 with differential

$$d(u \mapsto \mathrm{e}^{\frac{1}{2}u - \frac{1}{2}K_p(u)})v \Big|_{u=0} = \frac{1}{2}v \in B_p$$

- As $\int \left(\frac{1}{2}v\sqrt{p}\right)^2 = \frac{1}{4}\mathrm{E}_p\left[v^2\right]$, the embedding I is smooth, with differential $dI(p)v = \frac{1}{2}v\sqrt{p}$
- $\frac{1}{2}v\sqrt{p}$ belongs to the tangent at $\sqrt{p}$ of the sphere.

Parallel transport on the sphere



- $S_\mu = \left\{ f \in L^2(\mu) \middle| \|f\|_2 = 1 \right\}$ is the unit sphere of $L^2(\mu)$. It is an Hilbert manifold, sub-manifold of $L^2(\mu)$.
- $H_f = \left\{ h \in L^2(\mu) \middle| \int fh \; d\mu = 0 \right\}$ is the tangent space of S_μ at $f \in S_\mu$.
- Given $f, g \in S_\mu$, consider the circle $S_\mu \cap \mathrm{Span}\,(f, g)$.

- $U_f^g : H_f \to H_g$ is an isometric isomorphism if $f, g \in S$ and

$$U_f^g h = h - \left(1 + \int fg \; d\mu \right)^{-1} (f + g) \int gh \; d\mu$$

- If $h \perp f, g$, then $U_f^g h = h$, otherwise it is rotated.

Hilbert transport

$$\begin{array}{ccc} H_{\sqrt{p}} & \xrightarrow{U_{\sqrt{p}}^{\sqrt{q}}} & H_{\sqrt{q}} \\ {\scriptstyle v\mapsto \frac{1}{2}v\sqrt{p}}\downarrow & & \downarrow{\scriptstyle v\mapsto \frac{1}{2}v\sqrt{q}} \\ H_p = L_0^2(p) & \xrightarrow[\mathbb{U}_p^q]{} & H_q = L_0^2(q) \end{array}$$

$$\mathbb{U}_p^q u = \sqrt{\frac{p}{q}}u - \left(1 + \mathsf{E}_q\left[\sqrt{\frac{p}{q}}\right]\right)^{-1}\left(1 + \sqrt{\frac{p}{q}}\right)\mathsf{E}_q\left[\sqrt{\frac{p}{q}}u\right]$$

Proposition

1. The mapping $\mathbb{U}_p^q$ of is an isometry of H_p onto H_q.
2. In particular, if $u_1, \ldots, u_n$ is an orthonormal system in H_p, then $\mathbb{U}_p^q u_i$ in an orthonormal system in H_q.
3. $\mathbb{U}_q^p \mathbb{U}_p^q u = u$ if $u \in H_p$. Generally, $\mathbb{U}_q^r \mathbb{U}_p^q \neq \mathbb{U}_p^r$.

Hilbert bundle

Definition

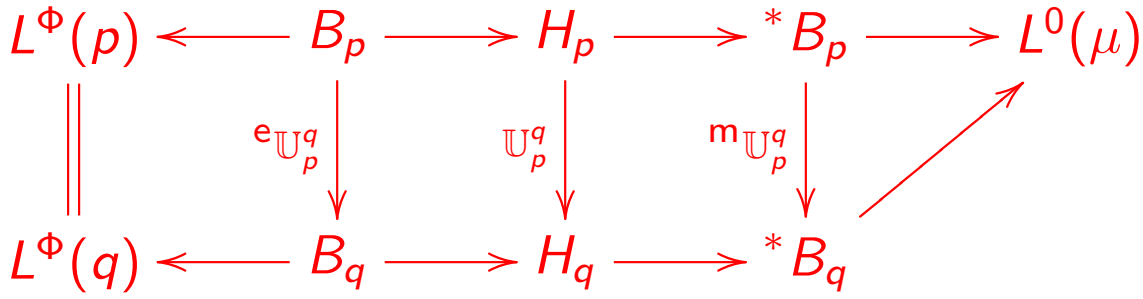
1. For each maximal exponential model $\mathcal{E}$ the Hilbert bundle $H\mathcal{E}$ is the set
$$H\mathcal{E} = \{(p, v) | p \in \mathcal{E}, v \in H_p\}$$
endowed with the atlas of charts
$$s_p : H\mathcal{E} \ni (q, v) \mapsto (s_p(q), \mathbb{U}_q^p v) \in \mathcal{S}_p \times H_p \subset B_p \times H_p$$
2. $H\mathcal{P}_> = \cup_{\mathcal{E}} H\mathcal{E}$

