

MACHT SCHOKOLADE SCHLAU?

Korrelation bedeutet nicht Kausalität, heißt es. Trotzdem lassen sich mit Hilfe mathematischer Methoden kausale Strukturen nachweisen.

TEXT JONAS PETERS

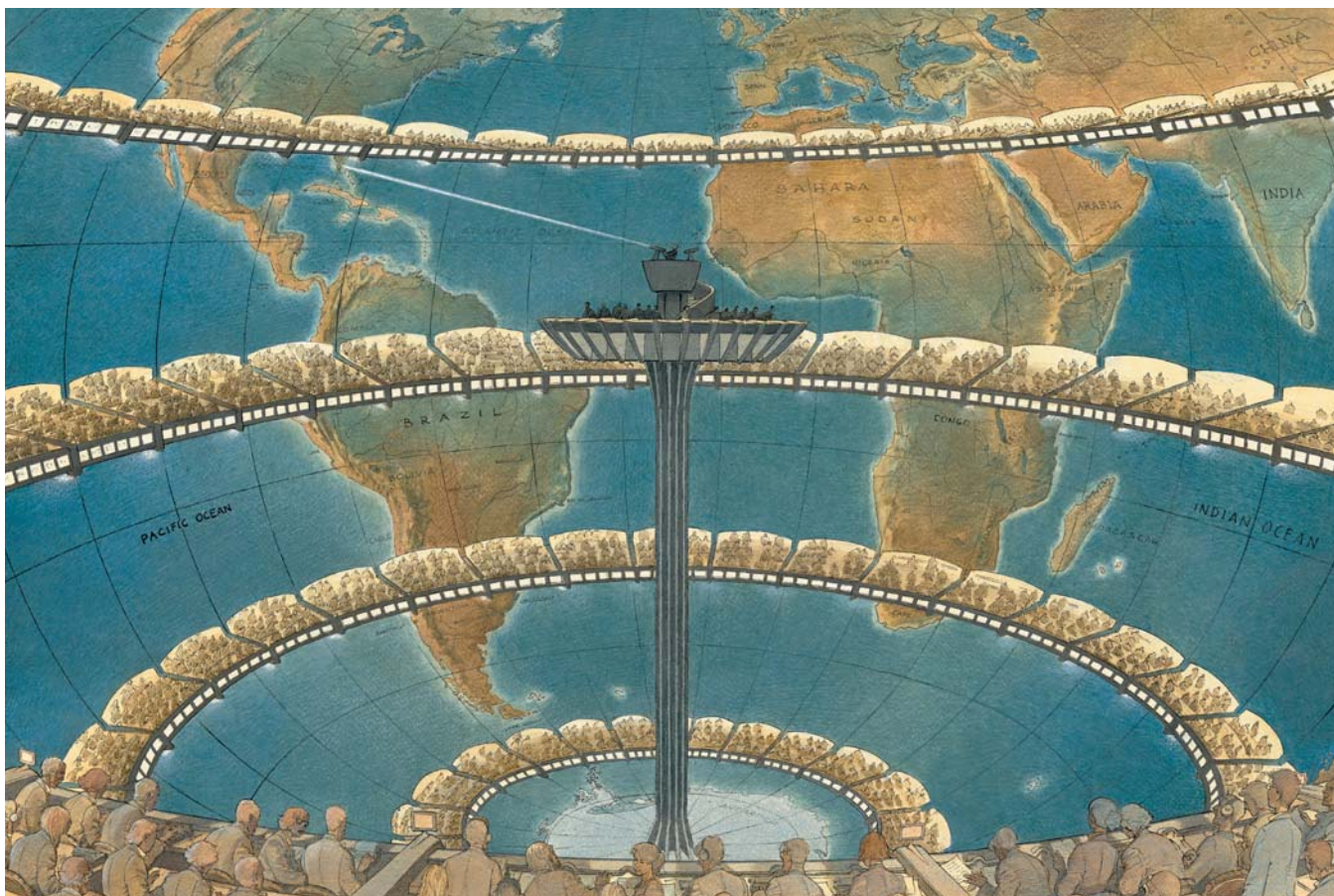
Je mehr Schokolade in einem Land gegessen wird, desto größer ist die Anzahl der Nobelpreisträgerinnen und -träger. Doch bedeutet dies, dass der Verzehr von Schokolade die Chance auf einen Nobelpreis erhöht? Um diese Frage zu beantworten, müsste man in einigen zufällig ausgewählten Ländern Schokolade verteilen und später untersuchen, ob sich die Anzahl der Nobelpreise für dieses Land erhöht hat. Auch wenn sich manch ein Studierender über die Einführung einer Zwei-Tafeln-Schokolade-pro-Tag-Politik freuen würde, ist solch ein Experiment kaum durchzusetzen.

In anderen Situationen scheitern derartige Interventionen aus ethischen, physikalischen oder finanziellen Gründen. Viele Forschende glauben, dass man ohne sorgfältig durchgeführte, randomisierte Experimente nichts über kausale Zusammenhänge herausfinden kann. Denn aus einem statistischen Zusammenhang lässt sich nicht so einfach auf einen kausalen Zusammenhang schließen – daran erinnert der Ausdruck „correlation does not imply causation“, Korrelation bedeutet nicht Kausalität. Aktuelle Forschungsprojekte zur kausalen Inferenz versuchen, diese Auffassung zu widerlegen. Es werden Methoden entwickelt, um kausale Beziehungen zwischen gleichzeitig beobachteten Größen zu erkennen, selbst ohne aktiv in das System einzugreifen. Diese Methoden basieren zum Beispiel auf der Idee, dass reale Zusammenhänge nicht beliebig komplex sind. Wären sie es, könnte Facebook nicht vorhersagen, welche Geschichten einen Nutzer interessieren. Google wüsste nicht, wonach wir suchen, und unser Smartphone würde unsere Spracheingabe nicht verstehen.

