

(1)

## Correlation index

Given joint distrib  $P_{ij, jk, \dots, l, m, \dots, n}$

We Wish a measure of the correlation of  $X^i$  with  
 $X^j X^k \dots$  and  $X^l$ , a subset of the remaining Variables.

We shall obtain this measure by imagining

that the random process selects a set of the random Variables,  $x_1, x_2, \dots, x_n$  according to the probability distribution  $P_{ij, \dots, n}$ , and that we are then told only the values  $x_j, x_k, \dots$  and  $x_l$ . This knowledge will then in general change our information about the probability of  $x^i$ , the change of information being the greater the more strongly are  $y^i$  and the  $x^j \dots x^l$

correlated. However, this information change will usually depend upon exactly which ~~values of~~<sup>values of</sup>  $x_j \dots x_e$

turn up, so that what we are really interested in is the expected information change

about  $x_i$  given that we are to be told the values  $x_j \dots x_e$ . We shall denote this quantity, called the Correlation index of  $x^i$  with  $x^j x^k \dots x^e$  by  $C(x^i, x^j, x^k, \dots, x^e)$ .

We shall use the definition of Information of a probability distrib  $P_i$  as

$$I = \sum_i p_i \ln p_i$$

(as usual in Information theory)

Now, since we are only interested in the variables  $X^i, X^j, X^K, X^L$  and not in the remainder

we define  $\bar{P}_{ijk..l} = \sum_{m..n} P_{i,j,k..l,m..n}$

so that  $\bar{P}_{ijk..l}$  is the joint (unconditional)

Probability distribution on  $X^i, X^j, X^L$  alone (Averaged over  $m..n$ )

We further define:

②  $\pi_i = \sum_{jkl} \bar{P}_{ijk..l}$  = a-priori (acond) distrib on  $X^i$

③  $\bar{\pi}_{jk..l} = \sum_i \bar{P}_{ijk..l}$  = Uncond distrib on  $X^j, X^K, \dots, X^L$

And finally, the conditional distribution on  $X^i$  given the values of  $X^j, \dots, X^L$  will be denoted by  $\pi_{ijk..l}$

④  $\pi_{ijk..l} = \frac{\bar{P}_{ijk..l}}{\sum_i \bar{P}_{ijk..l}} = \frac{\bar{P}_{ijk..l}}{\bar{\pi}_{jk..l}}$

Now, the A priori information on  $X^i$ ,  $I^o(X^i)$  is

$$⑤ \quad I^o(X^i) = \sum_i n_i \ln N_i$$

while the information on  $X^i$ , given the values of  $X^j, X^K, \dots, X^l$  (conditional)  $I_{(jk...l)}(X^i)$  is given by

$$⑥ \quad I_{(jk...l)}(X^i) = \sum_i n_{i/(jk...l)} \ln(n_{i/(jk...l)})$$

So that the change in information, <sup>upon</sup> being told that the values of  $X^j, \dots, X^l$  are  $x_j, \dots, x_l$ ,  $(\Delta I)_{jk\dots l}$  is

$$\begin{aligned} ⑦ \quad (\Delta I)_{jk\dots l} &= I_{(jk...l)}(X^i) - I^o(X^i) \\ &= \sum_i n_{i/(jk\dots l)} \ln(n_{i/(jk\dots l)}) - \sum_i n_i \ln N_i \end{aligned}$$

and now, since the values  $x_j, \dots, x_l$  will in fact turn up with probabilities  $\xi_{jk\dots l}$  we have

that the Expected change, (the correlation index) is:

$$\begin{aligned} ⑧ \quad C(X^i; X^j, X^K, \dots, X^l) &= \text{Exp}\{(\Delta I)_{jk\dots l}\} = \sum_{jk\dots l} \xi_{jk\dots l} (\Delta I)_{jk\dots l} \\ &= \sum_{jk\dots l} \left[ \xi_{jk\dots l} \left( \sum_i n_{i/(jk\dots l)} \ln(n_{i/(jk\dots l)}) - \sum_i n_i \ln N_i \right) \right] \end{aligned}$$

(4)

Computer  $C(X^i; X^0 X^K) = C(X^i; X^0) + C(X^i; X^K) - C(X^0; X^K)$

now, since by (4)  $\eta_{ijkl} \approx \frac{\theta_{ijkl}}{\xi_{jkl}}$

$$\Rightarrow (9) \quad C(X^i; X^0 X^K, X^l) = \sum_{jkl} \left[ \xi_{jkl} \left( \sum_i \left( \frac{\theta_{ijkl}}{\xi_{jkl}} \ln \left( \frac{\theta_{ijkl}}{\xi_{jkl}} \right) \right) - \sum_i n_i \ln n_i \right) \right]$$

$$= \sum_{ijkl} \theta_{ijkl} \ln \left( \frac{\theta_{ijkl}}{\xi_{jkl}} \right) - \sum_{ijkl} n_i \xi_{jkl} \ln n_i$$

but, since  $\sum_{jkl} \xi_{jkl} = 1 \Rightarrow \text{sum over } jkl \text{ gives 1}$