Hilbert bundle vs tangent and pretangent bundle

$$\mathbb{U}_q^p v = \sqrt{\frac{q}{p}} v - \left(1 + \mathrm{E}_p\left[\sqrt{\frac{q}{p}}\right]\right)^{-1} \left(1 + \sqrt{\frac{q}{p}}\right) \mathrm{E}_p\left[\sqrt{\frac{q}{p}} v\right]$$

- Assume $q, p \in \mathcal{E}$, $q = \mathrm{e}^{u - K_p(u)} \cdot p$. Define $p(t) = \mathrm{e}^{tu - K_p(tu)} \cdot p$, $p(1) = q$. Assume $v \in B_q \subset H_q$.
- $\sqrt{\frac{q}{p}} = \mathrm{e}^{\frac{1}{2}u - \frac{1}{2}K_p(u)} = \mathrm{e}^{-\frac{1}{2}K_p(u) + K_p(u/2)} \frac{p(1/2)}{p}$.
- $\mathrm{E}_p\left[\sqrt{\frac{q}{p}}\right] = \mathrm{e}^{-\frac{1}{2}K_p(u) + K_p(u/2)}$.
- $\sqrt{\frac{q}{p}} v = \mathrm{e}^{-\frac{1}{2}K_p(u) + K_p(u/2)} \frac{p(1/2)}{p} v =$
  $\mathrm{e}^{-\frac{1}{2}K_p(u) + K_p(u/2)} \left({}^{\mathrm{m}}\mathbb{U}^p_{p(1/2)} \, {}^{\mathrm{e}}\mathbb{U}^{p(1/2)}_q v + \frac{p(1/2)}{p} \mathrm{E}_{p(1/2)}[v]\right)$.
- $\mathrm{E}_p\left[\sqrt{\frac{q}{p}} v\right] = \mathrm{e}^{-\frac{1}{2}K_p(u) + K_p(u/2)} \mathrm{E}_{p(1/2)}[v]$.
- Now we compute $\left\langle \mathbb{U}_q^p v, w \right\rangle_p$, $w \in B_p$

Metric derivative in $H\mathcal{P}_>$



Definition

Let G, F be vector fields in $H\mathcal{P}_>$, i.e. $F(p), G(p) \in H_p\mathcal{P}_>$. We define $D_G F$ to be the vector field defined by $D_G F(p) = \frac{d}{dt}\mathbb{U}^p_{p(t)} F(t)\Big|_{t=0}$, where $p(t)$ is a curve such that $p(0) = p$ and $\delta p(0) = G(p)$.

Theorem

1. *$D_G F$ is a covariant derivative.*
2. *Let F_1, F_2, G be vector fields in $H\mathcal{P}_>$ such that the ordinary products GF_1 and GF_2 are vector fields in $H\mathcal{P}_>$ and*

$$D_G \mathrm{E}_p\left[F_1(p)F_2(p)\right] = \mathrm{E}_p\left[D_G F_1(p)F_2(p)\right] + \mathrm{E}_p\left[F_1(p)D_G F_2(p)\right],$$

Vector field

Definition

A **vector field** F of the the m-bundle ${}^*T(p)$, $p \in \mathcal{P}_>$, is a mapping which is defined on some domain $D \subset \mathcal{P}_>$ and it is a section of the m-bundle, that is $F(p) \in {}^*T(p)$, for all $p \in D \subset \mathcal{P}_>$.

