On the Chi Square and Higher-Order Chi Distances for Approximating f -Divergences
IEEE Signal Process. Lett. 21(1): 10-13 (2014). BIBTEX:`J2014-FDivergencesChiTaylor` [1]

The Pearson $\chi^2_P(X_1 : X_2) = \int \frac{(x_2(x)-x_1(x))^2}{x_1(x)} \mathrm{d}\nu(x)$ and Neyman chi square distances $\chi^2_N(X_1 : X_2) = \chi^2_P(X_2 : X_1)$ are generalized to higher-order chi (signed) divergences $\chi^k_P(X_1 : X_2) = \int \frac{(x_2(x)-x_1(x))^k}{x_1(x)^{k-1}} \mathrm{d}\nu(x)$, and we show how to compute f-divergences from those chi-type divergences. For random variables of the same exponential family with affine natural space (like isotropic Gaussians, Poissons or multinomials), the Pearson/Neyman Chi square distance between $X_1 \sim \mathcal{E}_F(\theta_1)$ and $X_2 \sim \mathcal{E}_F(\theta_2)$ is given by $\chi^2_P(X_1 : X_2) = e^{F(2\theta_2-\theta_1)-(2F(\theta_2)-F(\theta_1))} - 1 = \chi^2_N(X_2 : X_1)$. For Poisson distributions, we get $\chi^2_P(\lambda_1 : \lambda_2) = \exp\left(\frac{\lambda_2^2}{\lambda_1} - 2\lambda_2 + \lambda_1\right) - 1$ and for isotropic Gaussians $\chi^2_P(\mu_1 : \mu_2) = e^{(\mu_2-\mu_1)^\top(\mu_2-\mu_1)} - 1$. For members of the same exponential family with affine natural parameter space, $\chi^k_P(X_1 : X_2) = \sum_{j=0}^k (-1)^{k-j}\binom{k}{j} \frac{e^{F((1-j)\theta_1+j\theta_2)}}{e^{(1-j)F(\theta_1)+jF(\theta_2)}}$, so that $\chi^k_P(\lambda_1 : \lambda_2) = \sum_{j=0}^k (-1)^{k-j}\binom{k}{j} e^{\lambda_1^{1-j}\lambda_2^j - ((1-j)\lambda_1+j\lambda_2)}$, and $\chi^k_P(\mu_1 : \mu_2) = \sum_{j=0}^k (-1)^{k-j}\binom{k}{j} e^{\frac{1}{2}j(j-1)(\mu_1-\mu_2)^\top(\mu_1-\mu_2)}$. f-Divergences can be expressed as $I_f(X_1 : X_2) = \int x_1(x)\sum_{i=0}^\infty \frac{1}{i!} f^{(i)}(\lambda)\left(\frac{x_2(x)}{x_1(x)} - \lambda\right)^i \mathrm{d}\nu(x) = \sum_{i=0}^\infty \frac{1}{i!} f^{(i)}(\lambda)\ \chi^i_{\lambda,P}(X_1 : X_2)$. In particular, when X_1 is close to X_2, we have $I_f(X_1 : X_2) \sim \frac{f''(1)}{2}\chi^2_P(X_1 : X_2)$. The Kullback-Leibler divergence is expressed as $\mathrm{KL}(X_1 : X_2) = \sum_{j=2}^\infty \frac{(-1)^i}{i}\ \chi^j_P(X_1 : X_2)$. Series truncation for approximation and upper bounds of the Taylor remainder are studied with experiments.

References

[1] Frank Nielsen and Richard Nock. On the Chi square and higher-order Chi distances for approximating f-divergences. *Signal Processing Letters, IEEE*, 21(1):10–13, Jan 2014.