Statistik (/Stochastik) und Kausalität

Prof. Dr. Hans-Jörg Starkloff

Technische Universität Bergakademie Freiberg
Fakultät für Mathematik und Informatik
Institut für Stochastik

24.04.2024

letzte Änderung: 25. April 2024

Bemerkungen zur Kausalität in der Stochastik

- Ausgangspunkt/Anlass: Artikel von Koutsoyiannis et al "Revisiting causality using stochastics" (Proceedings of the Royal Society A, Mathematical, Physical and Engineering Sciences, 2022)
- dazu einige Bemerkungen zur genutzten Mathematik/Stochastik
- ein kritischer Kommentar auch: Leif Åsbrink "Revisiting causality using stochastics on atmospheric temperature and CO_2 concentration" (Proceedings of the Royal Society A, Mathematical, Physical and Engineering Sciences, 2023)

- etwas zum Simpson-Paradoxon
- etwas zu anderen stochastischen Kausalitätsmodellierungen

1. Stochastische Kausalität nach Koutsoyiannis et al

- **meine Einordnung:** Versuch mit statistischen Mitteln die etablierte wissenschaftliche Meinung, dass der menschengemachte steigende CO_2-Gehalt in der Atmosphäre Ursache für steigende mittlere Temperaturen der Atmosphäre ist, zu widerlegen, und zu zeigen, dass die umgekehrte Kausalrichtung richtig ist
- dazu Argumentation mit dem Begriff der Kausalität, Nutzung einer speziellen Definition, insbesondere für eine notwendige Bedingung für Kausalität, kein Anspruch auf Nachweis der Kausalität durch hinreichende Bedingungen dafür
- bei Nutzung von mathematischen Begriffen oft nicht sehr genau
- ein Beispiel dazu: (Abschnitt 3.2 Basic concepts and definitions) "... the celebrated Wold decomposition, proving that any stochastic process can be decomposed into a regular process (i.e., a process linearly equivalent to a white noise process) and a predictable process (i.e., a process that can be expressed deterministically in terms of its past values)."

Klassisches kausales System in Koutsoyiannis et al

- Definition ausgehend von Papoulis, Probability, Random Variables and Stochastic Processes, 3rd ed., 1991
- **Geg.:** $X(t)$, $Y(t)$ stochastische Prozesse in stetiger Zeit; diese bilden ein (klassisches) kausales System mit $X(t)$ als Ursache ("cause") und $Y(t)$ als Wirkung ("effect"), falls für sie gilt
$$Y(t) = \int_0^\infty g(h) X(t-h)\,\mathrm{d}h\,, \tag{kks}$$
die deterministische Funktion $g(h)$ wird Impulsantwortfunktion ("impulse response function", IRF) genannt, h ist die Zeitverschiebung ("lag")
- **hier eigentlich:** ein skalares, deterministisches, lineares, zeitinvariantes ("LTI") kausales System mit zusätzlichen Regularitätseigenschaften, im allgemeinen Begriffsgebrauch ist nicht das Paar $(X(t), Y(t))$ das System, sondern die Abbildung $X \mapsto Y = \mathbf{T}[X]$

Kausale Systeme in der Systemtheorie, ...

- viele verschiedene Systembegriffe, ganz oft Transformation eines "Eingangs (Eingangssignals)" in einen "Ausgang" (ein "Ausgangssignal")
- kausales System: (siehe z.B. UNBEHAUEN, Systemtheorie Eine Darstellung für Ingenieure, 1980, S.20)
  Der Verlauf des Ausgangssignals $y(t)$ bis zu einem beliebigen zulässigen Zeitpunkt t_1 hängt nur vom Verlauf des entsprechenden Eingangssignals $x(t)$ bis zu diesem Zeitpunkt t_1 ab, d.h. für zwei beliebige Eingangssignale $x^{(1)}(t), x^{(2)}(t)$ mit der Eigenschaft

  $$x^{(1)}(t) = x^{(2)}(t) \quad \text{für } t \leq t_1$$

  gilt

  $$y^{(1)}(t) := \mathbf{T}[x^{(1)}](t) = \mathbf{T}[x^{(2)}](t) =: y^{(2)}(t) \quad \text{für } t \leq t_1 \, .$$
- Asymmetrie in der Zeit, keine Wirkung vor der Ursache

