# 信息流的模式及其分解

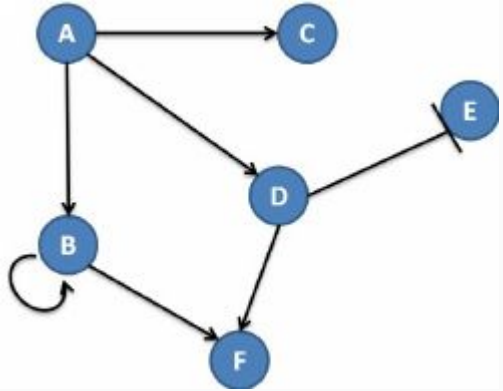
## 从互信息到密码学

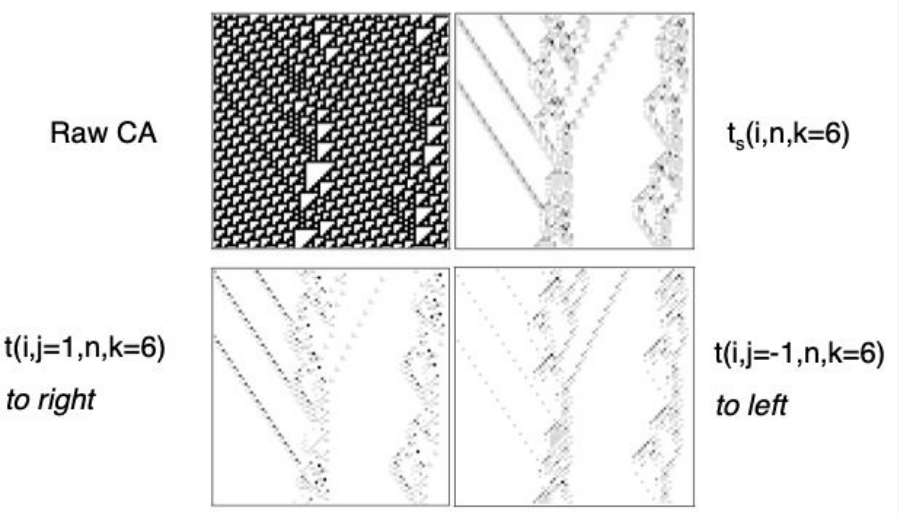
章彦博 Arizona State University

# 信息的流动

# 为什么要测量信息流？

- 对系统进行建模
- 从中观尺度观察复杂系统




Lizier, J. T., Prokopenko, M., & Zomaya, A. Y. (2007, December). Information transfer by particles in cellular automata. In Australian Conference on Artificial Life (pp. 49-60). Springer, Berlin, Heidelberg.

# 熵与信息

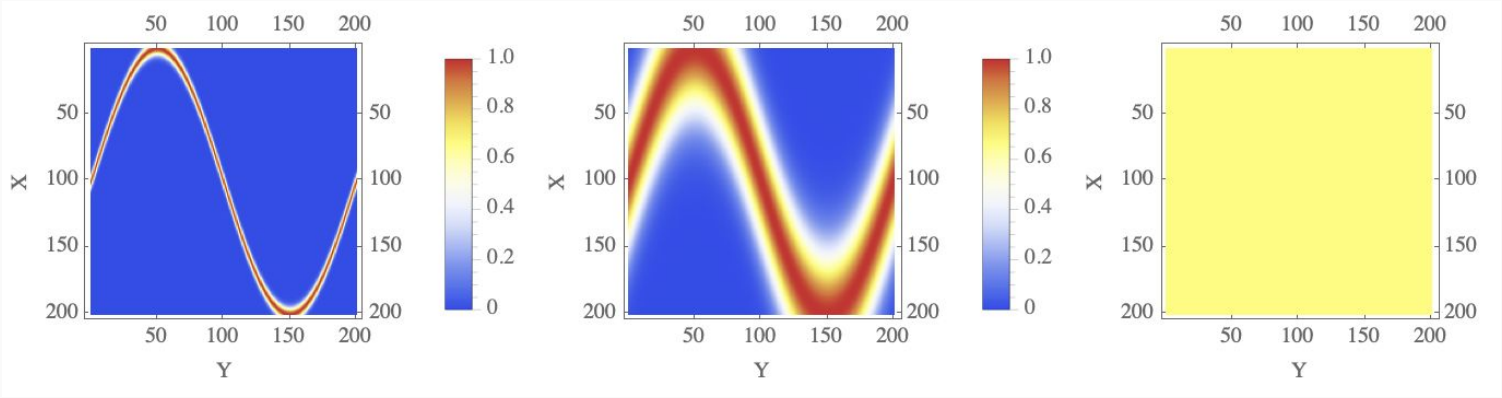
- 随机变量：X，在不掌握任何知识的情况下，X可以是0，也可以是1；
- X的熵H(X)=1 bit
- 信息：消除不确定性的程度
  - Y：明天是否下雨
    - 多雨的城市，可能是50%的概率，H(Y)=1bit：准确的天气预报提供了1bit的信息；
    - 干旱的城市，5%的概率，H(Y)=0.29bit：准确的天气预报提供了0.29bit的信息；

