

Além da correlação: detectando causalidade em séries de observações

Renato Mendes Coutinho

Santo André, 20 de junho de 2018

Centro de Matemática, Computação e Cognição

UFABC - Santo André

renato.coutinho@ufabc.edu.br



Introdução

Descobrir **mecanismos** – ou seja, relações de causa e efeito – é central à ciência

O que quer dizer “A causa B”?

Como se determinam relações causais

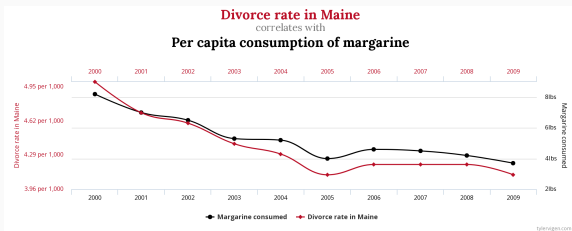
- *Observação direta*
- Manipulação das variáveis: experimentos
- Tratamento vs. controle
- Observações – “experimentos naturais”

- Indicador vs. mecanismo
 - exemplo: barômetro e chuva
- *Post hoc*: depois, logo porque
 - exemplo: vacinação e autismo

Causalidade em séries de observações (ou séries temporais)

Correlações espúrias: **correlação não implica causalidade**

Muitos exemplos esdrúxulos, mas também acontece em situações reais



<http://www.tylervigen.com/spurious-correlations>

Frequentemente há *variáveis de confusão*: variável A afeta tanto B quanto C, então B e C são correlacionados, mas não têm relação causal entre si.

Sistemas Dinâmicos e Causalidade

Sistemas dinâmicos

Sistemas que evoluem no tempo, de acordo com alguma regra ou equação

Equações de Lorenz:

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\rho - z) - y$$

$$\frac{dz}{dt} = xy - \beta z$$

Descrevem muito simplificada a convecção de ar na atmosfera.

Equação logística (em tempo discreto):

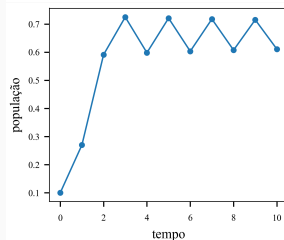
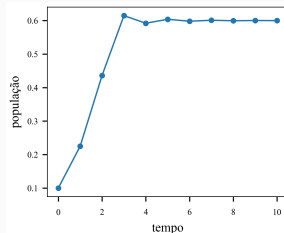
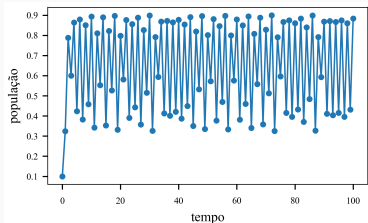
$$x_{t+1} = rx_t(1 - x_t)$$

Descreve uma população (por exemplo, de insetos) que se reproduz em gerações discretas e sob forte competição entre os indivíduos.

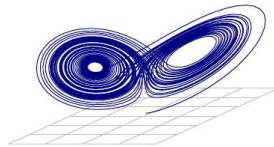
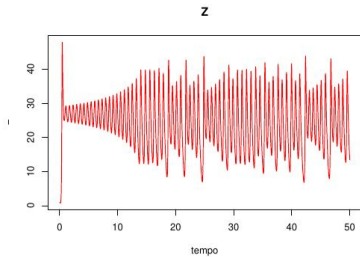
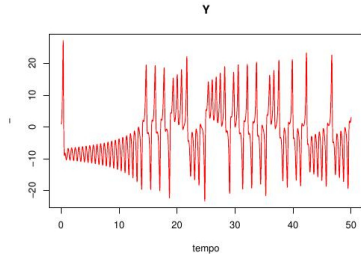
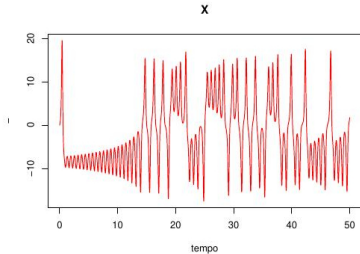
Atratores

Análise de sistemas dinâmicos: o que acontece no longo prazo?

- Equilíbrio (ponto fixo)
- Soluções periódicas
- Atratores estranhos (?)



