

①

discrete approximations to continuous distributions  
 Rigorous Def of Carr., Existence guaranteed

Given continuous  $P(x, y)$

divide into small intervals  $\Delta x_i, \Delta y_j$   
 drop  $x$  to  $x_i, y$  to  $y_j$

Prob of  $P(\Delta x_i, \Delta y_j) = P_{ij} = P(x_i, y_j) \Delta x_i \Delta y_j$

$$C^* = \sum_{i,j} P_{ij} \ln \frac{P_{ij}}{P_i P_j}$$

$$= \sum_{i,j} P(x_i, y_j) \Delta x_i \Delta y_j \ln \frac{P(x_i, y_j) \Delta x_i \Delta y_j}{\sum_k P(x_k, y_j) \Delta x_k \Delta y_j \sum_l P(x_i, y_l) \Delta x_i \Delta y_l}$$

$$= \sum_{i,j} P(x_i, y_j) \Delta x_i \Delta y_j \ln \frac{P(x_i, y_j)}{\sum_k P(x_k, y_j) \Delta x_k \sum_l P(x_i, y_l) \Delta y_l}$$

also, at any stage of approx is  $\leq$  ↓

try to show it is prove b/w always exists ↑ with other

and  
lim  
as  $\Delta x, \Delta y \rightarrow 0$

$$= \int P(x, y) \ln \frac{P(x, y)}{P(x) P(y)} dx dy = C(x, y)$$

so OK

Take the bain def of Correlation as lim of ↑ process, then  
works for continuous - discrete mixtures. Also works  
when  $I_{xy}$  and  $I_x$  &  $I_y$  are infinite and not finite as usual.

In fact, due to invariance property, says that info measure  
can be changed appropriately so that  $C$  is not undetermined  
(ie so that  $I_{xy}$  and  $I_x$  &  $I_y$  all finite  $\Rightarrow$  give correct  $C$ .)

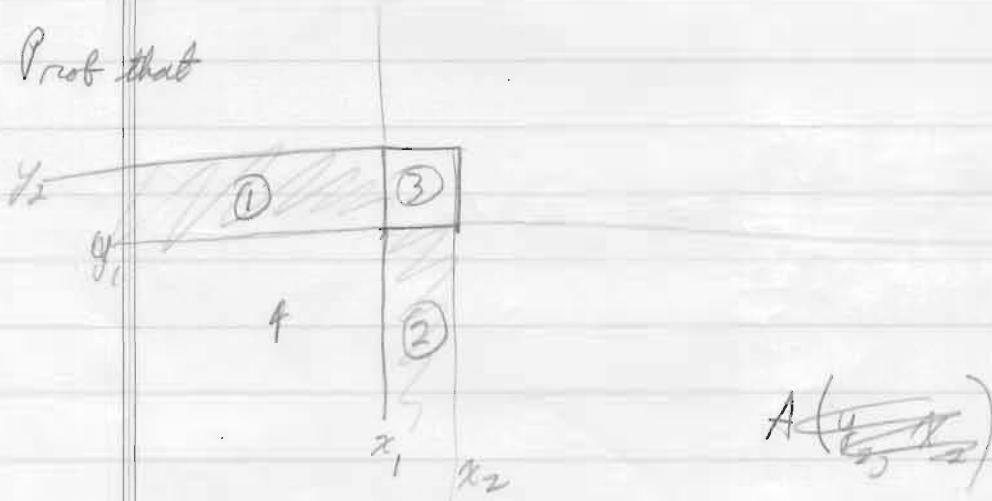
## Cumulative Distributions :

$$A(x', y') = \text{Prob } X < x' \text{ and } Y < y'$$

$$A(x', \infty) = \text{Prob } X < x' \quad (\text{marginal cumulative})$$

$$A(\infty, y') = \text{other marginal}$$

Prob that



$$A(x_1, y_1)$$

$$P(\text{region 4}) = A(x_2, y_2)$$

$$P(\text{region 1}) = A(x_1, y_2) - A(x_1, y_1)$$

$$(P\text{rot of 2}) = A(x_2, y_1) - A(x_1, y_1)$$

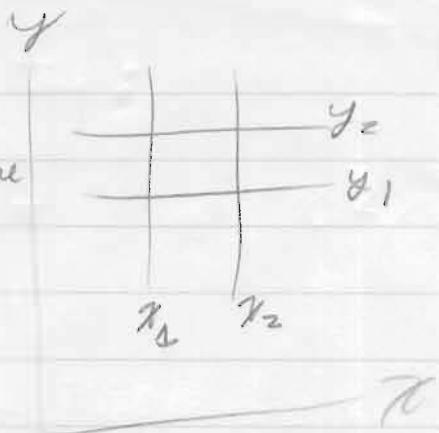
$$\underbrace{P\text{rot of 3}}_{=} = A(x_2, y_2) - \textcircled{1} - \textcircled{2} - \textcircled{3}$$