Begriff IRF und ein Hintergrund

- **formal** mit "$\delta-$Funktion"

$$g(t) = \int_0^\infty g(h)\delta(t-h)\,\mathrm{d}h\,,$$

- bestimmte Lösungen von linearen gewöhnlichen Differentialgleichungen mit deterministischen (konstanten) Koeffizienten und einem zufälligen Störterm können in der Form (kks) geschrieben werden

Beispiel Differentialgleichung 1. Ordnung

- inhomogene lineare Differentialgleichung 1. Ordnung mit konstanten Koeffizienten ($\lambda > 0$)

$$Y(t) := \int_{-\infty}^{t} \mathrm{e}^{-\lambda(t-s)h} X(s)\,\mathrm{d}s = \int_{0}^{\infty} \mathrm{e}^{-\lambda h} X(t-h)\,\mathrm{d}h$$

  ist unter geeigneten Voraussetzungen an $(X(t))_{t\in\mathbb{R}}$ Lsg. von

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} + \lambda Y(t) = X(t)\,.$$

- **Bem.:** Hier gilt $g(h) > 0$ für alle $h \geq 0$.

Beispiel Differentialgleichung 2. Ordnung

- inhomogene lineare Differentialgleichung 2. Ordnung mit konstanten Koeffizienten ($\theta_0 > 0\,, 0 < \delta < 1\,, \theta_d := \sqrt{1-\delta^2}\theta_0$)

$$Y(t) := \int_0^\infty \frac{1}{\theta_d} \mathrm{e}^{-\delta\theta_0 h} \sin(\theta_d h) X(t-h)\,\mathrm{d}h$$

  ist unter geeigneten Voraussetzungen Lsg. von

$$\frac{\mathrm{d}^2 Y(t)}{\mathrm{d}t^2} + 2\delta\theta_0 \frac{\mathrm{d}Y(t)}{\mathrm{d}t} + \theta_0^2 Y(t) = X(t)\,.$$

- **Bem.:** Hier nimmt die Impulsantwortfunktion g Werte mit unterschiedlichen Vorzeichen an.

Ausgangspunkt für Kausalitätsformen in Koutsoyiannis et al

- $X(t), Y(t)$ zwei stationäre stochastische Prozesse, dann existiert nach Ansicht der Autoren (mit Verweis auf Formel (14-12) in Papoulis) immer eine Beziehung

$$Y(t) = \int_{-\infty}^{\infty} g(h) X(t-h)\,\mathrm{d}h + V(t) \tag{7}$$

  mit einem stochastischen Prozess $V(t)$ unkorreliert zu $X(t)$

- zur Bewertung genutzt wird das erklärte Varianzverhältnis ("explained variance ratio")

$$e := 1 - \frac{\mathbb{V}\mathbf{ar}\,V}{\mathbb{V}\mathbf{ar}\,Y} \tag{evr}$$

- Eindeutigkeit von g ?

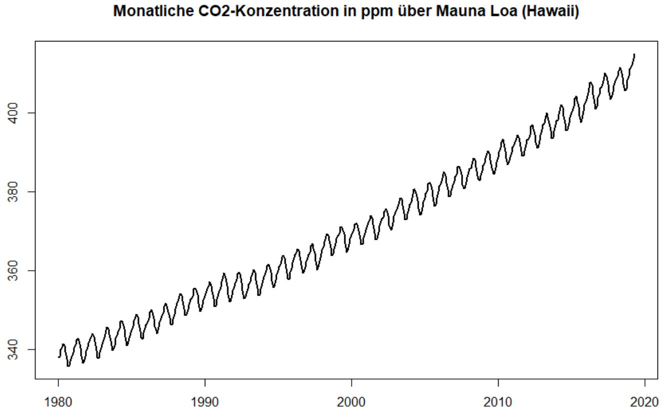
Genutzte Kausalitätsbegriffe in den Artikeln