O atrator de Lorenz – série temporal e espaço de fase



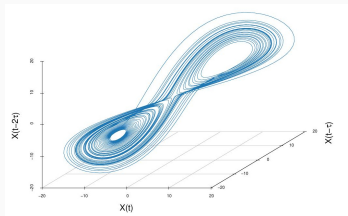
Atratores estranhos estão ligados ao caos.

Caos: sensibilidade às condições iniciais – incapacidade de previsão a longo prazo

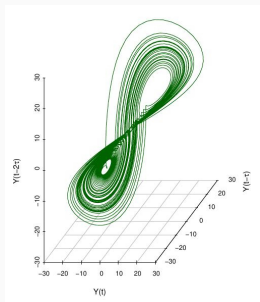
Difícil de comprovar definitivamente usando dados (caótico vs. ruidoso), mas considerado de grande importância em muitos sistemas

Reconstrução de atratores

Teorema de Takens: é possível reconstruir o atrator inteiro com apenas uma variável!



reconstrução usando a variável x . Retirado de [2].



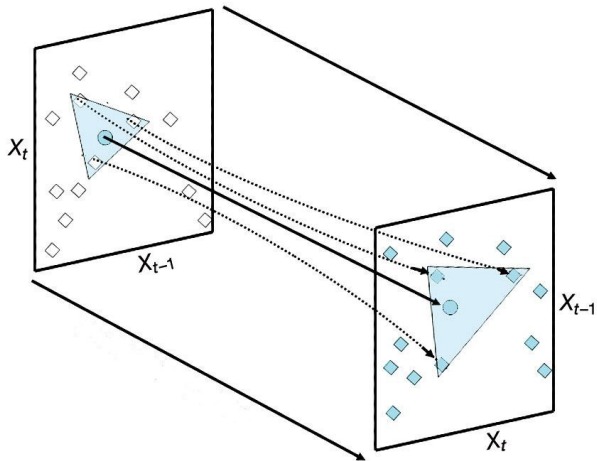
reconstrução usando a variável y

O atrator reconstruído não é idêntico, mas tem a **mesma forma** do original.

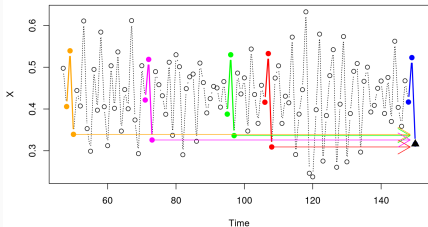
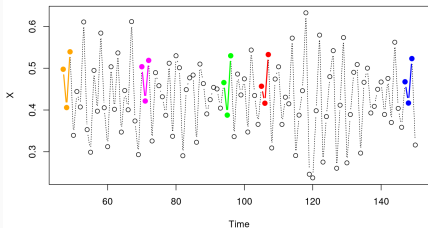
Cada ponto de um corresponde a um ponto do outro – em outras palavras, eles são **topologicamente** iguais

Projeção por meio do atrator reconstruído

Observando pontos próximos no atrator (no espaço de fase, não no tempo), podemos prever o ponto seguinte



Projeção por meio do atrator reconstruído



Observamos pontos do atrator próximos, e vemos aonde eles levam.

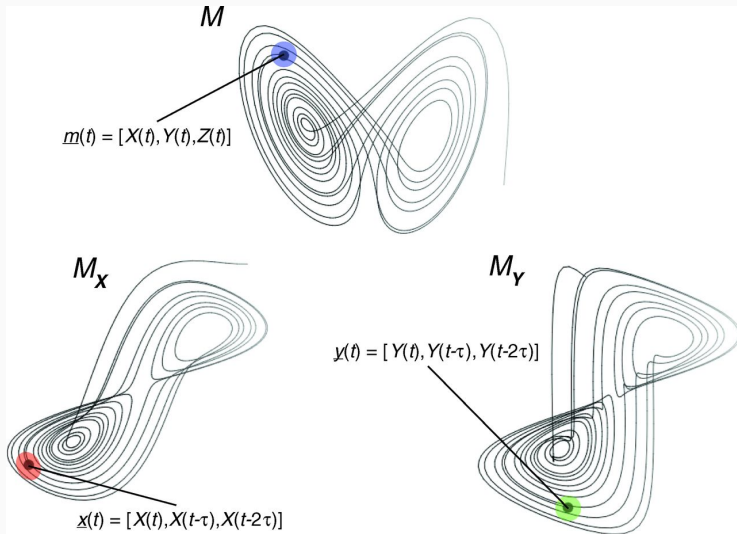
Isso nos dá um poder (limitado) de previsão

De atratores à causalidade

Duas variáveis têm uma relação causal se elas fazem parte do mesmo sistema dinâmico, e uma afeta a outra

Se duas variáveis pertencem ao mesmo sistema dinâmico, elas mapeiam o mesmo atrator!