Dieses Prinzip der einfachen Zusammenhänge kann man auch in der kausalen Inferenz ausnutzen: Wenn man einfach erklä-

ren kann, wie Y aus X entsteht, dann muss in vielen Fällen die Konstruktion von X aus Y besonders komplex sein. Dies ist etwas überraschend, lässt sich aber mathematisch beweisen. Das einfachere Modell schlagen wir dann als kausales Modell vor. In der Praxis können wir uns also auf einfache Modelle beschränken und schauen, ob Daten besser durch ein Modell von X nach Y, von Y nach X oder durch keines der beiden erklärt werden. Es ist nicht überraschend, dass im Schokoladenbeispiel die Daten keinen einfachen kausalen Zusammenhang nahelegen: Weder hat der Schokoladenkonsum einen kausalen Einfluss auf die Anzahl der Nobelpreise, noch beginnt man, riesige Mengen Schokolade zu essen, sobald man den Nobelpreis bekommt. Vielmehr erwarten wir eine unbeobachtete Variable wie die wirtschaftliche Stärke eines Landes, die einen Einfluss auf beide Größen hat – also auf Schokoladenkonsum und die Anzahl der Nobelpreisträger.

Interessante wissenschaftliche Probleme beschäftigen sich meist mit mehr als zwei Variablen. So versucht man in biologischen Interaktionsnetzwerken, die Folgen einer Intervention vorherzusagen: Was passiert, wenn man bestimmte Gene löscht oder Proteine deaktiviert? Hier sind kausale Methoden von Interesse, weil es oft zu viele mögliche Interventionen gibt, um sie alle auszuprobieren zu können. Basierend auf dem oben beschriebenen Prinzip des einfachen Modells möchten wir herausfinden, welche kausale Struktur die Daten am besten erklärt. Allerdings ist es unmöglich, alle Strukturen zu testen, da deren Anzahl schlicht zu groß ist – selbst dann, wenn man gerichtete Zyklen, also Feedback, und versteckte Variablen ausschließt: Bei zwei Variablen X und Y gibt es drei Möglichkeiten (X verursacht Y, Y verursacht X, kein kausaler Zusammenhang); bei drei Variablen X, Y und Z gibt es 25 mögliche Kausalstrukturen und bei 13 Variablen bereits



François Schuiten illustriert Lewis Fry Richardsons Idee der „forecast factory“: Tausende Menschen lösen gleichzeitig und gemeinsam Differentialgleichungen, um eine globale Wettervorhersage zu liefern.

18676600744432035186664816926721. Vor knapp hundert Jahren entwarf der Mathematiker, Physiker und Meteorologe Lewis Fry Richardson eine der ersten Architekturen zum sogenannten Parallelen Rechnen, mit der er die numerische Wettervorhersage verbessern wollte (die damals noch nicht rechnerbasiert stattfand). Er stellte sich eine riesige Halle vor, in der tausende menschliche „Computer“ gleichzeitig Differentialgleichungen zu Druck oder Temperatur für einen ihnen zugewiesenen kleinen Teil der Erde lösen, was François Schuiten in einem Bild illustrierte.

Mit dieser Idee war Richardson seiner Zeit weit voraus: Tatsächlich wurden für viele Jahre die größten Rechner der Welt für die Wettervorhersage eingesetzt. Seine Idee hilft uns auch dabei, das beste (und damit kausale) Modell zu finden. Eine oft angewandte Strategie besteht darin, mit einer beliebigen kausalen Struktur zu beginnen und mit jedem einzelnen Schritt zu überprüfen, ob man die erhobenen Daten noch besser erklären kann – etwa wenn die Struktur leicht verändert wird, indem man beispielsweise die Kausalbeziehungen zwischen zwei Variablen umdreht

oder weglässt. Und all diese möglichen Veränderungen werden gleichzeitig von den vielen Prozessoren eines Hochleistungsrechners überprüft. So lassen sich selbst Probleme nicht nur mit 13, sondern mit tausenden Variablen untersuchen. Die kommenden Jahre werden zeigen, inwieweit solche kausale Methoden dabei helfen können, reale Systeme besser zu verstehen. Richardson sah den Nutzen des Parallelen Rechnens in der Berechnung großer, deterministischer Systeme. Heutzutage ist Parallelisierung etwa aus der Datenverarbeitung (maschinelles Lernen, data science) nicht mehr wegzudenken. Firmen wie Google und Facebook könnten ihre riesigen Datenmengen ohne Parallelisierung nicht verarbeiten. Auch das Auffinden kausaler Strukturen wäre ohne Paralleles Rechnen unmöglich – es wird in diesen Gebieten wohl für viele Jahre eine große Rolle spielen. Dass eines Tages die Aufgabe der „Computer“ nicht von Menschen, sondern von winzigen elektronischen Prozessoren durchgeführt wird, damit hatte Richardson sicher nicht gerechnet. ❁

Jonas Peters, Mitglied der Jungen Akademie seit 2016, arbeitet am Department of Mathematical Sciences an der University of Copenhagen.