System $(X(t), Y(t))$ im Zusammenhang (7) ist

- potentiell kausal ("potentially causal"), falls $g(h) = 0$ für alle $h < 0$ und die erklärte Varianz ist nicht vernachlässigbar
- potentiell antikausal ("potentially anticausal"), falls $g(h) = 0$ für alle $h > 0$ und die erklärte Varianz ist nicht vernachlässigbar (dann ist $(Y(t), X(t))$ potentiell kausal)
- potentiell HOE-kausal ("potentially hen-or-egg (HOE) causal"), falls $g(h) \neq 0$ für einige $h < 0$ und einige $h < 0$ und die erklärte Varianz ist nicht vernachlässigbar
- nichtkausal ("noncausal"), falls die erklärte Varianz vernachlässigbar ist

Genutzte Daten

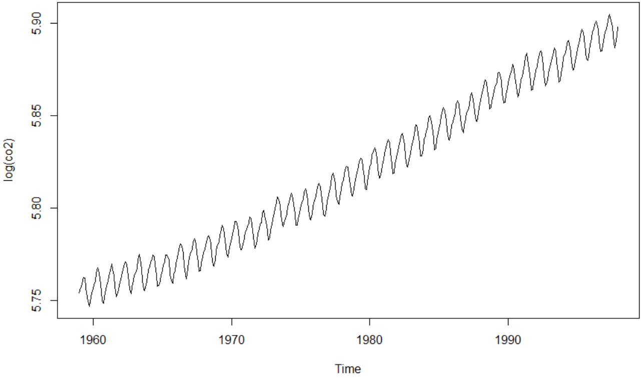
Monatliche CO_2-Konzentration in ppm über Mauna Loa (Hawaii) hier: 01/1980–05/2019, im Artikel 1979 bis 05/2021



(735 Werte) Datenquelle: Datensatz "MAUNALOA" im R-Paket "tsapp"

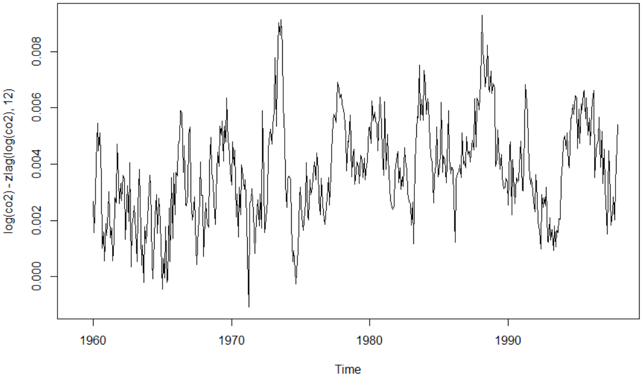
Fortsetzung CO_2-Daten, logarithmierte Werte

die Werte werden logarithmiert



Fortsetzung CO_2-Daten, 12-Monatsdifferenzen logarithmierte Werte

Differenzen um 12 Monate der logarithmierten Werte



Temperaturdaten

- Beobachtungen der globalen Temperatur durch Satelliten, entwickelt an der Universität von Alabama in Huntsville (UAH)
- Datensatz beginnt 1979, monatliche Daten

Inverses Problem

- ausgehend von (7) Berechnung (Schätzung) von $g(h)$, so dass Varianz von V minimal wird

$$Y(t) = \int_{-\infty}^{\infty} g(h)X(t-h)\,\mathrm{d}h + V(t)$$

- inkorrektes (schlecht gestelltes) Identifikationsproblem im Zusammenhang mit einer Integralgleichung; dazu
  - nur endlicher Träger von g wird betrachtet, nicht ganz $\mathbb{R}$;
  - Nichtnegativität von g (vgl. mit obigem Bsp. Differentialgleichung 2. Ordnung !)
  - Glattheitsbedingung: $E \leq E_0$ mit gewähltem E_0 und

$$E := \int_{-\infty}^{\infty} (g''(h))^2\,\mathrm{d}h$$

  - Minimum der Varianz von V
  - (Zeit-)Diskretisierung $\Rightarrow$ eigentlich wird mit Folgen (Vektoren) und Summen gearbeitet