$$H(X) = \mathbb{E}[-\log p(X)]$$

# 互信息

- 两个随机变量的关联程度：相较于独立分布多出来的信息

$$I(X;Y) = D_{\mathrm{KL}}(P_{(X,Y)} \| P_X \otimes P_Y)$$




从左到右，互信息逐渐降低

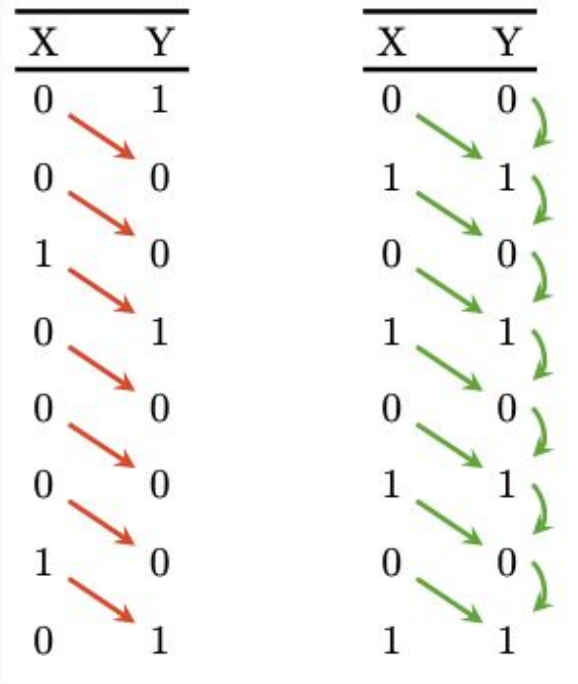
# 使用互信息度量信息流:时滞互信息

- 度量X→Y的信息流:当前的Y与过去的X的互信息

$$\text{flow} = I(X_{-1}; Y_0)$$

- 没有排除自身的作用:$X_{-1}$与$Y_{-1}$都可以预测$Y_0$

- 如何去除$Y_{-1}$的影响?



# 条件时滞互信息: 转移熵（transfer entropy）

- 「消去」自身因素的影响：引入条件

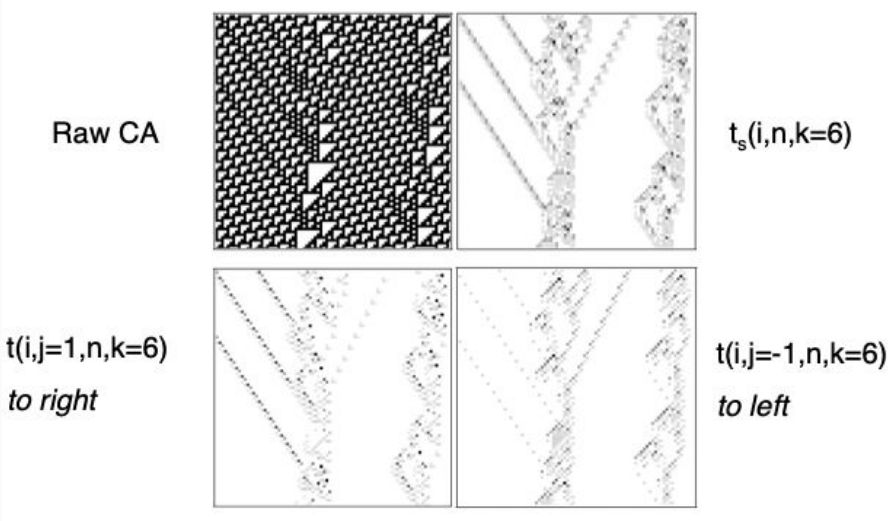
$$T_{X\to Y} = I(X_{-1}; Y_0|Y_{-1}) = H(Y_0|Y_{-1}) - H(Y_0|Y_{-1}, X_{-1})$$

