Code name: **JB**
Argument type: Parameter

1 Jensen-Bregman divergence

1.1 Definitions

The *Jensen-Bregman divergence*[1] (first introduced in April 2010[2]) between parameters p and q has been historically designed as a symmetrization of Bregman divergences generalizing the renown Jensen-Shannon divergence[3]. The Jensen-Bregman divergence defined by:

$$\mathrm{JB}_F(p;q) = \frac{1}{2}\left(\mathrm{B}_F\left(p : \frac{p+q}{2}\right) + \mathrm{B}_F\left(q : \frac{p+q}{2}\right)\right) \quad (1)$$

where $F \in \mathcal{F}_B$ is a Bregman function generator.

The Jensen-Bregman divergence amounts to a Jensen divergence (also called a Burbea-Rao divergence):

$$\mathrm{JB}_F(p;q) = \frac{F(p)+F(q)}{2} - F\left(\frac{p+q}{2}\right) = \mathrm{J}_F(p:q) = \mathrm{BR}_F(p;q) \quad (2)$$

Notice that the terms with the gradients $\nabla F\left(\frac{p+q}{2}\right)$ in the BD expressions cancel out (proof given later on for skew versions).

Choosing $F(x) = -H(x)$, the Shannon information, we recover the Jensen-Shannon divergence:

$$\mathrm{JB}_{-H}(p;q) = H\left(\frac{p+q}{2}\right) - \frac{H(p)+H(q)}{2} = \mathrm{JS}(p;q) \quad (3)$$

Let $\mathrm{K}_F(p:q) = \mathrm{B}_F\left(p : \frac{p+q}{2}\right)$ (generalizing Lin's K divergence for the Kullback-Leibler divergence). Then we have:

$$\mathrm{JB}_F(p;q) = \frac{1}{2}(\mathrm{K}_F(p:q) + \mathrm{K}_F(q:p)).$$

For $\alpha \in \mathbb{R}$, we define the *skew Jensen-Bregman divergence*[4]:

$$\mathrm{JB}_F^{(\alpha)}(p:q) = \alpha \mathrm{B}_F(p : \alpha p + (1-\alpha)q) + (1-\alpha)\mathrm{B}_F(p : \alpha p + (1-\alpha)q). \quad (4)$$

[1] Frank Nielsen and Sylvain Boltz. "The Burbea-Rao and Bhattacharyya centroids". In: *IEEE Transactions on Information Theory* 57.8 (2011), pp. 5455–5466.
[2] Frank Nielsen and Sylvain Boltz. "The Burbea-Rao and Bhattacharyya centroids". In: *CoRR* abs/1004.5049 (Apr. 2010). URL: http://arxiv.org/abs/1004.5049.
[3] Jianhua Lin. "Divergence measures based on the Shannon entropy". In: *IEEE Transactions on Information theory* 37.1 (1991), pp. 145–151.
[4] Frank Nielsen and Richard Nock. "Jensen-Bregman Voronoi diagrams and centroidal tessellations". In: *International Symposium on Voronoi Diagrams in Science and Engineering (ISVD)*. IEEE. 2010, pp. 56–65.

The skew Jensen-Bregman divergence amounts to a skew Jensen divergence (and vice-versa):

$$\mathrm{JB}_F^{(\alpha)}(p:q) = \alpha F(p) + (1-\alpha)F(q) - F(\alpha p + (1-\alpha)q), \quad (5)$$
$$= \mathrm{J}_F^{(\alpha)}(p:q). \quad (6)$$

Proof. Let $m_\alpha = \alpha p + (1-\alpha)q$ denote the weighted mixture of parameters p and q. A direct expansion of $\mathrm{JB}_F^{(\alpha)}(p:q)$ yields

$$\begin{aligned}\mathrm{JB}_F^{(\alpha)}(p:q) = \\ \alpha F(p) - \alpha F(m_\alpha) - \alpha\langle p - m_\alpha, \nabla F(m_\alpha)\rangle \\ +(1-\alpha)F(q) - (1-\alpha)F(m_\alpha) - (1-\alpha)\langle q - m_\alpha, \nabla F(m_\alpha)\rangle\end{aligned}$$