Example

1. For a given $u \in T_p$ and all $q \in \mathcal{E}(p)$

$$F : q \mapsto u - \mathrm{E}_q[u]$$

2. For all strictly positive density $q \in \mathcal{P}_>(\mathbb{R}) \cap C^1(\mathbb{R})$

$$F : q \mapsto \frac{q'}{q}$$

3. For all strictly positive $q \in \mathcal{P}_>(\mathbb{R}) \cap C^2(\mathbb{R})$

$$F : q \mapsto q''/q$$

Evolution equation

Definition

A one-parameter statistical model in $\mathcal{P}_>$, $p(\theta)$, $\theta \in I$, solves the evolution equation associated to the vector field F if $p(\theta) = \mathrm{e}^{u(\theta) - K_p(u(\theta))} \cdot p$:

1. the curve $\theta \mapsto u(\theta) \in T(p)$ is continuous in L^2;
2. the curve $\theta \mapsto p(\theta)/p - 1 \in {}^*T(p)$ is continuously differentiable;
3. for all $\theta \in I$ it holds

$$\boxed{\frac{\dot{p}(\theta)}{p(\theta)} = F(p(\theta))}$$

Evolution equation in the moving frame

Theorem (???)

Assume F is locally maximal monotone. Then the equation $\dot{p}/p + F(p) = 0$ has a solution which is unique.

- The evolution equation above is written with respect to the moving frame at p_θ because $\dot{p}(\theta)/p(\theta)$ is the representation of the velocity vector in ${}^*T(p(\theta))$.
- However, with respect to a fixed frame at p, we should have written

$$\begin{cases} \dot{u}_\theta = F(p(\theta)) - \mathrm{E}_p\left[F(p(\theta))\right] & \text{e-connection, assuming } \dot{u}_\theta \in T_{p_\theta} \\ \dot{l}_\theta = \frac{p}{p(\theta)} F(p(\theta)) & \text{m-connection} \end{cases}$$

Exponential family

Example

1. The exponential model $p_\theta = \mathrm{e}^{\theta F}/\Lambda(\theta)$ is a solution of the equation $\frac{\dot{p}_\theta}{p_\theta} = F - \mathrm{E}_{p_\theta}[F]$.
2. The second example follows by considering $\Omega = \mathbb{R}$ and taking for domain the class of C^2 positive densities q such that $F(q) = -q'/q \in {}^*T(f)$. We can therefore consider the evolution equation $\dot{p}_\theta/p_\theta = -F(p_\theta)$.

   Given any f in the domain, the statistical model $p_\theta(x) = f(x-\theta)$ is such that the score is

$$\frac{\dot{p}_\theta(x)}{p_\theta(x)} = -\frac{f'(x-\theta)}{p(f-\theta)} = F(p(\cdot-\theta))(x)$$

   and therefore is a solution of the evolution equation. The classical Pearson classes of distributions are related to this equation.

Heat equation

The heat equation $\frac{\partial}{\partial t}p(t,x) - \frac{\partial^2}{\partial x^2}p(t,x) = 0$ is an interesting example of evolution equation in $\mathcal{M}_>$. In fact, we can consider the vector field

$$F(p)(x) = \frac{\frac{\partial^2}{\partial x^2}p(x)}{p(x)}$$

Upon division of both sides of of the heat equation by $p(t,x)$, we obtain an equation of our type, whose solution is the solution of the heat equation. Moreover, the heat equation has a variational form. For each $v \in D$

$$\mathrm{E}_p\left[F(p)v\right] = \int p''(x)v(x)\,dx = -\int p'(x)v'(x)\,dx = -\mathrm{E}_p\left[\frac{p'}{p}v'\right]$$

from which the weak form of the evolution equation follows.
as

$$\mathrm{E}_{p_\theta}\left[\frac{\dot{p}_\theta}{p_\theta}v\right] + \mathrm{E}_{p_\theta}\left[F_0(p_\theta)v\right] = 0 \quad v \in D$$

where F_0 is the vector field associated to the translation model.