  - 已知$Y_{-1}$，此时$X_{-1}$还能带来的信息
  - 格兰杰因果：转移熵在线性自回归过程上的特例
- Causation Entropy (Sun, Collt)：进一步消除第三方影响
  - $C_{X\to Y|(Y,Z)} = I[Y_t : X_{0:t} | Y0_{:t'} Z_{0:t}]$
  - 仍然使用「条件」来排除影响

# 条件时滞互信息: 转移熵（transfer entropy）

- 研究元胞自动机中的信息流动
  - Local transfer entropy
  - 展现复杂系统中的信息流动

- 条件互信息真的消去了自身的影响了吗？并非如此！




条件时滞互信息: 转移熵

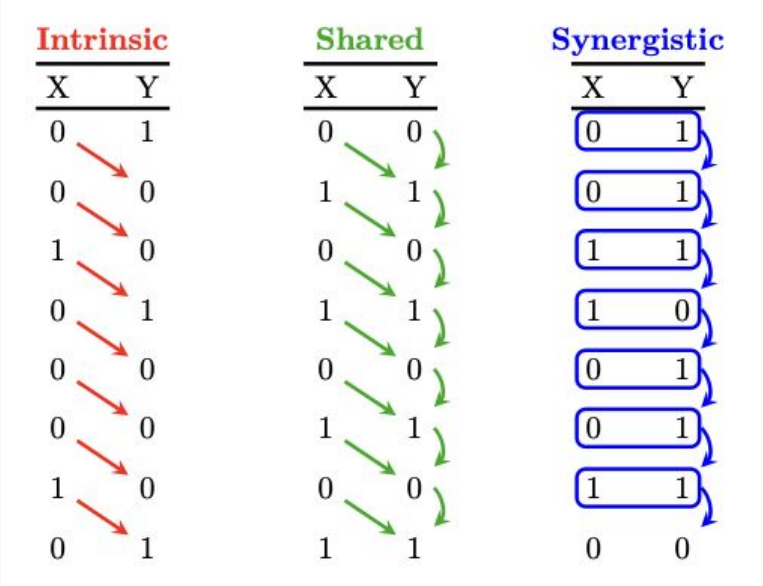
- 「消去」自身因素的影响：引入条件

# 信息流的分类

# 三种信息流

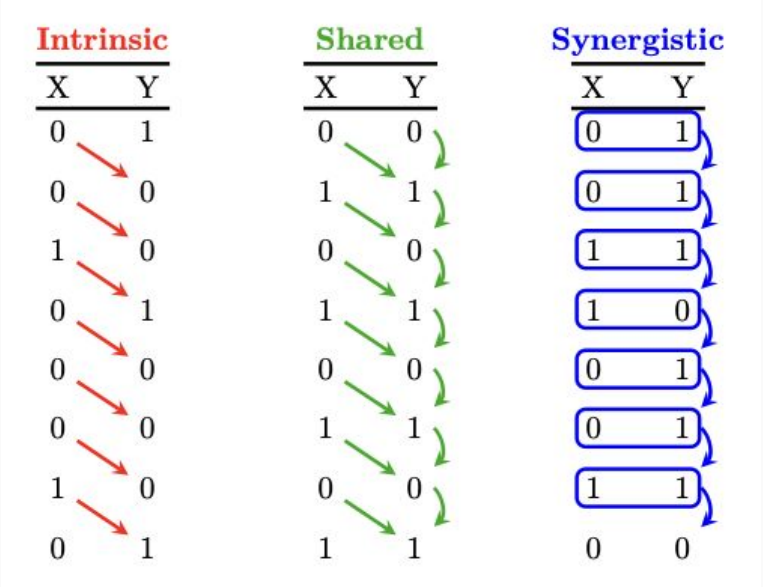
- 固有信息流（Intrinsic flow）
- 共享信息流（Shared flow）
- 协同信息流（Synergistic flow）

