# Divergences de Bregmann, $\phi-$entropies, lois $\phi-$exponentielles...

JFB & VG's tribulations

May 28, 2014

Soit une fonction $f$ strictement convexe, de classe $C^1(\mathbb{R})$ (ou au moins continue et dérivable), et la fonction de deux variable

$$d_f(u_1,u_0) = f(u_1) - f(u_0) - (u_1 - u_0)f'(u_0)$$

(divergence de Bregman, définies dans le cadre plus général de vecteurs et de matrices [1]). On définie la famille des divergences de Bregmann entre deux densités de probabilité $p_0$ et $p_1$ sous la forme

$$D_f(p_1\|p_0) = \int_{\mathbb{R}^d} d_f(p_1(x),p_0(x))\,dx = \int_{\mathbb{R}^d} \left(f(p_1(x)) - f(p_0(x)) - (p_1(x)-p_0(x))f'(p_0(x))\right)\,dx \tag{1}$$

[2–4], l'entropie associée (cf. $h-\phi-$entropies de Salicrú [5, 6] avec $h=-id$, $f$ supposée $C^3(\mathbb{R}_+)$)

$$H_f(p) = -\int_{\mathbb{R}^d} f(p(x))\,dx \tag{2}$$

et donc

$$D_f(p_1\|p_0) = H_f(p_0) - H_f(p_1) - \int_{\mathbb{R}^d} (p_1(x)-p_0(x))f'(p_0(x))\,dx$$

A noter que la forme symétrique des divergences de Bregman s'écrit $D_f^s(p_1\|p_0) = D_f(p_1\|p_0) + D_f(p_0\|p_1)$, i.e.,

$$D_f^s(p_1\|p_0) = \int_{\mathbb{R}^d} (p_1(x)-p_0(x))\left(f'(p_1(x)) - f'(p_0(x))\right)\,dx \tag{3}$$

En raison de la convexité de $f$, $d_f(x_1,x_0) \ge 0$ avec égalité ssi $x_1 = x_0$ et par conséquent

$$D_f(p_1\|p_0) \ge 0 \qquad \text{avec égalité ssi} \qquad p_1 = p_0 \text{ (p. p.)} \tag{4}$$

- pour $f(u) = u\log(u)$ on retrouve comme cas particulier la divergence de Kullback-Leibler et l'entropie de Shannon
- pour $f(u) = \frac{u^\alpha - 1}{\alpha - 1}$ on retrouve l'entropie de Tsallis en la divergence de Bregamnn associée.

(voir [**?** ] pour un paronama complet).

A l'image des $q-$logarithmes à la Tsallis, on défini une extension du logarithme sous la forme **(cf. J. Naudts)**

$$\begin{array}{rccc} \log_\phi & : & \mathbb{R}_+^* & \to & \mathbb{R} \\ & & u & \mapsto & \displaystyle\int_1^u \frac{1}{\phi(v)}\,dv \end{array} \tag{5}$$

avec $\phi$ strictement positive (sauf éventuellent sur un ensemble de mesure nulle), assurant que $\log_\phi$ soit strictement croissante, et on notera $\exp_\phi$ sa fonction réciproque,

$$\exp_\phi \circ \log_\phi(u) = u \tag{6}$$

- Pour $\phi(u) = u$ on retrouve le logarithme usuel
- Pour $\phi(u) = u^q$ on retrouve le $q-$logarithme de Tsallis

On définit enfin une fonctionnelle de divergence sous la forme

$$f_\phi(u) = \int_1^u \log_\phi(v)\, dv \tag{7}$$

qui, par construction même, est strictement convexe. On s'intéresse alors à la divergence de Bregmann

$$D_{f_\phi}(p_1\|p_0) = H_{f_\phi}(p_0) - H_{f_\phi}(p_1) - \int_{\mathbb{R}^d} (p_1(x) - p_0(x)) \log_\phi(p_0(x))\, dx$$

Plaçons nous à présent dans le cadre scalaire, $d = 1$ et imaginons que l'on travaille avec une loi référence de type $\phi-$exponentielle

$$p_0(x) = \exp_\phi(T(x))$$

Alors pour $p_1$ et $p_0$ de même moment $E_{p_0}[T(X)] = E_{p_1}[T(X)]$ a

$$D_{f_\phi}(p_1\|p_0) = H_{f_\phi}(p_0) - H_{f_\phi}(p_1) \geq 0$$

avec égalité ssi $p_1 = p_0$ (p. p.).

Fixons nous $p(x)$ et cherchons $\phi$ et $T$ tels que $p(x) = \exp_\phi(T(x))$. Nécessairement sur les intervalles où $p$ est bijective $T$ est bijective et,

$$p \circ T^{-1} = \exp_\phi$$

c'est-à-dire

$$T \circ p^{-1} = \log_\phi$$

et donc, par dérivation,

$$\phi(y) = \frac{T'}{p'}(p^{-1}(y))$$

Il faut nécessairement que $T'/p'$ soit positive pour que $p$ puisse être sous forme $\phi-$exponentielle avec $f_\phi$ convexe. **voir, mais $\phi$ définie de manière unique sur l'image de $p$.**

# References


[1] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problem in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

[2] I. Csiszàr. Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2031–2066, December 1991.

[3] I. Csiszàr and F. Matúš. Generalized minimizers of convex integral functionals, Bregman distance, Pythgorean identities. *Kybernetica*, 48(4):637–689, 2012.

[4] M. Basseville. Divergence measures for statistical data processing – an annotated bibliography. *Signal Processing*, 93(4):621–633, April 2013.

[5] M. Salicrú. Funciones de entropía asociada a medidas de Csiszár. *Qüestiió*, 11(3):3–12, 1987.

[6] M. Salicrú. Measures of information associated with Csiszár's divergences. *Kybernetica*, 30(5):563–573, 1994.