Annotated selected works

Frank Nielsen

We highlight the main result of each selected work as follows:

- Nielsen, F. and Okamura, K. (2023). On f-Divergences between Cauchy distributions. *IEEE Transactions on Information Theory*, 69(5):3150–3171 : The main result is that all f-divergences (3) $I_f(p : q) = \int p(x) f\left(\frac{q(x)}{p(x)}\right) \mathrm{d}x$ between univariate Cauchy distributions $p_{l_1,s_1}(x)$ and $p_{l_2,s_2}(x)$ are symmetric by showing that the χ^2-divergence is a *maximal invariant* (4) for the linear fractional transform action of $\mathrm{SL}(2,\mathbb{R})$ (special linear/real fractional linear group) when Cauchy distributions $p_{l,s}$ are parametrized by complex numbers $\theta = l + is$. That is $a.x \mapsto \frac{ax+b}{cx+d}$ and $A.X \sim \mathrm{Cauchy}(A.\theta)$ when $X \sim \mathrm{Cauchy}(\theta)$ for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Since all f-divergences are invariant under this group action, they can be expressed as a scalar function $h_f(u)$ of the maximal invariant $\chi(l_1, s_1; l_2, s_2) = I_{\chi^2}(p_{\theta_1} : p_{\theta_2}) = \frac{(l_1-l_2)^2}{2s_1s_2}$ divergence:
$$I_f(p_{l_1,s_1} : p_{l_2,s_2}) = h_f(\chi(l_1, s_1; l_2, s_2)) = I_f(p_{l_2,s_2} : p_{l_1,s_1}).$$

f-divergence name	$f(u)$	$I_f(p : q)$	$h_f(u)$
Chi-squared divergence	$(u-1)^2$	$\int \frac{(p(x)-q(x))^2}{p(x)} \mathrm{d}x$	u
Total variation distance	$\frac{1}{2}\lvert u-1\rvert$	$\int \frac{1}{2}\lvert p(x)-q(x)\rvert \mathrm{d}x$	$\frac{2}{\pi}\arctan\left(\sqrt{\frac{u}{2}}\right)$
Kullback-Leibler divergence	$-\log u$	$\int p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x$	$\log(1+\frac{1}{2}u)$
Jensen-Shannon divergence	$\frac{u}{2}\log\frac{2u}{1+u} - \frac{1}{2}\log\frac{1+u}{2}$	$\int \left(p(x)\log\frac{2p(x)}{p(x)+q(x)} + q(x)\log\frac{2q(x)}{p(x)+q(x)}\right)\mathrm{d}x$	$\log\left(\frac{2\sqrt{2+u}}{\sqrt{2+u}+\sqrt{2}}\right)$
Taneja T-divergence	$\frac{u+1}{2}\log\frac{u+1}{2\sqrt{u}}$	$\int \frac{p(x)+q(x)}{2}\log\frac{p(x)+q(x)}{2\sqrt{p(x)q(x)}}\mathrm{d}x$	$\log\left(\frac{1+\sqrt{1+\frac{u}{2}}}{2}\right),$
LeCam-Vincze divergence	$\frac{(u-1)^2}{1+u}$	$\int \frac{(p(x)-q(x))^2}{p(x)+q(x)}\mathrm{d}x$	$2-4\sqrt{\frac{1}{2(u+2)}}$
squared Hellinger divergence	$\frac{1}{2}(\sqrt{u}-1)^2$	$\frac{1}{2}\int\left(\sqrt{p(x)}-\sqrt{q(x)}\right)^2\mathrm{d}x$	$1-\frac{2K\left(1-\left(1+u+\sqrt{u(2+u)}\right)^{-2}\right)}{\pi\sqrt{1+u+\sqrt{u(2+u)}}}$

- Nielsen, F. (2023). A Simple Approximation Method for the Fisher–Rao Distance between Multivariate Normal Distributions. *Entropy*, 25(4):654 : The Fisher-Rao distance (Rao; 9) $\rho(p_{\theta_1}, p_{\theta_2})$ between two distributions p_{θ_1} and p_{θ_2} belonging to a parametric family of distributions $\{p_\theta\}$ is the geodesic Riemannian distance with respect to the Fisher metric. To calculate the Fisher-Rao distance in closed-form one needs to get closed-form of the Fisher-Rao geodesics and perform integration along the geodesics with boundary conditions. We show that the Fisher-Rao distance is upper bounded by the square root of their Jeffreys divergence: $\rho(p_{\theta_1}, p_{\theta_2}) \leq D_J\rho(p_{\theta_1}, p_{\theta_2})$. We then consider proxy curves of the Fisher-Rao geodesics obtained by orthogonal projections by using an isometric embedding of the Fisher-Rao manifold into the high-dimensional cone of symmetric positive-definite matrix (2).

