

# Taxonomy of principal distances and divergences



Sony CSL

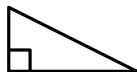


Bolyai  
(1802-1860)



Lobachevsky  
(1792-1856)

## Euclidean geometry



## Hyperbolic/spherical 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  
 $LP_\rho(p, q) = \inf_{\epsilon > 0} \{p(A) \leq q(A^\epsilon) + \epsilon \forall A \in \mathcal{B}(\mathcal{X})\}$   
 $A^\epsilon = \{y \in \mathcal{X}, \exists x \in A : \rho(x, y) < \epsilon\}$

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

## Riemannian geometry



Riemannian metric tensor  
 $\int \sqrt{g_{ij} \frac{dx_i}{ds} \frac{dx_j}{ds}} ds$   
(B. Riemann 1826-1866.)



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



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



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



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

## Affine differential geometry

Logarithmic divergence

$L_{G, \alpha}(\theta_1 : \theta_2) = \frac{1}{\alpha} \log(1 + \alpha \nabla G(\theta_2)^T (\theta_1 - \theta_2)) + G(\theta_2) - G(\theta_1)$

$\alpha \rightarrow 0, F = -G$

Bregman divergences (1967):

$B_F(\theta_1||\theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2)$

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

Itakura-Saito divergence

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

Bregman-Csiszár divergence (1991)

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

Generalized Pythagoras' theorem  
(Generalized projection)



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

Non-additive entropy

Tsallis entropy (1998)

(Non-additive entropy)

$T_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \left( \int p^\alpha d\mu - 1 \right)$

$T_\alpha(p||q) = \frac{1}{1-\alpha} \left( 1 - \int \frac{p^\alpha}{q^{1-\alpha}} d\mu \right)$



Earth mover distance  
(EMD 1998)

$\rho = L_1$

Wasserstein distances

$W_{\alpha, \rho}(p, q) = \left( \inf_{\gamma \in \Gamma(p, q)} \int \rho(x, y)^\alpha d\gamma(x, y) \right)^{\frac{1}{\alpha}}$

## Optimal transport geometry

Log Det divergence

$D(\mathbf{P}||\mathbf{Q}) = \langle \mathbf{P}, \mathbf{Q}^{-1} \rangle - \log \det \mathbf{P} \mathbf{Q}^{-1} - \dim \mathbf{P}$

Integral probability metrics

IPMs

MMD

Maximum Mean

Discrepancy

(between compact metric spaces)

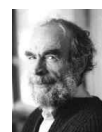
$d_{GH}(X, Y) = \inf_{\phi_X: X \rightarrow Z, \phi_Y: Y \rightarrow Z} \{ \rho_H^Z(\phi_X(X), \phi_Y(Y)) \}$   
 $\phi_X, \phi_Y$ : isometric embeddings

Sinkhorn divergence ( $h$ -regularized OT)

Stein discrepancies



Gromov-Hausdorff distance



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



Von Neumann divergence

$D(\mathbf{P}||\mathbf{Q}) = \text{Tr}(\mathbf{P}(\log \mathbf{P} - \log \mathbf{Q}) - \mathbf{P} + \mathbf{Q})$

## Quantum geometry

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



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

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

Generalized  
 $f$ -means  
duality...

$\beta = 1$

## Information geometries



Hellinger

$H(p||q) = \sqrt{\int (\sqrt{p} - \sqrt{q})^2 d\mu}$   
 $= \sqrt{2(1 - \int \sqrt{pq} d\mu)}$



Neyman

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



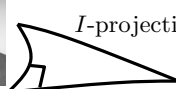
Kolmogorov  
 $K(p||q) = \int |q - p| d\mu$   
(Kolmogorov-Smirnov max  $|p - q|$ )



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

Bhattacharya distance (1967)

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



$I$ -projection  
Jeffrey divergence  
(Jensen-Shannon)

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



$H(p) = KL(p||u)$



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



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



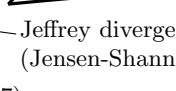
Information entropy  
 $H(p) = -\int p \log p d\mu$   
(C. Shannon 1948)

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

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



$I$ -projection



Jeffrey divergence  
(Jensen-Shannon)

Bhattacharya distance (1967)

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

Kolmogorov  
 $K(p||q) = \int |q - p| d\mu$   
(Kolmogorov-Smirnov max  $|p - q|$ )

Matsushita distance (1956)

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

Hellinger

$H(p||q) = \sqrt{\int (\sqrt{p} - \sqrt{q})^2 d\mu}$   
 $= \sqrt{2(1 - \int \sqrt{pq} d\mu)}$

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

Neyman

Csiszár'  $f$ -divergence

$D_f(p||q) = \int p f\left(\frac{p}{q}\right) d\mu$   
(Ali& Silvey 1966, Csizsár 1967)

Dual div. \*-conjugate ( $f^*(y) = y f(1/y)$ )

$D_{f^*}(p||q) = D_f(q||p)$

Generalized  $f$ -means duality...

$\beta = 1$

Amari  $\alpha$ -divergence (1985)

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

Quantum entropy

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

Von Neumann divergence

$D(\mathbf{P}||\mathbf{Q}) = \text{Tr}(\mathbf{P}(\log \mathbf{P} - \log \mathbf{Q}) - \mathbf{P} + \mathbf{Q})$

Gromov-Hausdorff distance

$d_{GH}(X, Y) = \inf_{\phi_X: X \rightarrow Z, \phi_Y: Y \rightarrow Z} \{ \rho_H^Z(\phi_X(X), \phi_Y(Y)) \}$   
 $\phi_X, \phi_Y$ : isometric embeddings

Sinkhorn divergence ( $h$ -regularized OT)