# Taxonomy of principal distances and divergences

Sony CSL

## Euclidean geometry

Euclidean distance
$d_2(\mathbf{p},\mathbf{q}) = \sqrt{\sum_i (p_i - q_i)^2}$ (Pythagoras' theorem circa 500 BC)

Hamming distance
$(|\{i : p_i \neq q_i\}|)$

Manhattan distance
$d_1(\mathbf{p},\mathbf{q}) = \sum_i |p_i - q_i|$
(city block-taxi cab)

Minkowski distance ($L_k$-norm)
$d_k(\mathbf{p},\mathbf{q}) = \sqrt[k]{\sum_i |p_i - q_i|^k}$
(H. Minkowski 1864-1909)

Lévy-Prokhorov distance
$\mathrm{LP}_\rho(p,q) = \inf_{\epsilon>0}\{p(A) \leq q(A^\epsilon) + \epsilon \forall A \in \mathcal{B}(\mathcal{X})\}$
$A^\epsilon = \{y \in \mathcal{X}, \exists x \in A : \rho(x,y) < \epsilon\}$

Quadratic distance
$d_{\mathbf{Q}} = \sqrt{(\mathbf{p}-\mathbf{q})^T \mathbf{Q} (\mathbf{p}-\mathbf{q})}$

## Hyperbolic/spherical geometry

Bolyai (1802-1860)

Lobachevsky (1792-1856)

## Statistical geometry

Physics entropy $JK^{-1}$
$-k \int p \log p \mathrm{d}\mu$
(Boltzmann-Gibbs 1878)

**Additive entropy**
cross-entropy
conditional entropy
mutual information
(chain rules)

Mahalanobis metric (1936)
$d_{\mathbf{\Sigma}} = \sqrt{(\mathbf{p}-\mathbf{q})^T \mathbf{\Sigma}^{-1} (\mathbf{p}-\mathbf{q})}$

Information entropy
$H(p) = -\int p \log p \mathrm{d}\mu$
(C. Shannon 1948)

Haussdorf set distance
$d_H(X,Y) = \max\{\sup_x \rho(x,Y), \sup_y \rho(X,y)\}$

$H(p) = \mathrm{KL}(p||u)$

Kullback-Leibler divergence
$\mathrm{KL}(\mathbf{p}||\mathbf{q}) = \int p \log \frac{p}{q} \mathrm{d}\mu = \mathrm{E}_p[\log \frac{P}{Q}]$
(relative entropy, 1951)

$I$-projection

Jeffrey divergence
(Jensen-Shannon)

## Non-Euclidean geometries

## Riemannian geometry

Fisher information (local entropy)
$\mathbf{I}(\theta) = \mathrm{E}[(\frac{\partial}{\partial\theta} \ln p(X|\theta))^2]$
(R. A. Fisher 1890-1962)

Riemannian metric tensor
$\int \sqrt{g_{ij} \frac{\mathrm{d}x_i}{\mathrm{d}s} \frac{\mathrm{d}x_j}{\mathrm{d}s}} \mathrm{d}s$
(B. Riemann 1826-1866,)

Finsler metric tensor
$g_{ij} = \frac{1}{2} \partial^2 \frac{F^2(x,y)}{\partial y^i \partial y^j}$

Aitchison distance
Probability simplex

Hilbert
log-ratio metric

Fisher-Rao distance:
$\mathrm{d}s^2 = g_{ij} \mathrm{d}\theta^i \mathrm{d}\theta^j = \mathrm{d}\theta^\top I(\theta) \mathrm{d}\theta$
$\rho_{FR}(p,q) = \min_\gamma \int_0^1 \sqrt{\dot\gamma(t) I(\theta) \dot\gamma(t)} \mathrm{d}t$

## Cone geometry

J.M. Souriau

J.L. Koszul

E. Vinberg

Bhattacharya distance (1967)
$d(p,q) = -\log \sqrt{\int \sqrt{p}\sqrt{q} \mathrm{d}\mu}$

Bhat.

Kolmogorov
$K(p||q) = \int |q-p| \mathrm{d}\mu$
(Kolmogorov-Smirnoff $\max |p-q|$)

## Symplectic geometry

$\alpha = -1$

Matsushita distance (1956)
$M_\alpha(p,q) = \sqrt[\alpha]{\int |q^{\frac{1}{\alpha}} - p^{\frac{1}{\alpha}}| \mathrm{d}\mu}$

Chernoff divergence (1952)
$C_\alpha(p||q) = -\ln \int p^\alpha q^{1-\alpha} \mathrm{d}\mu$
$C(p,q) = \max_{\alpha\in(0,1)} C_\alpha(p||q)$

Hellinger
$H(p||q) = \sqrt{\int (\sqrt{p} - \sqrt{q})^2}$
$= \sqrt{2(1-\sqrt{fg}}$

$\times \alpha(1-\alpha)$

Rényi divergence (1961)
$H_\alpha = \frac{1}{\alpha(1-\alpha)} \log \int f^\alpha \mathrm{d}\mu$
$\mathrm{R}_\alpha(\mathbf{p}|\mathbf{q}) = \frac{1}{\alpha(\alpha-1)} \ln \int p^\alpha q^{1-\alpha} \mathrm{d}\mu$
(additive entropy)