- 时滞互信息：混淆了固有、共享信息流，忽略了协同信息流；
- 转移熵：混淆了固有、协同信息流



# 时滞互信息的问题

- 无法区分固有信息和共享信息:因 为它没有考虑别的变量也可以预测$Y_0$



# 转移熵的问题

- 无法区分固有信息和 协同信息
- 引入「条件Y」，并非真的消去了Y的影响，同时也引入了Y的另一种影响

- $I(Y_0;X_{-1})=0$
- $I(Y_0;Y_{-1})$
- $I(Y_0;X_{-1}|Y_{-1})=1$ bit ← 「条件Y」引入了新的信息
  - PID的视角：去掉了共享信息的同时，又引入了协同信息

转移熵也可能来自$X_{-1}+Y_{-1}$，如何真正消除$Y_{-1}$的影响？

**Synergistic**

| $X_{-1}$ | $Y_0$ | $Y_{-1}$ | Pr |
|---|---|---|---|
| 0 | 0 | 0 | 1/4 |
| 0 | 1 | 1 | 1/4 |
| 1 | 0 | 1 | 1/4 |
| 1 | 1 | 0 | 1/4 |

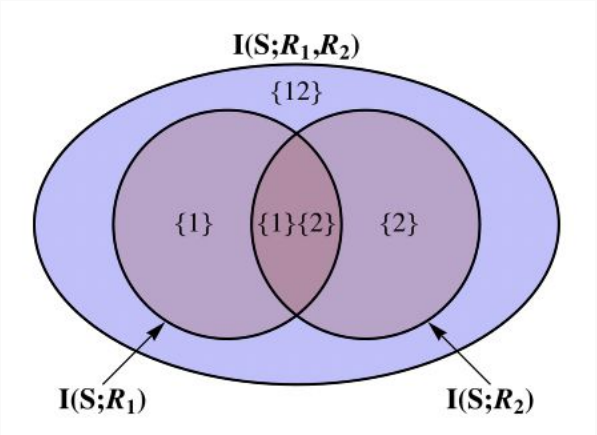
intrinsic = 0

shared = 0

synergistic = 1

# PID的视角

- 要去掉的恰恰就是{1}{2}这部分，在PID中称为「冗余信息」
- 如果将$R_1$对应$X_{-1}$，$R_2$对应$Y_{-1}$，而S对应$Y_0$，那么本文是直接对{1}进行估计。