(10)

$$\Rightarrow C(X^i; X^0 X^K, X^l) = \sum_{ijkl} \theta_{ijkl} \ln \left( \frac{\theta_{ijkl}}{\xi_{jkl}} \right) - \sum_i n_i \ln n_i$$

$$\text{but } N_i = \sum_{ijkl} f_{ijkl}$$

$$\Rightarrow \sum_i N_i \ln N_i = \sum_i \sum_{jkl} f_{ijkl} \ln N_i$$

$$= \sum_{ijkl} f_{ijkl} \ln N_i$$

$$\Rightarrow C = \sum_{ijkl} f_{ijkl} \left( \ln \left( \frac{f_{ijkl}}{\sum_{jkl} f_{jkl}} \right) - \ln N_i \right)$$

$$\Rightarrow \textcircled{II} \quad C(x^i_j x^j_k \dots x^l) = \sum_{ijkl} f_{ijkl} \ln \left( \frac{f_{ijkl}}{\sum_{jkl} N_i} \right)$$

(5)

$$\text{conjecture } C(X^i; X^j \dots X^l)$$

$$+ C(X^m; X^i \dots X^l) - C(X^i; X^m) = \text{Expected gain on } X^i \text{ and } X^m$$

Must really generalize a little more to get

$$C(\bar{X}^{\alpha} \dots \bar{X}^{\delta}; X^{\delta} \dots X^{\omega}) \quad (\text{simply convolution in joint distributions})$$

should get formula:

$$C(\bar{X}^{\alpha} \dots \bar{X}^{\delta}; X^{\delta} \dots X^{\omega}) = \sum_{\alpha, \beta, \gamma, \delta, \omega} q_{\alpha, \beta, \gamma, \delta, \omega} \ln \left( \frac{q_{\alpha, \beta, \gamma, \delta, \omega}}{\sum_{\delta, \omega} q_{\alpha, \beta, \gamma, \delta, \omega}} \right)$$

then Work out Algebra of C's

$$C(X_1 X_2 \dots X_n) = \sum_{\alpha, \beta} f_{\alpha, \beta} \ln \left( \frac{f_{\alpha, \beta}}{\sum_{\gamma, \delta} n_{\gamma, \delta}} \right)$$

Alternate form

$$= \sum_{\alpha, \beta} f_{\alpha, \beta} \ln f_{\alpha, \beta} - \sum_{\alpha, \beta} f_{\alpha, \beta} \ln \sum_{\gamma, \delta} n_{\gamma, \delta} - \sum_{\alpha, \beta} f_{\alpha, \beta} \ln \frac{n_{\alpha, \beta}}{f_{\alpha, \beta}}$$

$$\sum_{\gamma, \delta} \left( \sum_{\alpha, \beta} f_{\alpha, \beta} \right) \ln \frac{n_{\gamma, \delta}}{\sum_{\alpha, \beta} f_{\alpha, \beta}}$$

$$C = \sum_{\alpha, \beta} f_{\alpha, \beta} \ln f_{\alpha, \beta} - \sum_{\gamma, \delta} \sum_{\alpha, \beta} n_{\gamma, \delta} \ln \frac{n_{\gamma, \delta}}{\sum_{\alpha, \beta} n_{\gamma, \delta}} - \sum_{\alpha, \beta} n_{\alpha, \beta} \ln \frac{n_{\alpha, \beta}}{\sum_{\gamma, \delta} n_{\gamma, \delta}}$$

