

Correlation Index:

Given two random variables X, Y ranges $\{x_i\}, \{y_j\}$

and Joint probability distribution $P_{ij} = \text{Prob of } X_i \text{ and } Y_j$

Then Correlation index is defined to be the expected information gain (on X) given that y is told (according to its a priori distribution):

$$\text{ie } C(X, Y) = \sum_{ij} P_{ij} \ln \left[\frac{(\sum_j P_{ij})(\sum_i P_{ij})}{P_{ij}} \right]$$

More Generally, we may have a joint distribution

on n -random Variables X^1, X^2, \dots, X^n
ranges $\{x_i^1\} \{x_j^2\} \dots$

with joint distribution $P_{i_1 i_2 \dots i_n}$

= Prob $\{X^1 = x_i^1 \text{ and } X^2 = x_j^2 \text{ and } \dots \text{ and } X^n = x_k^n\}$

and again we can define a correlation index for any

single Variable X^* and any subset of the remaining variables

$$C(X^*; X_1, X_2, \dots, X^5)$$

to be the expected

information change about X^* given that we are told
 X_1, X_2, \dots, X^5 .

$$\text{Let } N_i = \sum_{j, k, \dots, l} P_{ijk\dots l} = \text{a priori distrib on } X^*$$

$$\left\{ \sum_{j, k, \dots, l} P_{ijk\dots l} = \sum_i P_{ijk\dots l} = \text{a priori distrib on } X_1, X_2, \dots, X^5 \right.$$

$$\text{Initial Information } I^*(X) = \sum_i N_i \ln N_i$$

Conditional Information given x_1, x_2, \dots, x_l

$$I_{x_1, x_2, \dots, x_l}(X) = \sum_i P_{ijk\dots l} \ln P_{ijk\dots l}$$

information change cond on x_1, x_2, \dots, x_l

$$\Delta I_{x_1, x_2, \dots, x_l}(X) = \sum_i P_{ijk\dots l} \ln P_{ijk\dots l} - \sum_i N_i \ln N_i$$

(3)

(nearly give to
2 variables)

Expected info change

$$= \overline{\Delta I} = \sum_{j \neq i, l} \{ \Delta I_{x_i x_k \dots x_l}$$

$$= \sum_{j \neq i, l} \{ \left(\sum_i p_{ijk\dots l} \ln p_{ijk\dots l} - \sum_i n_i \ln n_i \right)$$

Cond prob dist

one given j & k

$$= \frac{p_{ijk\dots l}}{\sum_i p_{ijk\dots l}} = \frac{p_{ijk\dots l}}{\sum_{j \neq i, l}}$$

handed for subset of remaining variables

i.e. Let $P_{i, j, k, l, m, \dots, n}$

↑ ↑ ↑
information known remainder

Let $\pi_i = \sum_{j, k, l, m, \dots, n} P_{i, j, k, l, m, \dots, n} = \text{apriori on } i$

$\sum_{j, k, \dots, l} P_{i, j, k, l, m, \dots, n} = \text{a-priori on } j, k, l, \dots$

$\pi_{i, j, k, l} = \sum_{m, \dots, n} P_{i, j, k, l, m, \dots, n}$
conditional if $\pi_{i, j, k, l}$ = $\frac{P_{i, j, k, l, m, \dots, n}}{\text{Total}}$