Ergebnisse für Temperatur und CO_2

Ergebnisse der Berechnung der Impulsantwortfunktionen

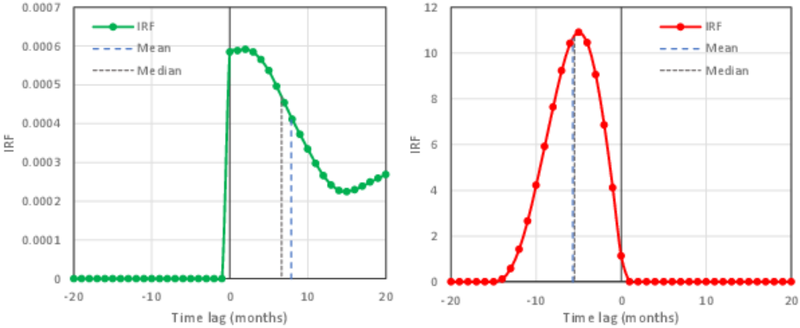


Figure 14 IRFs for of temperature – CO_2 concentration based on the modern time series. (left) Case study #23 ($\Delta T \rightarrow \Delta\ln[CO_2]$); (right) case study #24 ($\Delta\ln[CO_2] \rightarrow \Delta T$).

Aus: R.P. Feynman, Sie belieben wohl zu scherzen, Abschnitt: Cargo-Kult-Wissenschaft

Als ich in Cornell war, habe ich mich oft mit den Leuten aus dem Fachbereich Psychologie unterhalten. Eine der Studentinnen erzählte mir, sie wolle ein Experiment machen, das etwa folgendermaßen ablaufen sollte: Es war herausgefunden worden, daß Ratten unter bestimmten Bedingungen, X, etwas Bestimmtes, A, tun. Sie wollte nun wissen, ob die Ratten immer noch A tun, wenn sie die Bedingungen nach Y hin veränderte. Sie schlug also vor, das Experiment unter den Bedingungen Y durchzuführen und zu sehen, ob die Ratten auch dann noch A tun.

Ich erklärte ihr, sie müsse zunächst im Labor das Experiment wiederholen, das von anderen angestellt worden war - das heißt unter den Bedingungen X überprüfen, ob sie ebenfalls das Resultat A bekäme, und dann die Bedingungen nach Y hin verändern und sehen, ob A sich änderte. Dann wisse sie, ob das, was sie unter Kontrolle zu haben glaube, der wirkliche Unterschied sei.

Sie freute sich sehr über diese neue Idee und ging zu ihrem Professor. Und dessen Antwort war nein, sie könne das nicht machen, denn das Experiment sei ja bereits gemacht worden und sie vergeude damit nur Zeit.

Das war so um 1947, und es scheint damals die Regel gewesen zu sein, psychologische Experimente nicht zu wiederholen, sondern nur die Bedingungen zu ändern und zu sehen, was dann geschieht.

. . .

Freilich sind nicht alle Experimente in der Psychologie von dieser Art. Es sind zum Beispiel viele Experimente durchgeführt worden, bei denen Ratten durch alle möglichen Labyrinthe und dergleichen geschickt wurden - und zwar ohne daß sich dabei sonderlich klare Resultate ergeben hätten. Doch 1937 hat ein Mann namens Young ein sehr interessantes Experiment angestellt. Er arbeitete mit einem langen Gang, bei dem sich auf der einen Seite die Türen befanden, durch welche die Ratten hereinkamen, und auf der anderen Seite jene, hinter denen sich das Futter befand. Er wollte wissen, ob er die Ratten darauf trainieren konnte, jeweils durch die dritte Tür zu laufen, und zwar ganz gleich, wo er sie losließ.

