

# Annexe N

## Indice de Rand

L'indice de Rand est une **mesure de similarité** entre deux partitions d'un ensemble [Rand, 1971]. Son principe est de mesurer la consistance entre deux partitions.

Soit  $\pi = \{u_i\}_I$  une partition de l'ensemble  $E$ , deux éléments  $e_1$  et  $e_2$  de  $E$  sont dits « groupés » dans  $\pi$  s'ils appartiennent à un même sous-ensemble de  $\pi$ , c'est-à-dire  $\exists i \in I (e_1, e_2) \in U_i \times U_i$ . On dit que deux éléments  $e_1$  et  $e_2$  sont séparés dans  $\pi$  s'ils appartiennent à deux sous-ensembles distincts de  $\pi$ .

Soient deux partitions  $\pi_1$  et  $\pi_2$  de  $E$  et soient les comptages :

- $a$  le nombre de paires éléments de  $E$  groupés dans  $\pi_1$  et dans  $\pi_2$  ;
- $b$  le nombre de paires éléments de  $E$  groupés dans  $\pi_1$ , mais séparés dans  $\pi_2$  ;
- $c$  le nombre de paires éléments de  $E$  groupés dans  $\pi_2$ , mais séparés dans  $\pi_1$  ;
- $d$  le nombre de paires éléments de  $E$  séparés dans  $\pi_1$  et dans  $\pi_2$ .

La somme  $a + d$  représente la **consistance** entre deux partitions.

La somme  $b + c$  représente l'**inconsistance** entre deux partitions.

L'indice de Rand vaut :

$$RI(\pi_1, \pi_2) = \frac{a + d}{a + b + c + d} \quad (\text{N.1})$$

c'est-à-dire la proportion des paires d'éléments qui sont conjointement groupées ou conjointement séparées.

La distance de Rand vaut :

$$d = 1 - RI(\pi_1, \pi_2) \quad (\text{N.2})$$



# Bibliographie

[Rand, 1971] RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66(336):846–850.