

DISTANCE ENTRE LOIS (A4, A7, C, G8)

(25 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **distance entre lois (de probabilité)** n'est autre qu'une **distance** définie sur une **famille de lois** (cf aussi **distance entre probabilités**). Une notion voisine est celle de **distance entre variables aléatoires**.

(i) En **Statistique**, on utilise souvent :

(a) la distance de KULLBACK-LEIBLER (cf **statistique de KULLBACK-LEIBLER**) ;

(b) la distance de SHANNON (cf **information de SHANNON**) ;

(c) la **distance de HELLINGER**.

Ce type de distances intervient dans l'étude des **modèles** (cf **comparaison de modèles**), dans l'**analyse des proximités** (cf **proximité**), dans la théorie de la **robustesse**, etc.

(ii) Du point de vue géométrique (H. JEFFREYS, C.R. RAO), la notion de **matrice d'information** (cf **information de FISHER**) permet de définir une **métrique riemannienne** sur un ensemble de **lp** (cf **variété de RIEMANN**).

Cette métrique est définie par la **forme quadratique fondamentale** :

$$(1) \quad q(\theta) = (d\theta_q)' I_{qr}(\theta) (d\theta_r), \quad \forall \theta \in \Theta,$$

dans laquelle $\Theta \in \mathcal{O}(\mathbf{R}^Q)$ (**ouvert** de \mathbf{R}^Q) et $I(\theta) = (I_{qr}(\theta))_{(q,r)}$ est la (Q,Q)-matrice d'information de FISHER.

La distance précédente permet de calculer eg la **distance géodésique** entre deux lois $P_{\theta}^{X'}$ et $P_{\theta}^{X''}$.

(iii) La distance utilisée n'est pas toujours une « vraie » distance (ie ne vérifie pas les trois axiomes d'une distance) (cf **semi-distance**, **écart**, **fonction d'écart**, **entropie relative**, **estimateur à distance minimale**, **estimateur à distance minimum**).