Reordering the terms, we get:

$$\mathrm{JB}_F^{(\alpha)}(p:q) = \alpha F(p) + (1-\alpha)F(q) - F(m_\alpha) + \langle \nabla F(m_\alpha), -\alpha(p-m_\alpha) - (1-\alpha)(q-m_\alpha)\rangle$$

Since $p - m_\alpha = (1-\alpha)(p-q)$ and $q - m_\alpha = -\alpha(p-q)$, we have $-\alpha(p - m_\alpha) - (1-\alpha)(q-m_\alpha) = -\alpha(1-\alpha)(p-q) + (1-\alpha)\alpha(p-q) = 0$ in the second argument of the inner product, and the inner product is 0. Thus $\mathrm{JB}_F^{(\alpha)}(p:q) = \alpha F(p) + (1-\alpha)F(q) - F(\alpha p + (1-\alpha)q) = \mathrm{J}_F^{(\alpha)}(p:q)$. □

Define the *scaled skew Bregman-Jensen divergences* and *scaled skew Jensen divergences* for $\alpha \in \mathbb{R}\backslash\{0,1\}$:

$$\mathrm{JB}'^{(\alpha)}_F(p:q) = \frac{1}{\alpha(1-\alpha)}\mathrm{JB}_F^{(\alpha)}(p:q) = \mathrm{JB}_F^{(\alpha)}(p:q) \quad (7)$$

In the limit cases $\alpha \to 0$ or $\alpha \to 1\}$, the scaled skew JBD tend to a sided Bregman divergence[5]:

$$\lim_{\alpha\to 0} \mathrm{JB}'^{(\alpha)}_F(p:q) = \mathrm{J}'^{(\alpha)}_F(p:q) = \mathrm{B}_F(p:q) \quad (8)$$
$$\lim_{\alpha\to 1} \mathrm{JB}'^{(\alpha)}_F(p:q) = \mathrm{J}'^{(\alpha)}_F(p:q) = \mathrm{B}_F(q:p) \quad (9)$$

(A straightforward proof for the scaled skew Jensen divergence follows from the equivalence with the scaled skew Jensen-Bregman divergence, and by using the definition of the scaled skew Jensen-Bregman divergence.)

Those definitions extend straightforwardly to matrix types[6]: For example, by considering the von Neumann information (negative of the von Neumann entropy) $F(X) = \mathrm{tr}(X\log X - X)$ or the log-det information $F(X) = -\log|X|$.

[5] Nielsen and Boltz, "The Burbea-Rao and Bhattacharyya centroids".
[6] Nielsen and Nock, "Jensen-Bregman Voronoi diagrams and centroidal tessellations".

Remark 1. *Although* $\mathrm{B}_F(p:q) = \mathrm{B}_{F^*}(q^*:p^*)$ *(from Young's equality* $F(x) + F^*(x^*) = \langle x, x^*\rangle$*), we cannot easily expressed a relationship between* $\mathrm{JB}_F^{(\alpha)}(p:q)$ *and* $\mathrm{JB}_{F^*}^{(\alpha)}(p^*:q^*)$ *because* $m_\alpha^* = \nabla F(m_\alpha)$ *does not express in terms of* p^* *and* q^* *(except for the case of the quadratic information* $F(x) = \frac{1}{2}\langle x, x\rangle$*).*

We extend the skew Jensen-Bregman divergence to weighted convex combination of points using the Jensen-Bregman diversity:

$$\mathrm{JB}_F(p_1, \dots p_n; w_1, \dots, w_n) := \sum_{i=1}^n w_i \mathrm{B}_F(p_i : m_w), \quad m_w = \sum_{i=1}^n w_i p_i \tag{10}$$