# 来自密码学的方法

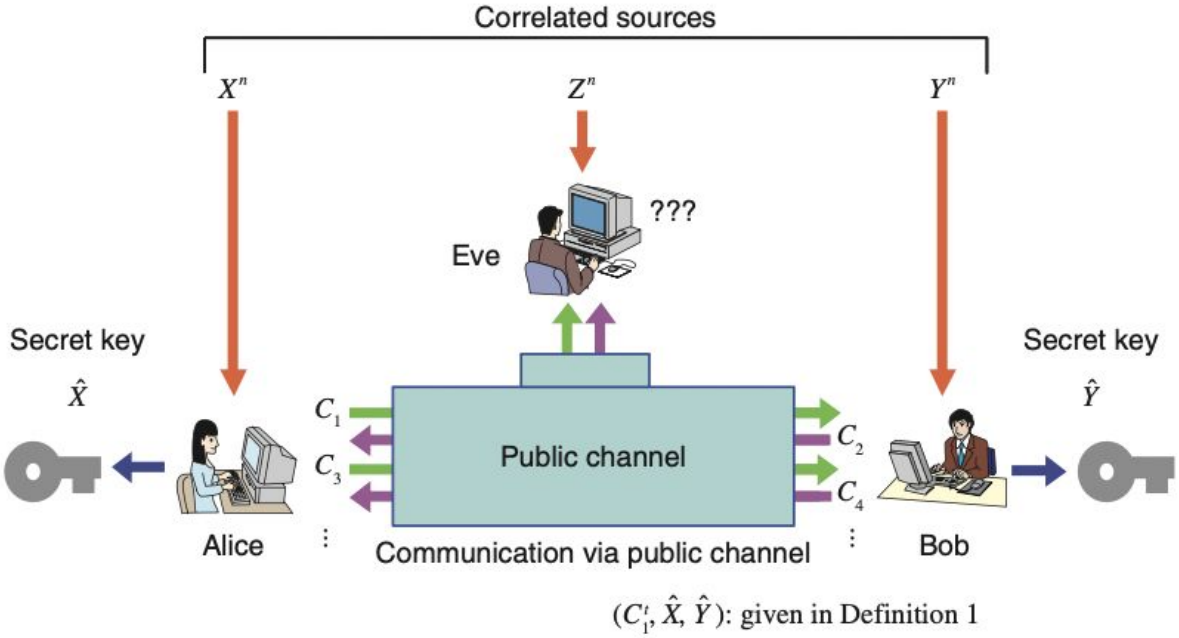
# 信息流与加密通信

想要把密钥分发给Alice和Bob，但又不想让窃听者Eve获得信息：

- 在Alice与Bob的信息中，减去Eve可以获得的那一部分
- Alice = $X_{-1}$, Bob=$Y_0$, Eve=$Y_{-1}$
- 问题：在Eve无法窃听的条件下，Alice与Bob之间最多能共享多少信息？

这个信息，就是密钥一致率（secret-key agreement rate），或密钥容量（？）

- 核心思想：做一个真正的减法




密钥传输协议：

| i | 方向 | 内容 |
|---|---|---|
| 1 | Alice→Bob | $C_1 = h_X(C_1^0; X^n) = h_X(;X^n)$ |
| 2 | Alice←Bob | $C_2 = h_Y(C_1^1; Y^n)$ |
| 3 | Alice→Bob | $C_3 = h_X(C_1^2; X^n)$ |
| ... | | |

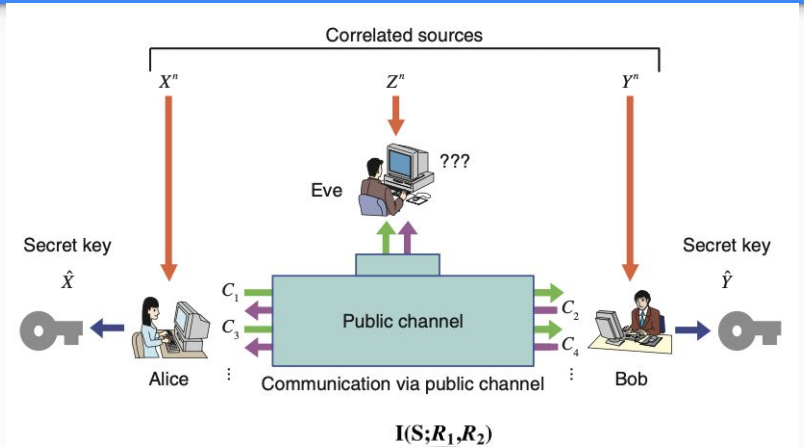
传输完成之后，计算密钥：

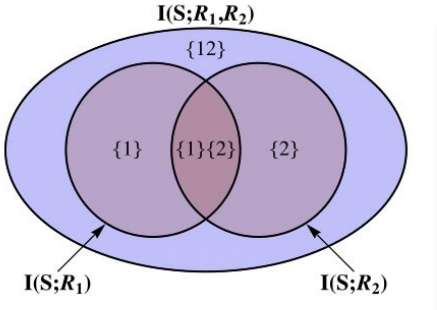
- $\hat{X} = f(C_1^t; X^n)$;
- $\hat{Y} = g(C_1^t; Y^n)$;