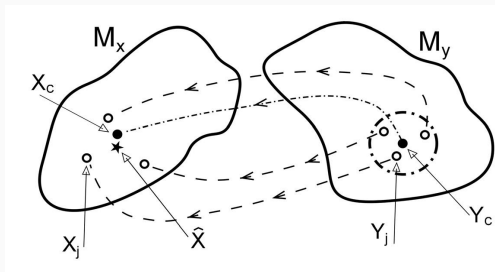
Atratores equivalentes



Retirado de [1].

Mapeamento cruzado: projeção de um atrator sobre o outro

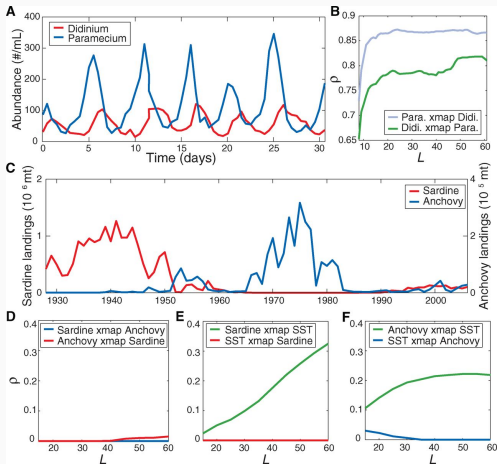
Usamos o atrator reconstruído a partir da variável Y para projetar o valor de X num certo instante – \hat{X}



Retirado de [1].

Repetimos esse procedimento para todos os pontos da série temporal. Se a projeção for boa, haverá alta correlação entre os valores projetados e os observados – ρ – “habilidade de mapeamento”

Mapeamento cruzado convergente (CCM)



Retirado de [1].

Se há causalidade, então a habilidade de mapeamento deve

- ser maior do que a correlação simples
- aumentar à medida que usamos mais dados para reconstruir o atrator

O método de mapeamento cruzado convergente é uma nova ferramenta para investigar causalidade em séries temporais

Ele tem encontrado aplicações em diversas áreas: ecologia, epidemiologia, genética, e mais – mas ainda tem potencial pra muito mais

Obrigado pela atenção!

Perguntas?

`http://professor.ufabc.edu.br/~renato.coutinho
renato.coutinho@ufabc.edu.br`

Recursos

Alguns pontos de partida para leituras e recursos importantes, nos quais baseei a apresentação.

- <https://mathbio.github.io/edmTutorials/>
Introdução à Modelagem Dinâmica Empírica na Escola de Verão do ICTP-SAIFR em janeiro/2018 (<http://www.ictp-saifr.org/school-on-physics-applications-in-biology-2/>), com um enfoque instrumental, mostrando com dados as ideias e a aplicação do método em casos simples. A seção 4, “Readings and resources”, contém as referências básicas da literatura no assunto.
- <https://repositorio.unesp.br/handle/11449/153844>
Dissertação de mestrado de Rafael Lopes Paixão da Silva, defendida no IFT em abril/2018 – “Aplicação de teoria de sistema dinâmicos para inferência de causalidade entre séries temporais sintéticas e biológicas”. Uma boa introdução ao assunto em português, com aplicação a um sistema ecológico de borboletas observadas na USP (campus Butantã).

- [1] Sugihara, G et al. **Detecting Causality in Complex Ecosystems** *Nature*, **338**, pp. 496–500 (2012)
- [2] da Silva, RLP **Aplicação de teoria de sistema dinâmicos para inferência de causalidade entre séries temporais sintéticas e biológicas**, dissertação de mestrado *IFT-Unesp* (2018)
- [3] Cabella, B et al, **Introduction to rEDM: Empirical Dynamics Modeling in R** workshop do ICT-SAIFR (2018)
- [4] Liu, Hui, et al. **Nonlinear dynamic features and co-predictability of the Georges Bank fish community** *Marine Ecology Progress Series* **464** pp. 195–207 (2012)