Nein. Die Ratten liefen sofort zu der Tür, hinter der beim vorigen Mal das Futter gewesen war. Da der Gang sehr schön gleichmäßig gebaut war, stellte sich die Frage, woher die Ratten wußten, daß es die gleiche Tür wie vorher war. Offenkundig war an der Tür etwas anders als an den anderen Türen.

Also strich er die Türen sehr sorgfältig an und richtete es so ein, daß die Oberflächenbeschaffenheit der Türen genau gleich war. Trotzdem fanden die Ratten es heraus. Dann dachte er, vielleicht könnten die Ratten das Futter riechen, und er verwendete Chemikalien, um den Geruch nach jedem Durchlauf zu verändern. Aber die Ratten erkannten die Tür immer noch. Dann fiel ihm ein, daß die Ratten, wie jede vernünftige Person, in der Lage sein könnten, die Tür durch das Licht und die Einrichtung im Labor zu finden. Er deckte den Gang ab, aber die Ratten erkannten die Tür nach wie vor. Schließlich fand er heraus, daß sie die Tür an dem Schall wiedererkannten, der entstand, wenn sie über den Boden liefen. Und das konnte er nur ausräumen, indem er seinen Gang in Sand bettete.

Er beseitigte also nacheinander alle möglichen Anhaltspunkte und brachte es schließlich fertig, die Ratten so zu täuschen, daß sie lernen mußten, durch die dritte Tür zu laufen. Wenn er nur eine der Bedingungen nicht aufrechterhielt, wußten die Ratten Bescheid. Vom wissenschaftlichen Standpunkt aus ist dies ein erstklassiges und absolut mustergültiges Experiment. Es ist das Experiment, das Experimente sinnvoll macht, bei denen man Ratten laufen läßt, denn es deckt die Anhaltspunkte auf, die die Ratten wirklich verwenden - und nicht die, von denen man das glaubt.

Und es ist das Experiment, aus dem exakt hervorgeht, welche Voraussetzungen man schaffen muß, um genau zu sein und bei einem Experiment mit laufenden Ratten alles unter Kontrolle zu behalten.

Ich habe mir die spätere Geschichte dieser Forschungen angesehen. Das nächste Experiment und das darauffolgende nahm nicht einmal auf Mr. Young Bezug. Sie verwendeten kein einziges von seinen Kriterien, die dazu geführt hatten, den Gang in Sand zu betten oder sehr sorgfältig vorzugehen. Man fuhr einfach fort, die Ratten genau wie früher laufen zu lassen, und schenkte den großen Entdeckungen von Mr. Young keine Beachtung; seine Artikel werden nicht zitiert, weil er ja nichts über die Ratten herausgefunden hat. Tatsächlich hat er alles herausgefunden, was man tun muß, um etwas über Ratten herauszufinden. Doch solchen Experimenten keine Beachtung zu schenken, ist charakteristisch für die Cargo-Kult-Wissenschaft.

Aus: E. FROMM, Überfluss und Überdruss in unserer Gesellschaft, Abschnitt: 1. Der passive Mensch

Ich möchte ein Beispiel geben für den Unterschied zwischen Aktivität und Passivität. Es hat eine große Rolle in der amerikanischen Industrie-Psychologie gespielt. Professor Elton Mayo hat folgendes Experiment angestellt, als er von der Western Electric Company beauftragt wurde, zu prüfen, wie man die Produktivität von Arbeiterinnen, ungelernten übrigens, in den Hawthorne-Werken in Chicago steigern könne. Man war damals der Meinung, vielleicht arbeiten sie besser, wenn man ihnen am Morgen zehn Minuten frei gibt und vielleicht zehn weitere Minuten als Kaffeepause et cetera. Diese ungelernten Arbeiterinnen mußten etwas tun, was sehr monoton ist, nämlich Spulen aufwickeln. Dazu gehört keine Kunst, keine Anstrengung, es ist das Passivste und Eintönigste, was man sich vorstellen kann.

