Bhattacharyya distance for hurdle gamma distributions

Stefano Mezzini

2023-02-02

For discrete variables with PMFs $P(\mathcal{X})$ and $Q(\mathcal{X})$ over the common domain $\mathcal{X}$, the Bhattacharyya distance is

$$D_B = -\log\left(\sum_{x\in\mathcal{X}} \sqrt{P(x)\ Q(x)}\right).$$

For continuous variables with PDFs $P(\mathcal{X})$ and $Q(\mathcal{X})$ over the common domain $\mathcal{X}$, the Bhattacharyya distance is

$$D_B = -\log\left(\int_{\mathcal{X}} \sqrt{P(x)\ Q(x)}\ dx\right).$$

We can approximate it using a sum over a finite set $\mathcal{X}' \in \mathcal{X}$ with finite sampling intervals dx:

$$D_B = -\log\left(\sum_{x\in\mathcal{X}'} \sqrt{P(x)\ Q(x)}\ dx\right),$$

provided the size of $\mathcal{X}'$ is sufficiently large to ensure a reasonably stable and accurate estimate.

Given two random variables following **Bernoulli distributions** with success probabilities p_1 and p_2, the Bhattacharyya distance is

$$D_B = -\log\left(\sum_{x\in\{0,1\}} \sqrt{P(x)\ Q(x)}\right) = -\log\left(\sqrt{(1-p_1)(1-p_2)} + \sqrt{p_1\ p_2}\right).$$

Given two random variables following **hurdle gamma distributions** with probabilities of success p_1 and p_2, shapes k_1 and k_2, and scales θ_1 and θ_2, the approximate Bhattacharyya distance is

$$D_B \approx -\log\left(\sum_{x\in\mathcal{X}'} \sqrt{P(x)\ Q(x)}\right) = -\log\left(\sqrt{(1-p_1)(1-p_2)} + \sum_{x\in\mathcal{X}'-\{0\}} \sqrt{(p_1\ P_\Gamma(x))\ (p_2\ Q_\Gamma(x))}\right)$$