$\alpha = 0$

$\chi^2$ test
$\chi^2(p||q) = \int \frac{(q-p)^2}{p} \mathrm{d}\mu$
(K. Pearson, 1857-1936 )

Pearson

Vajda

Neyman

L. LeCam

## Affine differential geometry

K. Nomizu

Logarithmic divergence
$L_{G,\alpha}(\theta_1 : \theta_2) =$
$\frac{1}{\alpha} \log\left(1 + \alpha \nabla G(\theta_2)^\top (\theta_1 - \theta_2)\right) + G(\theta_2) - G(\theta_1)$

$\downarrow \alpha \to 0,\ F = -G$

Bregman divergences (1967):
$B_F(\theta_1||\theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^\top \nabla F(\theta_2)$

$\cap$ Kullback-Leibler

**Csiszár**' $f$-divergence
$D_f(p||q) = \int p f(\frac{q}{p}) \mathrm{d}\mu$
(Ali& Silvey 1966, Csiszár 1967)

Dual div. (Legendre) $D_{F*}(\nabla F(\theta_1)||\nabla F(\theta_2)) = D_F(\theta_2||\theta_1)$

Dual div.$_{*\text{-conjugate}}$ $(f^*(y) = y f(1/y))$
$D_{f^*}(p||q) = D_f(q||p)$

## Information geometries

F. Itakura

Itakura-Saito divergence
$\mathrm{IS}(\mathbf{p}|\mathbf{q}) = \sum_i (\frac{p_i}{q_i} - \log \frac{p_i}{q_i} - 1)$
(Burg entropy)

**Bregman-Csiszár** divergence (1991)
$F_\alpha(x) = \begin{cases} x - \log x - 1 & \alpha = 0 \\ x \log x - x + 1 & \alpha = 1 \\ \frac{1}{\alpha(1-\alpha)}(-x^\alpha + \alpha x - \alpha + 1) & 0 < \alpha < 1 \end{cases}$

Amari $\alpha$-divergence (1985)
$f_\alpha(x) = \begin{cases} x \log x & \alpha = 1 \\ -\log x & \alpha = -1 \\ \frac{4}{1-\alpha^2}(1 - x^{\frac{1+\alpha}{2}}) & -1 < \alpha < 1 \end{cases}$

$\nabla$

$\nabla^*$

Generalized
$f$-means
duality...

M. Nagumo

B. De Finetti

$\beta = 1$

**Generalized Pythagoras' theorem**
(Generalized projection)

## Quantum & matrix geometry

Fröbenius & Hilbert-Schmidt norm

Quantum entropy
$S(\rho) = -k \mathrm{Tr}(\rho \log \rho)$
(Von Neumann 1927)

**Burbea-Rao or Jensen**
(incl. Jensen-Shannon)
$J_F(p;q) = \frac{f(p)+f(q)}{2} - f\left(\frac{p+q}{2}\right)$

J. Jensen

Quantum $f$-divergences
(Dénes Petz)

$\beta \to \alpha$

**Sharma-Mittal entropies**
$h_{\alpha,\beta}(p) = \frac{1}{1-\beta}\left(\left(\int p^\alpha \mathrm{d}\mu\right)^{\frac{1-\beta}{1-\alpha}} - 1\right)$

**Non-additive entropy**

Tsallis entropy (1998)
(Non-additive entropy)
$T_\alpha(\mathbf{p}) = \frac{1}{1-\alpha}(\int p^\alpha \mathrm{d}\mu - 1)$
$T_\alpha(p||q) = \frac{1}{1-\alpha}(1 - \int \frac{p^\alpha}{q^{\alpha-1}} \mathrm{d}\mu)$

Log Det divergence
$D(\mathbf{P}||\mathbf{Q}) = <\mathbf{P}, \mathbf{Q}^{-1}> - \log\det \mathbf{P}\mathbf{Q}^{-1} - \dim \mathbf{P}$

Von Neumann divergence
$D(\mathbf{P}||\mathbf{Q}) = \mathrm{Tr}(\mathbf{P}(\log \mathbf{P} - \log \mathbf{Q}) - \mathbf{P} + \mathbf{Q})$

G. Monge

L. Kantorovich

Integral probability metrics
IPMs

Stein discrepancies

Earth mover distance
(EMD 1998)

$\rho = L_1$

MMD
Maximum Mean
Discrepancy

**Gromov-Haussdorf distance**
(between compact metric spaces)
$d_{GH}(X,Y) = \inf_{\phi_X : X \to Z, \phi_Y : Y \to Z}\{\rho_H^Z(\phi_X(X), \phi_Y(Y))\}$
$\phi_X, \phi_Y$: isometric embeddings

M. Fréchet

Wasserstein distances
$W_{\alpha,\rho}(p,q) = (\inf_{\gamma\in\Gamma(p,q)} \rho(p,q)^\alpha \mathrm{d}\gamma(x,y))^{\frac{1}{\alpha}}$

## Optimal transport geometry

Sinkhorn divergence ($h$-regularized OT)

©2023 Frank Nielsen