$$= A(x_2, y_2) - A(x_1, y_2) + A(x_1, y_1)$$

$$- A(x_2, y_1) + \cancel{A(x_1, y_1)} - \cancel{A(x_2, y_1)}$$

(3)

$\therefore$  Prob of square



$$\text{in } P = A(x_2, y_2) - A(x_1, y_2) - A(x_2, y_1) + A(x_1, y_1)$$

so, if  $X$  divided into ordered sequence  $x_i$   
 $y$  into  $y_j$

and interpret  $P_{i,j}$  as prob of square  $x_i$  to  $x_{i+1}$   
 and  $y_j$  to  $y_{j+1}$

$$P_{i,j} = A(x_{i+1}, y_{j+1}) - A(x_i, y_{j+1}) - A(x_{i+1}, y_j) + A(x_i, y_j)$$

$$\text{while } P_i = A(x_{i+1}, \infty) - A(x_i, \infty)$$

$$P_j = A(\infty, y_{j+1}) - A(\infty, y_j)$$

~~Q~~ Does refinement always increase C?

affine only ie  $P_{ij} \rightarrow P'_{i,j}$  and  $P'_{i+1,j}$

such that

$$P_{i-} + P_{i+} = P_{ij}$$

$$P_{i-} = \sum_j P_{i-j}$$

$$P_{i+} = \sum_j P_{i+j}$$

otherwise some

$$P_{i-} + P_{i+} = P_i$$

$$I' = \sum_{i+j} P'_{ij} \ln \frac{P'_{ij}}{P_i P_j}$$

$$= \sum_i P'_{i-} \ln \frac{P'_{i-}}{P_i P_j} + \sum_i P'_{i+} \ln \frac{P'_{i+}}{P_i P_j} + \sum_j P'_{i+} \ln \frac{P'_{i+}}{P_i P_j}$$

compared to  $\sum_i P'_{i-} \ln \frac{P'_{i-}}{P_i P_j} + \sum_j P'_{i+} \ln \frac{P'_{i+}}{P_i P_j}$

(5)

turn to be

$$\text{compared} = \sum_j p_{ij} \ln \frac{p_{ij}}{p_i p_j} \quad i = 1 \\ i = 2$$

$$= \sum_j (p_{1j} + p_{2j}) \ln \frac{p_{1j} + p_{2j}}{(p_1 + p_2) p_j}$$

compared to  $\sum_j p_{1j} \ln \frac{p_{1j}}{p_1 p_j} + \sum_j p_{2j} \ln \frac{p_{2j}}{p_2 p_j}$

$$= \sum_j (p_{1j} + p_{2j}) \ln \left( \frac{p_{1j} + p_{2j}}{(p_1 + p_2) p_j} \right) + \sum_j (p_{1j} + p_{2j}) \ln \frac{p}{(p_1 + p_2) p_j}$$

in particular

need that

$$(p_{1j} + p_{2j}) \ln \frac{p_{1j} + p_{2j}}{(p_1 + p_2) p_j} \leq p_{1j} \ln \frac{p_{1j}}{p_1 p_j} + p_{2j} \ln \frac{p_{2j}}{p_2 p_j}$$

$$\text{Now, } (P_{1j} + P_{2j}) \ln \frac{P_{1j} + P_{2j}}{(P_1 + P_2) P_j} \leq (P_{1j} + P_{2j}) \ln \frac{P_{1j} + P_{2j}}{P_1 P_j}$$

and  $\leq$

$$\frac{1}{\frac{P_1 P_j}{P_2}}$$

$$\text{ie } (\alpha_1 + \alpha_2) \ln \frac{\alpha_1 + \alpha_2}{(P_1 + P_2) P_3} \leq \alpha_1 \ln \frac{\alpha_1}{P_1 P_3} + \alpha_2 \ln \frac{\alpha_2}{P_2 P_3}$$