- Information in Joint distrib - Inform, uncond  $X_1 \dots X_n$  - Inform uncond  $X_1 \dots X_n$
- Inform in Joint distrib - (sum of inform in unconditional)

$$\text{Notation: } I(X; X_1 \dots X^8) = I_{\alpha, \beta, \dots, \gamma}$$

= unconditional inform in distribution

i.e. if Total Joint distrib  
is  $P_{\alpha, \beta, \dots, \gamma, \delta, \dots, \omega}$

then Unconditional Joint distrib on  $\alpha, \beta, \dots, \gamma$  only

$$\text{is } f_{\alpha, \beta, \dots, \gamma} = \sum_{\delta, \dots, \omega} P_{\alpha, \beta, \dots, \gamma, \delta, \dots, \omega}$$

$$\text{and } I_{\alpha, \beta, \dots, \gamma} = \sum_{\alpha, \beta, \dots, \gamma} f_{\alpha, \beta, \dots, \gamma} \ln f_{\alpha, \beta, \dots, \gamma}$$

$$= \sum_{\alpha, \beta, \dots, \gamma} \left( \sum_{\delta, \omega} P_{\alpha, \beta, \dots, \gamma, \delta, \omega} \right) \ln \left( \sum_{\delta, \omega} P_{\alpha, \beta, \dots, \gamma, \delta, \omega} \right)$$

$$= \sum_{\alpha, \beta, \dots, \gamma} P_{\alpha, \beta, \dots, \gamma} \ln \left( \sum_{\delta, \omega} P_{\alpha, \beta, \dots, \gamma, \delta, \omega} \right)$$

$$\ln \left( P_{\alpha, \beta, \dots, \gamma} \cdot \frac{\sum_{\delta, \omega} P_{\alpha, \beta, \dots, \gamma, \delta, \omega}}{P_{\alpha, \beta, \dots, \gamma}} \right)$$

$$= \sum_{\alpha, \beta, \delta, \omega} P_{\alpha, \beta, \delta, \omega} \ln \frac{P_{\alpha, \beta, \delta, \omega}}{P_{\alpha, \beta, \dots, \gamma}} + \sum_{\alpha, \beta, \delta, \omega} P_{\alpha, \beta, \delta, \omega} \ln \left( \frac{P_{\alpha, \beta, \delta, \omega}}{\sum_{\delta, \omega} P_{\alpha, \beta, \delta, \omega}} \right)$$

$$= I_{\alpha, \beta, \delta, \omega} - \sum_{\alpha, \beta, \delta, \omega} P_{\alpha, \beta, \delta, \omega} \ln \left( \frac{P_{\alpha, \beta, \delta, \omega}}{\sum_{\delta, \omega} P_{\alpha, \beta, \delta, \omega}} \right)$$

but  $\frac{P_{\alpha\beta\delta\omega}}{\sum_{\delta\omega} P_{\alpha\beta\delta\omega}} = \text{cond: } P_{\delta\omega/\alpha\beta\gamma}$

so that  $I_{\alpha\beta\gamma} = I_{\alpha\beta\delta\omega} - \sum_{\delta\omega} P_{\alpha\beta\delta\omega} \ln P_{\delta\omega/\alpha\beta\gamma}$

$$\begin{aligned} & \sum_{\delta\omega} (\sum_{\alpha\beta\gamma} P_{\alpha\beta\delta\omega}) \ln P_{\delta\omega/\alpha\beta\gamma} \\ &= \sum_{\delta\omega} P_{\delta\omega} \ln P_{\delta\omega/\alpha\beta\gamma} \end{aligned}$$

now,  $\sum_{\alpha\beta\delta\omega} P_{\alpha\beta\delta\omega} \ln P_{\delta\omega/\alpha\beta\gamma}$

General Formula  $C(\alpha\beta\gamma; \delta\omega) = I_{\alpha\beta\delta\omega} - I_{\alpha\beta\gamma} - I_{\delta\omega}$

$$\Rightarrow C(\alpha\beta\gamma\delta) = I_{\alpha\beta\delta} - I_{\alpha\beta\gamma} - I_{\beta\gamma\delta}$$

$$\Rightarrow \cancel{I_{\beta\gamma\delta}} = I_{\alpha\beta\delta} - I_{\alpha} - C(\alpha\beta\delta)$$

$$C(\alpha, \beta) = I_{\alpha\beta} - I_\alpha - I_\beta$$

$$C(\beta, \gamma) = I_{\beta\gamma} - I_\beta - I_\gamma$$

$$C(\gamma\delta) = I_{\gamma\delta} - I_\gamma - I_\delta$$

Conditional inform  $I_{\alpha\beta|\gamma\delta}$

set by  $I_{\alpha\beta}^{\gamma\delta} = \sum_{\alpha\beta} P_{\alpha\beta|\gamma\delta} \ln P_{\alpha\beta|\gamma\delta}$

or let  $P_{\alpha\beta}^{\gamma\delta}$  be cond prob distrib on  $\alpha, \beta$   
given  $\gamma, \delta$

$\Rightarrow$  orig distrib  $P_{\alpha\beta|\gamma\delta}$

$$\Rightarrow P_{\alpha\beta|\gamma\delta} = \sum_{\alpha\beta} P_{\alpha\beta|\gamma\delta}$$