- Nielsen, F. (2022b). Statistical divergences between densities of truncated exponential families with nested supports: Duo Bregman and duo Jensen divergences. *Entropy*, 24(3):421 : Consider two truncated densities $p_{\theta_1}^{R_1}$ and $p_{\theta_2}^{R_2}$ of an exponential family $\{p_\theta(x) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\mu}(x) = 1_{\mathcal{X}}(x)\exp(\langle\theta, t(x)\rangle - F(\theta) + k(x))\}$ where R_1 and R_2 are the supports of $p_{\theta_1}^{R_1}$ and $p_{\theta_2}^{R_2}$, respectively. A density p_θ^R of a truncated exponential family belongs to another exponential family with log-normalizer $F_R(\theta) = F(\theta) + \log Z_R(\theta)$ where $Z_R(\theta) = \int_R p_\theta(x)\mathrm{d}\mu(x)$. When $R_1 \subset R_2$ (nested support), we show that
$$D_{\mathrm{KL}}[p_{\theta_1}^{R_1} : p_{\theta_2}^{R_2}] = \int_{R_1} p_{\theta_1}^{R_1}(x)\log\frac{p_{\theta_1}^{R_1}(x)}{p_{\theta_2}^{R_2}(x)}\mathrm{d}\mu(x) = B_{F_{R_2},F_{R_1}}(\theta_2 : \theta_1),$$

where B_{F_1,F_2} is a duo Bregman pseudo-divergence:

$$B_{F_1,F_2}(\theta : \theta') = F_1(\theta) - F_2(\theta') - \langle \theta - \theta', \nabla F_2(\theta') \rangle \geq 0.$$

This is a pseudo-divergence because when $R_1 \neq R_2$, $B_{F_{R_1},F_{R_2}} > 0$. As an example, we report the formula for the Kullback-Leibler divergence between truncated normal distributions.

- Nielsen, F. (2022a). Generalizing the alpha-divergences and the oriented kullback–leibler divergences with quasi-arithmetic means. *Algorithms*, 15(11):435 : By observing that the α-divergences are scaled differences of the arithmetic (A) minus geometric (G) means (with $A \geq G$), we define (M, N) α-divergences for pairs of weighted means such that $M \geq N$. In the limit case of $\alpha \pm 1$, we get generalizations of the forward and reverse Kullback-Leibler divergences. We report these novel divergences for quasi-arithmetic means $M = A_f$ and $N = A_g$ where $A_h(a, b; \alpha) = h^{-1}(\alpha h(a) + (1 - \alpha)h(b))$.

References

[Rao]

[2] Calvo, M. and Oller, J. M. (1990). A distance between multivariate normal distributions based in an embedding into the Siegel group. *Journal of multivariate analysis*, 35(2):223–242.

[3] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observation. *studia scientiarum Mathematicarum Hungarica*, 2:229–318.

[4] Eaton, M. L. (1989). *Group invariance applications in statistics.*

[5] Nielsen, F. (2022a). Generalizing the alpha-divergences and the oriented kullback–leibler divergences with quasi-arithmetic means. *Algorithms*, 15(11):435.

[6] Nielsen, F. (2022b). Statistical divergences between densities of truncated exponential families with nested supports: Duo Bregman and duo Jensen divergences. *Entropy*, 24(3):421.

[7] Nielsen, F. (2023). A Simple Approximation Method for the Fisher–Rao Distance between Multivariate Normal Distributions. *Entropy*, 25(4):654.

[8] Nielsen, F. and Okamura, K. (2023). On f-Divergences between Cauchy distributions. *IEEE Transactions on Information Theory*, 69(5):3150–3171.

[9] Rao, C. R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in statistics*, pages 235–247. Springer.