Data Science PhD

Statistical Causality

Asking "What if?" to data

Six lectures held by Marco Malvaldi

Introduction
May 4th 2021, 09:00 - 11:00

A gentle introduction: forecasting the future, explaining the past
What causes what? A (necessary) philosophical digression
Why you should care about causality: Simpson's paradox
Elements of probability theory and Bayes' theorem.
Difference between p(A|B) and p(A|do(B)): spurious correlation - or, selling ice cream do not cause woodfires.
How to define causality: B causes A if p(A|do B) > p(A|B). Is the altitude that causes
temperature, or is the cold that causes temperature?
Directed acyclic Graphs - for lazy people, DAG. A DAG is not a Bayesian net.
How to build a DAG. Conditional independence, or: is my DAG correct?

Dealing with DAG and reality
May 7th 2021, 09:00-11:00

A system of two variables: How to establish if A causes B, or B causes A, or none of them.
How to establish causality in a system of n variables, all known and measurable. 2.0 Virtual
intervention on a DAG: the do-operator.
Simpson paradox again, this time solved.
A drug good for men, good for women, harmful for humans.
How can you test a goalkeeper?

What if I can't measure a variable?
May 11th 2021, 09:00-11:00

Building a DAG with hidden variables:
Surgery on a DAG: blocking a pipe is the best way to test where is the leakage...
Constraints in hidden variables DAG
How to establish if X causes Y in a system of n variables – some of them unknown.
First attempt: the Back-door criterion.
Second attempt: the Front-door criterion.

Let's generalize
May 14th 2021, 09:00-11:00

General methods for operating on DAGs
A general criterion: if no child of X can be reached with a bidirectional trajectory, we can state if X causes, Y, that is, we can always express p(Y| do(X)) as a linear combination of conditional probabilities, and the procedure can be cast into an algorithm.

Causality in time series
May 18th 2021, 09:00-11:00

Granger causality
Transfer entropy
Where these two concepts fail (but they are nevertheless mandatory to learn)

Hands on
May 21st 2021, 09:00-11:00

Asking questions to the data: how to use the do-operator to build hypothetical worlds.
Direct effects of X on Y. Is the Covid fatality higher in China or in Italy?
Indirect effects of X on Y. Having made the political choices of Sweden, would fatality rate be different in Italy?
Natural effects vs. Induced effects. How would Belotti play in Milan coached by Pioli?
Questions, hope as many as possible.

About Marco Malvaldi

Marco Malvaldi is a popular Italian author of novels, short stories and scientific essays for the general public on popularizing physics-chemistry, computing, mathematics and the scientific aspects of human communication.

Malvaldi holds a PhD in chemistry from the University of Pisa (2004) and, beyond a novelist and a teacher, is also an active researcher in physical chemistry and in sport data analytics.

Malvaldi is currently a research collaborator of the department of Computer Science, University of Pisa,

This short course is organized for Ph.D. students in Data Science and other programs of the organizing institutions.
Registration is mandatory.
Other interested people can register, but their admission is subject to approval (based on the total number of participants).
Registration form: https://datasciencephd.eu/events/statistical-causality/registration
More details: https://datasciencephd.eu/events/statistical-causality

Scuola Normale Superiore · Università di Pisa · IMT School for Advanced Studies Lucca · Sant'Anna School of Advanced Studies – Pisa · Consiglio Nazionale delle Ricerche · SoBigData

Images by Anna Nekrashevich: https://www.pexels.com/it-it/foto/penna-business-occhiali-ricerca-6801648 and Nicola Ughi: https://it.wikipedia.org/wiki/Marco_Malvaldi#/media/File:MarcoMalvaldi_(cropped).jpg