此时，窃听者只能看到$C_1^t$，而不能看到$X^n$和$Y^n$。

# 密码学视角下的信息流

- 一个类比
  - Alice: $X_{-1}$;
  - Bob: $Y_0$;
  - Eve: $Y_{-1}$;
- 去掉$Y_{-1}$的影响：在保证信息不被Eve窃听的情况下，Alice与Bob可以通信的信息量



# 估算密钥一致率

- 计算其上界：偷听者不光能获得Z，还能获得所有做了局域修改的Z
  - 遍历所有可能的Z，条件互信息的最小值，就是密钥一致率的上界
  - 如何遍历Z bar: 将Z中的元素，按照任意条件概率，映射到某个 值。例如{0→1}.

$$S(X:Y\|Z) \leq \min_{\Pr(\bar{z}|z)} I[X:Y|\bar{Z}]$$

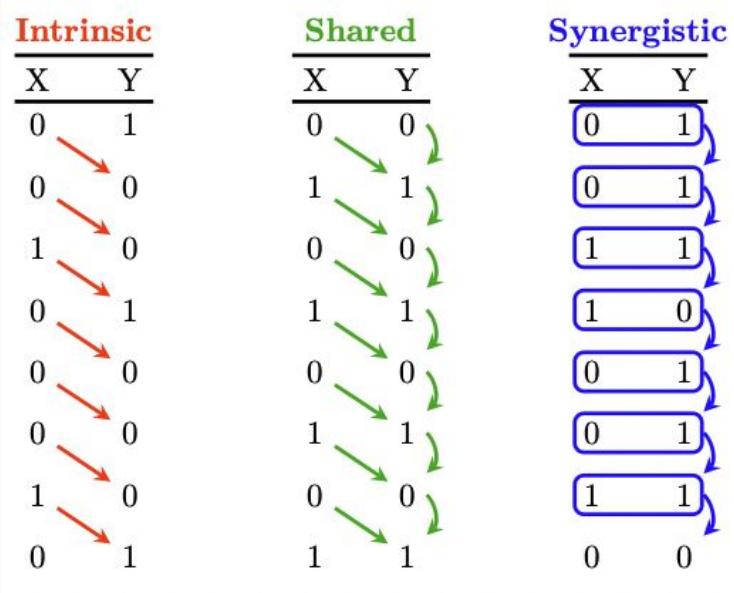
$$= I[X:Y \downarrow Z]$$

$$I[X:Y \downarrow Z] \to I[X_{-1}:Y_0 \downarrow Y_{-1}]$$

$$\bar{Z} \sim P_{\bar{z}|z}$$

# 真正地排除不需要的信息

- 固有信息流：$S(X_{-1}:Y_0||Y_{-1})$ —— 条件互信息 - Y的信息
- 共享信息流：$I(X_{-1};Y_0)-S(X_{-1}:Y_0||Y_{-1})$ —— $X_{-1}$与$Y_0$的互信息减去固有信息流
- 协同信息流：$I(X_{-1};Y_0|Y_{-1})-S(X_{-1}:Y_0||Y_{-1})$ —— 转移熵 - 固有信息
  - 转移熵 = 固有信息流+协同信息流



| $X$ | $Y$ | $Z$ | $\overline{Z}$ | Pr |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | $1/8$ |
| 0 | 1 | 1 | 0 | $1/8$ |
| 1 | 0 | 1 | 0 | $1/8$ |
| 1 | 1 | 0 | 0 | $1/8$ |
| 2 | 2 | 2 | 2 | $1/4$ |
| 3 | 3 | 3 | 3 | $1/4$ |

$$\mathrm{I}[X:Y] = 3/2\,\mathrm{bit}$$

$$\mathrm{I}[X:Y \mid Z] = 1/2\,\mathrm{bit}$$

$$\mathrm{I}[X:Y \downarrow Z] = 0\,\mathrm{bit}$$

$$a \to b$$

$$b \begin{bmatrix} & a=0 & a=1 & a=2 & a=3 \\ b=0 & \frac{1}{2} & 0 & 1 & 1 \\ b=1 & 0 & \frac{1}{2} & 1 & 1 \\ b=2 & 1 & 1 & \frac{1}{2} & 1 \\ b=3 & 1 & 1 & 1 & \frac{1}{2} \end{bmatrix}$$