Example: Decision geometry (Dawid&Lauritzen) I

If the sample space is $\mathbb{R}$ and $p, q \in \mathcal{P}_>$, write $q = \mathrm{e}^{u - K_p(u)} \cdot p$, so that

$$\log q - \log p = u - K_p(u).$$

Assume u belongs to the Sobolev space

$$W^{\Phi,1} = \left\{ u \in L_0^\Phi(p) \middle| \nabla u \in L_0^\Phi(p) \right\}.$$

It follows

$$\begin{aligned} d(p,q) &= \frac{1}{4} \mathsf{E}_p \left[\| \nabla \log q - \nabla \log p \|^2 \right] \\ &= \frac{1}{4} \mathsf{E}_p \left[\| \nabla u \|^2 \right]. \end{aligned}$$

Example: Decision geometry (Dawid&Lauritzen) II

For $u, v \in W^{\Phi,1}$ we have a bilinear form

$$\begin{aligned}\langle u, v\rangle_p &= \mathsf{E}_p\left[\nabla u \nabla v\right] = \int u_x(x) v_x(x) p(x) dx \\ &= -\int \nabla(u_x(x) p(x)) v(x) dx \\ &= -\int (\Delta u(x) p(x) + \nabla u(x) \nabla p(x)) v(x) dx \\ &= \mathsf{E}_p\left[(-\Delta u - \nabla \log p \nabla u) v\right]\end{aligned}$$

We have

$$\mathsf{E}_p\left[\nabla u \nabla v\right] = \mathsf{E}_p\left[F_p u v\right], \qquad F_p u \in {}^*W^{\Phi,1}$$

i.e a classical setting for evolution equations $\partial_t u_t = F_p(u_t)$.

Example: binary case (L. Malagò)

Consider the optimization of $\theta \mapsto \mathrm{E}_\theta[F]$ along a binary exponential family

$$p_\theta = \mathrm{E}\left(\sum_{j=1}^{d} \theta_j T_j - K(\theta)\right) \cdot p, \quad T_j^2 = 1.$$

- $\partial_j \mathrm{E}_\theta[F] = \mathrm{Cov}_\theta(F, T_j) = \mathrm{E}_\theta[FT_j] - \mathrm{E}_\theta[F]\,\mathrm{E}_\theta[T_j]$
- θ is is a critical point if $\mathrm{Cov}_\theta(F, T_j) = 0$, $j = 1, \ldots, d$. This is not possible if F is a linear combination of the T_j's or if the remainder after projection on the tangent space is small enough.
- At the critical point the Hessian matrix is

  $$\partial_i \partial_j = \mathrm{Cov}_\theta(F, T_i T_j)$$

  which is not zero if F is a linear combination of the T_j's and the interactions $T_i T_j$'s.
- The diagonal elements of the Hessian matrix are $\partial_i^2 = \mathrm{Cov}_\theta\left(F, T_i^2\right) = \mathrm{Cov}_\theta(F, 1) = 0$. The Hessian matrix is not sign-defined at the critical point.

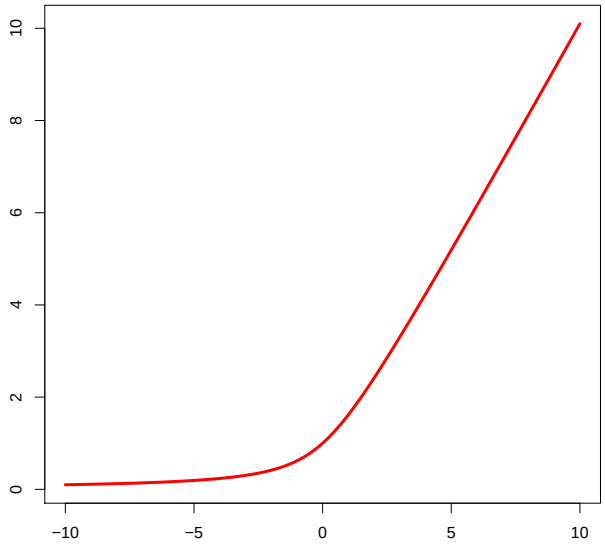
Part III Deformed exponentials

Kaniadakis' exponential

Definition ($\exp_\kappa$ with $\kappa = \frac{1}{2}$)

