

DISTANCE DE BHATTACHARYYA (A4, C10, G, H, I)

(14 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

La **distance de BHATTACHARYYA** est définie sur un ensemble de **lois de probabilité** à partir d'une distance entre leurs **densités**.

(i) Soit $(\Omega, \mathcal{F}, \mathcal{P})$ un **modèle statistique**, $\xi : \Omega \mapsto \mathbf{R}^K$ un **vecteur aléatoire** et $P_1^\xi = \xi(P_1)$ (resp $P_2^\xi = \xi(P_2)$) la **mesure image** de P_1 (resp de P_2) par ξ . On suppose que P_1^ξ (resp P_2^ξ) admet une **densité de probabilité** f_1 (resp f_2) par à une **mesure abstraite** positive σ -finie μ définie sur $\mathcal{B}(\mathbf{R}^K)$ (cf **mesure positive, mesure σ -finie**).

On note \mathcal{P}^ξ la **famille** des lois qui sont les images des **mesures de probabilité** $P \in \mathcal{P}$ par ξ .

On appelle alors **distance de G.K. BHATTACHARYYA** sur \mathcal{P}^ξ la **distance** d définie par :

$$(1) \quad d(P_1^\xi, P_2^\xi) = \text{Arc cos} \left\{ \int_{\mathbf{R}^K} (f_1(x) \cdot f_2(x))^{1/2} d\mu(x) \right\}.$$

Autrement dit :

$$(2) \quad d(P_1^\xi, P_2^\xi) = \text{Arc cos} \left\{ \int_{\mathbf{R}^K} (f_2 / f_1)^{1/2} dP_1^\xi \right\} = \text{Arc cos} \left\{ \int_{\mathbf{R}^K} (f_1 / f_2)^{1/2} dP_2^\xi \right\}.$$

On note \mathbf{R}^K pour désigner \mathbf{R}^K .

(ii) La distance d permet de définir la **statistique de G.K. BHATTACHARYYA**.

Ainsi, dans le cadre du problème à deux échantillons X^1 et X^2 (cf **problème à plusieurs échantillons**), de tailles resp N_1 et N_2 , on note resp f_1^\sim et f_2^\sim des estimateurs des densités f_1 et f_2 (dérivées des **fonctions de répartition** F_1 et F_2 dont sont issus les échantillons).

La **statistique** de BHATTACHARYYA s'écrit :

$$(3) \quad B_{N(1)N(2)} = \text{Arc cos} \left\{ \int_{\mathbf{R}^K} (f_1^\sim(x) f_2^\sim(x))^{1/2} d\mu(x) \right\},$$

où f_1^\sim dépend du N_1 -échantillon X^1 , f_2^\sim dépend du N_2 -échantillon X^2 , et où $N(i)$ désigne N_i ($i = 1, 2$).

Cette statistique intervient dans divers **tests d'hypothèses d'homogénéité** portant sur les deux **populations** considérées.