$\Pr(\bar{z}|z)$不同时，对应的I[X:Y|Z]

$$\Pr_{\bar{z}|z}(0|1) = 1$$

$$\Pr_{\bar{z}|z}(0|0) = 1$$

$$\Pr_{\bar{z}|z}(2|2) = 1$$

$$\Pr_{\bar{z}|z}(3|3) = 1$$

**Intrinsic**

| $X_{-1}$ | $Y_0$ | $Y_{-1}$ | Pr |
|---|---|---|---|
| 0 | 0 | 0 | 1/4 |
| 0 | 0 | 1 | 1/4 |
| 1 | 1 | 0 | 1/4 |
| 1 | 1 | 1 | 1/4 |

intrinsic = 1
shared = 0
synergistic = 0

**Shared**

| $X_{-1}$ | $Y_0$ | $Y_{-1}$ | Pr |
|---|---|---|---|
| 0 | 1 | 0 | 1/2 |
| 1 | 0 | 1 | 1/2 |

intrinsic = 0
shared = 1
synergistic = 0

**Synergistic**

| $X_{-1}$ | $Y_0$ | $Y_{-1}$ | Pr |
|---|---|---|---|
| 0 | 0 | 0 | 1/4 |
| 0 | 1 | 1 | 1/4 |
| 1 | 0 | 1 | 1/4 |
| 1 | 1 | 0 | 1/4 |

intrinsic = 0
shared = 0
synergistic = 1

$$\Pr_{\bar{z}|z}(0|1) = 1 \text{ or } \Pr_{\bar{z}|z}(1|0) = 1$$

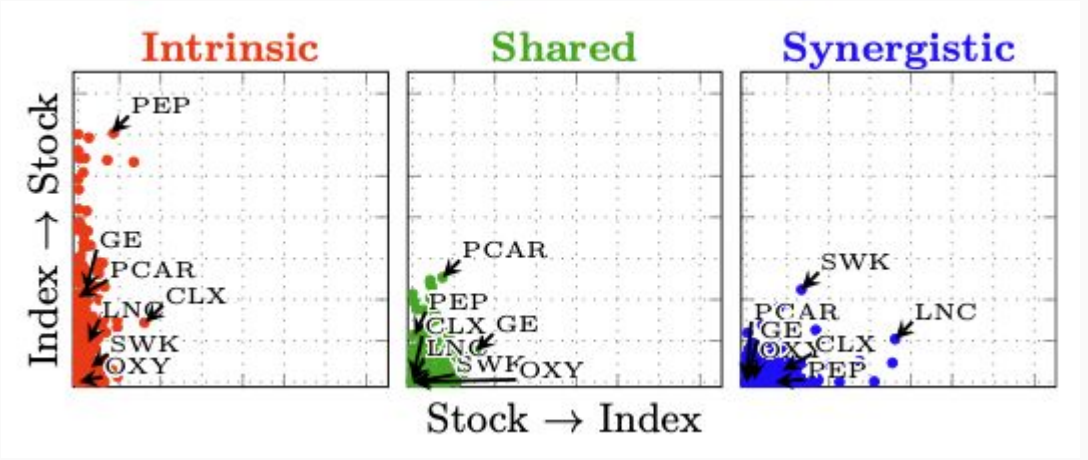
都能得到 $I[X_{-1}; Y_0 \downarrow Y_{-1}] = 0$

# 信息流拆分的一个应用:标普五百数据

- 信息流存在方向
  - X→Y和Y→X是不一样的
- 研究S&P500与个股之间的信息流
  - S&P 500 → stock[i];
  - stock[i] → S&P 500

# 信息流拆分的一个应用:标普五百数据

- (S&P 500 → stock) >> (stock → S&P 500)
  - 一种因果倒置？



# 信息流分类的启发

- 除了固有信息流，其他两种（共享、协同）都是多对一的关系
- 网络科学默认使用「一对一」的结构，从而会忽略多对一的关系
  - NLP中也有这种现象，如word2vec（看似多对一，实则一对一）
- 如何表示多对多的结构？
  - Wolfram: Hypergraph
  - Petri Net
  - 守恒量：c=f(X,Y,Z)
  - 高阶网络