Easily calculating and programming the skew Bhattacharyya coefficients and related divergences

Frank Nielsen
Frank.Nielsen@acm.org

May 19, 2022

The α-skew Bhattacharyya coefficient [?] (for $\alpha \in (0,1)$ is a similarity measure defined by

$$\rho_\alpha[p:q] = \int p^\alpha(x) q^{1-\alpha}(x) \mathrm{d}\mu(x) = \rho_{-\alpha}[q:p].$$

We have $0 < \rho_\alpha[p:q] \leq 1$.

The Bhattacharyya coefficient can be used to define dissimilarities like the α-skew Bhattacharyya distances:

$$D_{\mathrm{Bhat}}[p:q] = -\log \rho_\alpha[p:q],$$

or the α-divergences for $\alpha \in \mathbb{R}$:

$$D_\alpha[p:q] = \left\{ \begin{array}{ll} \frac{4}{1-\alpha^2}\left(1-\rho_{\frac{1-\alpha}{2}}(p:q)\right), & \alpha \notin \{-1,1\} \\ D_{\mathrm{KL}}[p:q], & \alpha = -1 \\ D_{\mathrm{KL}}[q:p], & \alpha = 1. \end{array} \right. ,$$

where D_{KL} denotes the Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(p:q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}\mu(x).$$

We shall consider the α-Bhattacharyya coefficient between multivariate Gaussian distributions with the same covariance matrix, where the probability density of a multivariate Gaussian distribution $p_{\mu,\Sigma}(x)$ with mean $\mu \in \mathbb{R}^d$ and covariance matrix Σ is:

$$p_{\mu,\Sigma}(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

It is reported in [?] (page 46):

$$\rho_\alpha[p_{\mu,\Sigma_1} : p_{\mu,\Sigma_2}] = \frac{|\Sigma_1|^{\frac{1-\alpha}{2}} |\Sigma_2|^{\frac{\alpha}{2}}}{|(1-\alpha)\Sigma_1 + \alpha\Sigma_2|^{\frac{1}{2}}}.$$

We give the following calculation recipe in [?]:

- Since $\rho_\alpha[p_{\theta_1} : p_{\theta_2}] = \exp(-J_{F,\alpha}(\theta_1 : \theta_2))$ for densities $p_\theta(x) = \exp(\theta^\top t(x) - F(\theta) + k(x))$ belonging to an exponential family (where $J_{F,\alpha}$ is a skew Jensen divergence) and since $J_{F,\alpha} = J_{G,\alpha}$ for $G(\theta) = F(\theta) + a^\top\theta + b$, let us choose $G(\theta) = -\log p_\theta(x)$. Furthermore, the densities are parameterized by their usual parameters λ which may differ from their natural parameters θ. Thus we have
$$\rho_\alpha[p_{\theta_1} : p_{\theta_2}] = \frac{p_{\lambda_\alpha}(\omega)}{p_{\lambda_1}(\omega)^\alpha \, p_{\lambda_2}(\omega)^{1-\alpha}}, \quad \forall\omega \in \mathbb{R}^d,$$
where $\lambda_\alpha = \lambda((1-\alpha)\theta_1 + \alpha\theta_2)$.

- To get λ_α, let us write $p_{\lambda_1}^\alpha(\omega) p_{\lambda_2}^{1-\alpha}(\omega) \propto \exp(\lambda_\alpha^\top t(\omega))$. Thus we do not need to explicitly calculate the log-normalizer $F(\theta)$ to get λ_α.

For the α-Bhattacharyya coefficient between same-mean Gaussian distributions, let us choose $\omega = \mu$ so that
$$p_{\mu,\Sigma}(\mu) = \frac{1}{\sqrt{\det(2\pi\Sigma)}}.$$

Therefore
$$p_{\mu,\Sigma_1}(\mu)^\alpha p_{\mu,\Sigma_2}(\mu)^{1-\alpha} = \frac{1}{\sqrt{(2\pi I)}|\Sigma_1|^\alpha|\Sigma_2|^{1-\alpha}}.$$

Hence
$$\Sigma_\alpha =$$

Overall, we have
$$...$$

References