

FONCTION DISCRIMINANTE DE FISHER (I7, K7, K9)

(10 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **fonction discriminante de FISHER** est une **fonction discriminante** utilisée dans le cas où l'on veut imputer un **échantillon** à une **population** parmi deux.

(i) Soit $(\Omega, \mathcal{F})$ un **espace probabilisable** et $\mathcal{P} = \{P_{\theta(1)}, P_{\theta(2)}\}$ une **famille** constituée de deux **mesures de probabilité** définies sur $\mathcal{F}$, où $\theta(i)$ désigne, par commodité, θ_i ($i = 1, 2$). On note souvent $\Theta = \{1, 2\}$ au lieu de $\Theta = \{\theta_1, \theta_2\}$ et l'on munit $\mathcal{P}$ d'une **probabilité a priori** $\Pi = \pi_1 \delta_1 + \pi_2 \delta_2$ (avec $\pi_1 + \pi_2 = 1$, $\pi_1 > 0$ et $\pi_2 > 0$), où δ_i désigne la **masse de DIRAC** placée en i ($i = 1, 2$).

D'autre part, soit $\xi : \Omega \mapsto \mathbf{R}^K$ un **vecteur aléatoire** réel donné. Le **modèle image** par ξ du modèle de base $(\Omega, \mathcal{F}, \mathcal{P})$ précédent est donc $(\mathbf{R}^K, \mathcal{B}(\mathbf{R}^K), (P_1^\xi, P_2^\xi))$.

Π exprime alors que ξ est tiré selon la loi P_1^ξ avec la probabilité π_1 ou selon la loi P_2^ξ avec la probabilité complémentaire $\pi_2 = 1 - \pi_1$.

(ii) Le **problème de discrimination** consiste ici à déterminer la P_{θ}^ξ qui a généré l'observation ξ . Soit p_{12} (resp p_{21}) la probabilité de décider, à tort, que ξ a été tiré selon la loi P_2^ξ (resp P_1^ξ) alors que la loi qui l'a engendré est P_1^ξ (resp P_2^ξ), ie :

$$\begin{aligned} p_{21} &= P(\theta = 2 / \theta = 1), \\ (1) \quad p_{12} &= P(\theta = 1 / \theta = 2). \end{aligned}$$

La probabilité d'une imputation erronée est donc :

$$(2) \quad p = p_{21} \pi_1 + p_{12} \pi_2.$$

(iii) On suppose que les lois P_{θ}^ξ admettent resp pour **densités** pr à la **mesure de LEBESGUE** λ_K les fonctions f_{θ} ($\theta = 1, 2$).

On appelle alors **fonction discriminante de R.A. FISHER** le logarithme du **rapport des vraisemblances** (ou **rapport de vraisemblance**) f_1 / f_2 , ie :

$$(3) \quad d(\xi) = \ln \{f_1(\xi) / f_2(\xi)\} = \ln f_1(\xi) - \ln f_2(\xi).$$

Si $P_{\theta}^\xi = \mathcal{N}_K(\mu_{\theta}, \Sigma_{\theta})$ ($\theta = 1, 2$) (**loi normale multidimensionnelle**), on appelle **fonction discriminante quadratique de R.A. FISHER** la **statistique** résultant de (3), ie :

$$(4) \quad q(\xi) = (\xi - \mu_1)' \Sigma_1^{-1} (\xi - \mu_1) - (\xi - \mu_2)' \Sigma_2^{-1} (\xi - \mu_2).$$

Par suite, en posant :

$$(5) \quad \Lambda = \text{Log} (|\Sigma_2| / |\Sigma_1|) + 2 \text{Log} (\pi_1 / \pi_2),$$

la **décision** d'affectation (imputation de ξ à l'une des deux populations ou à l'une des deux lois) est déterminée comme suit :

$$(a) \text{ si } q(\xi) < \Lambda, \text{ alors } \mathcal{L}(\xi) = P_1^\xi ;$$

$$(b) \text{ si } q(\xi) \geq \Lambda, \text{ alors } \mathcal{L}(\xi) = P_2^\xi.$$

Cette règle de décision se résume selon :

$$(6) \quad \mathcal{L}(\xi) = \mathbf{1}(q(\xi) < \Lambda) \cdot P_1^\xi + \mathbf{1}(q(\xi) \geq \Lambda) \cdot P_2^\xi,$$

où $\mathbf{1}(A)$ désigne la **fonction indicatrice** d'une **partie** A.