~~$$-(\alpha_1 + \alpha_2) \ln \frac{(\alpha_1 + \alpha_2)}{(P_1 + P_2)} - (\alpha_1 + \alpha_2) \ln P_3 \leq \alpha_1 \ln \frac{\alpha_1}{P_1} + \alpha_2 \ln \frac{\alpha_2}{P_2}$$~~
~~$$-\alpha_1 \ln P_3 - \alpha_2 \ln P_3$$~~

need  
only  
prove:

$$(\alpha_1 + \alpha_2) \ln \frac{\alpha_1 + \alpha_2}{P_1 + P_2} \leq \alpha_1 \ln \frac{\alpha_1}{P_1} + \alpha_2 \ln \frac{\alpha_2}{P_2}$$

$$= (\alpha_1 + \alpha_2) \ln(\alpha_1 + \alpha_2) - (\alpha_1 + \alpha_2) \ln(P_1 + P_2)$$

True by #117 of Hardy, Littlewood & Polya, namely

$$x \ln \frac{x}{a+b} + y \ln \frac{y}{a+b} \geq (x+y) \ln \frac{x+y}{a+b}$$

So Prove as Theorem

given discrete joint distribution, Correlation  
of any blocks ~~always~~ never increases  
correlation.

Then reverse to prove that  
Correlation is always defined for any ~~block~~

Jid Prob. Distribution simply as limit of approximations  
by discrete distributions. Since each refinement  
does not decrease correlation, the sequence of  
approximating correlations is monotone ↑, hence  
approaches a limit (which may of course be infinite) nevertheless  
correlation always defined. Moreover this is independent  
of the relative sizes of the patches, hence invariant  
to any scale changes (not only is measure of joint  
is product measure of margins)

hence given joint distrib with spikes, approx by patches  
gives in limit C for the continuous part  
+ C for the discrete part



(single spike  $\frac{1}{4}$  size gives  $\frac{1}{4}$  spike in each  
final Marginal gives  $\frac{1}{4} \ln \frac{1}{16}$   
 $= \frac{1}{4} \ln 4$  to go)

Isolated Spike, magnitude  $\propto$

contributes  $\propto \ln \frac{\alpha}{\alpha^2} = \propto \ln \frac{1}{\alpha} = -\alpha \ln \alpha$

~~so sum~~

Correlation

Note since inequality

$$x \ln \frac{x}{a} + y \ln \frac{y}{b} > (x+y) \ln \frac{x+y}{a+b}$$

$$= x \ln \frac{x}{a} + y \ln \frac{y}{b} + z \ln \frac{z}{c} \geq (x+y) \ln \frac{(x+y)}{a+b} + z \ln \frac{z}{c}$$

$$\Rightarrow 3(x+y+z) \ln \frac{x+y+z}{a+b+c} \text{ by iteration}$$

$\Rightarrow$  by induction

$$\sum_i x_i \ln \frac{x_i}{a_i} \geq \left( \sum_i x_i \ln \frac{\sum x_i}{\sum a_i} \right)$$

so that def holds for group correlations

Q

## 51 Probability distributions

with  
respect  
 $P_{X,Y}$

Let  $P(x, y, \dots, z)$  represent a joint probability density for the random variables  $X, Y, \dots, Z$ , whose values are real numbers, so that  $P(x, y, \dots, z) dx dy \dots dz$  is interpreted to mean the probability that  $X$  will take a value in the interval  $[x, x+dx]$  and  $Y$  will take a value in  $[y, y+dy]$  and ... and  $Z$  will take a value ~~between~~ in  $[z, z+dz]$ .

We now define a marginal probability for any subset of the original random variables, say  $X, Y$  to be the integral of the total joint density with respect to the remaining variables, i.e.

$$P(x, y) = \int_{z \dots w} P(x, y, z \dots w) dz \dots dw$$

which represents the density for the joint event  $X \in [x, x+dx]$  and  $Y \in [y, y+dy]$  with no restrictions on the remaining variables. It is the relative density over  $X \dots Y$  when we have no information about the other variables.

Finally, we define a conditional density for any subset  $X \dots Y$ , conditioned on any remaining variables having prescribed values, say  $Z = z \dots W = w$ , denoted by  $P_{x \dots y}^{z \dots w}$  to be:

$$P_{x \dots y}^{z \dots w} = \frac{P_{x \dots y \dots z \dots w}}{P_{z \dots w}}$$

which represents the density for  $x \dots y$  relative to the knowledge that  $Z \dots W$  have the definite values  $z \dots w$ .