The Jensen-Bregman diversity amounts to a Jensen diversity:

$$\mathrm{JB}_F(p_1, \dots p_n; w_1, \dots, w_n) = \mathrm{J}_F(p_1, \dots p_n; w_1, \dots, w_n), \tag{11}$$

with the Jensen diversity defined as:

$$\mathrm{J}_F(p_1, \dots p_n; w_1, \dots, w_n) := \sum_{i=1}^n w_i F(p_i) - F\left(\sum_{i=1}^n w_i p_i\right) \tag{12}$$

Proof. We have $\sum_{i=1}^n w_i \mathrm{B}_F(p_i : m_w)$, $m_w = \sum_{i=1}^n w_i p_i = (\sum_i w_i F(p_i)) - F(\bar{p}) - \sum_{i=1}^n w_i \langle (\sum_i w_i(p_i - \bar{p}), \nabla F(\bar{p})\rangle$. Since $\sum_i w_i(p_i - \bar{p}) = (\sum_i w_i p_i) - (\sum_i w_i)\bar{p} = 0$, the inner product terms vanish, and we have the Jensen-Bregman diversity that equals the Jensen diversity. □

In Bregman k-means clustering[7], the Jensen(-Bregman) diversity is called the *Bregman information*: It generalizes the notion of variance of a cluster (that is obtained by considering the squared Euclidean distance, a Bregman divergence).

Sided information radius: For a given weighted set, let $R_F(x) = \sum_i w_i \mathrm{B}_F(p_i : x)$ and $L_F(l) = \sum_i w_i \mathrm{B}_F(x : p_i)$. Let $r = \arg\min R_F(x)$ and $l = \arg\min L_F(x)$. Then $R_F(r; p_i, w_i) = L_{F^*}(r^*; p_i^*, w_i)$ and $L_F(l; p_i, w_i) = R_{F^*}(l^*; p_i^*, w_i)$. In general, the sided information radii do not coincide (except when the generator is $F(x) = \frac{1}{2}\langle x, x\rangle$)

1.2 Properties

- $\sqrt{\mathrm{JB}_F(p;q)}$ is a metric iff $F(p+q)$ is a Conditionally Positive Definite (CPD) kernel[8].

[7] Frank Nielsen and Richard Nock. "Sided and symmetrized Bregman centroids". In: *IEEE transactions on Information Theory* 55.6 (2009), pp. 2882–2904.

[8] Sreangsu Acharyya, Arindam Banerjee, and Daniel Boley. "Bregman Divergences and Triangle Inequality". In: *Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA.*. 2013, pp. 476–484. DOI: 10.1137/1.9781611972832.53. URL: http://dx.doi.org/10.1137/1.9781611972832.53.

1.3 Further extensions

Jensen-Bregman log det divergence[9], total Jensen-Bregman divergences[10] that are equivalent to total Jensen divergences.

1.4 Historical notes

See:[11], matrix Jensen-Bregman log-det divergence[12]

See also:[13]

---

[9] Anoop Cherian et al. "Jensen-Bregman LogDet divergence with application to efficient similarity search for covariance matrices". In: *IEEE transactions on pattern analysis and machine intelligence* 35.9 (2013), pp. 2161–2174.

[10] Meizhu Liu et al. "Shape retrieval using hierarchical total Bregman soft clustering". In: *IEEE transactions on pattern analysis and machine intelligence* 34.12 (2012), pp. 2407–2419.

[11] Nielsen and Boltz, "The Burbea-Rao and Bhattacharyya centroids"; Nielsen and Nock, "Jensen-Bregman Voronoi diagrams and centroidal tessellations"; Frank Nielsen and Richard Nock. "Skew Jensen-Bregman Voronoi diagrams". In: *Transactions on Computational Science XIV*. Springer, 2011, pp. 102–128.

[12] Cherian et al., "Jensen-Bregman LogDet divergence with application to efficient similarity search for covariance matrices".

[13] Frank Nielsen. "A family of statistical symmetric divergences based on Jensen's inequality". In: *CoRR* abs/1009.4004 (2010). URL: `http://arxiv.org/abs/1009.4004`.