$$\exp_\kappa(y) = \left(\frac{1}{2}y + \sqrt{1 + \frac{1}{4}y^2}\right)^2$$

$$\ln_\kappa(x) = x^{1/2} - x^{-1/2}$$

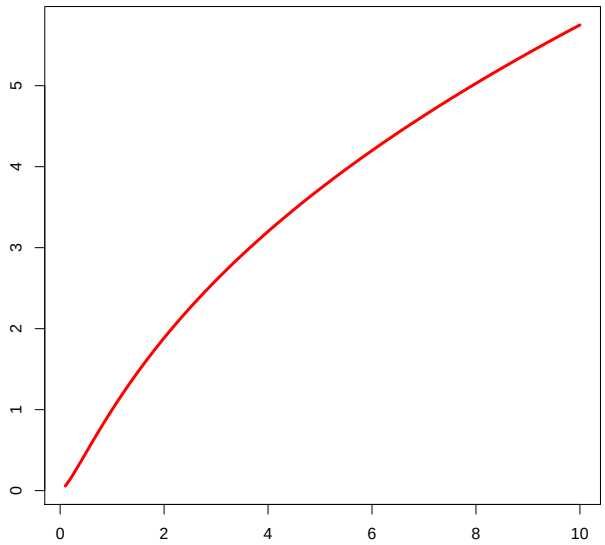


Deformed logarithm formalism

$$\ln_\kappa(x) = \int_1^x \frac{du}{\chi(u)}$$

$$\chi(u) = 2\frac{u^{3/2}}{u+1}$$

$$\frac{d}{dy}\exp_\kappa(y) = \chi(\exp_\kappa(y))$$



Musielak-Orlicz space

Definition ($L^{\kappa}(p)$)

For each $p \in \mathcal{P}_{>}$, the set

$$L^{\kappa}(p) = \left\{ v \middle| \int \exp_{\kappa}\left(\alpha\,|v| + \ln_{\kappa}(p)\right)\ d\mu < +\infty, \text{some } \alpha > 0 \right\}$$

is a Banach space with closed unit ball

$$S_{\kappa} = \left\{ v \in L^{\kappa}(p) \middle| \int \exp_{\kappa}\left(|v| + \ln_{\kappa}(p)\right)\ d\mu - 1 \leq 1 \right\}$$

Integrability

From the convexity

$$\exp_{\kappa}\left(\alpha\,|v| + \ln_{\kappa}(p)\right) - \exp_{\kappa}\left(\ln_{\kappa}(p)\right) \geq \exp_{\kappa}{}'\left(\ln_{\kappa}(p)\right)|v|\,,$$

, i.e.

$$\exp_{\kappa}\left(\alpha\,|v| + \ln_{\kappa}(p)\right) - p \geq \chi(p)\,|v|\,,$$

then $L^{\kappa}(p) \subset L^{1}(\chi(p) \cdot d\mu)$.

Deformed exponential arc

Example

Assume $p = 1$, so that $\ln_\kappa(p) = 0$. As for $y \geq 0$

$$y^2 \leq \left(\frac{1}{2}y + \sqrt{1 + \frac{1}{4}y^2}\right)^2 \leq (1+y)^2$$

we have

$$L^\kappa(1) = \left\{ v \middle| \int \exp_\kappa(\alpha |v|) \ d\mu \leq +\infty, \text{for some } \alpha > 0 \right\} = L^2(\mu)$$

Definition

As $\exp_\kappa$ is convex, for $p, q \in \mathcal{P}_>$

$$\int \exp_\kappa((1-\theta)\ln_\kappa(p) + \theta \ln_\kappa(q)) \ d\mu < +\infty, \quad \theta \in [0,1].$$

If it holds for $\theta \in I$, open $I \supset [0,1]$, then $p \smile q$.

Deformed model

Example

If $p = 1$, we want to check

$$\int \exp_\kappa (\theta \ln_\kappa (q)) \ d\mu < +\infty, \quad \theta > 1,$$

which is equivalent to

$$p^{1/2} + p^{-1/2} \in L^2(\mu)$$

If

$$q = \exp_\kappa (u - K_p(u) + \ln_\kappa (p)), \quad u \in L^\kappa(\mu) \cap L_0^1(\chi(p))$$

then $\ln_\kappa (p) - \ln_\kappa (q) \in L^\kappa(\mu)$ and

$$K_p(u) = \mathrm{E}_{\chi(p)} [\ln_\kappa (p) - \ln_\kappa (q)] = D_\kappa (p \| q)$$
$$u = \ln_\kappa (q) - \ln_\kappa (p) + D_\kappa (p \| q)$$