Da hat Elton Mayo ihnen sein Experiment erklärt und zunächst einmal die Kaffeepause am Nachmittag eingeschaltet. Sofort stellte sich heraus, daß die Produktivität stieg.

Nun, das war ja wunderbar, weil sich zeigte, wie gut diese Methode wirkt. Dann hat er zusätzlich die Pause am Vormittag installiert, und wieder stieg die Produktivität. Weitere Vergünstigungen hatten weitere Produktivität zur Folge, so daß die Rechnung voll aufging.

Ein gewöhnlicher Professor hätte an dieser Stelle das Experiment beendet und den Direktoren der Western Electric Company empfohlen, durch einen Zeitverlust von zwanzig Minuten eine höhere Produktivität zu erzielen. Anders Elton Mayo, der ein einfallsreicher Mann war. Er hat sich nämlich gefragt, was geschehen würde, wenn er die Vorteile wieder streichen würde. So hat er zunächst die Kaffeepause rückgängig gemacht – und die Produktionssteigerung ging weiter. Und so fort. Womöglich hätten an dieser Stelle einige Professoren achselzuckend festgestellt: Naja, man sieht eben, das Experiment ist nicht aussagekräftig ... Aber in unserem Falle tauchte plötzlich der Gedanke auf: Womöglich haben die ungelernten Arbeiterinnen zum ersten Mal in ihrem Leben Interesse gewonnen an dem, was sie in der Fabrik taten.

Das Spulenaufrollen blieb langweilig und eintönig wie eh und je; aber man hatte sie in das Experiment eingeweiht, und so fühlten sie, daß sie in einem Zusammenhang wirkten, daß sie etwas beitrugen, das nicht nur für den Profit des ananymen Unternehmers, sondern für die ganze Belegschaft von Bedeutung war.

Mayo konnte nachweisen, daß es dieses unerwartete Interesse, dieses Dabeiseinkönnen war, was die Produktivität der Arbeit gesteigert hatte, und nicht etwa die Pausen am Vor- oder Nachmittag. Das war Anlaß und Anstoß zu einer neuen Denkweise: daß das Motiv für die Produktivität mehr im *Interesse* an der Arbeit selbst, als in Pausen, Gehaltserhöhungen und weiteren Bequemlichkeiten zu suchen ist. Ich komme später noch einmal darauf zurück. Ich wollte hier nur den entscheidenden Unterschied zwischen Aktivität und Passivität aufzeigen. Solange die Arbeiterinnen kein Interesse hatten, waren sie passiv. In dem Augenblick, als ihnen Anteil an dem Experiment gegeben wurde, erwachte in ihnen ein Gefühl für Mitarbeit, sie wurden aktiv und änderten ihre Haltung grundlegend.

2. Zum Simpson-Paradoxon

- **Geg.** Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$, $A, B, C \in \mathcal{F}$,
  $\mathbf{P}(B \cap C) \cdot \mathbf{P}(B^c \cap C) \cdot \mathbf{P}(B \cap C^c) \cdot \mathbf{P}(B^c \cap C^c) > 0$
- es kann sein, dass gilt

  $$\mathbf{P}(A|B) < \mathbf{P}(A|B^c)$$

  obwohl

  $$\mathbf{P}(A|B \cap C) > \mathbf{P}(A|B^c \cap C) \quad \text{und} \quad \mathbf{P}(A|B \cap C^c) > \mathbf{P}(A|B^c \cap C^c)$$

- möglich, da gewichtete Mittel mit unterschiedlichen Gewichten:
  $$\mathbf{P}(A|B) = \mathbf{P}(C|B)\, \mathbf{P}(A|B \cap C) + \mathbf{P}(C^c|B)\, \mathbf{P}(A|B \cap C^c)$$
  $$\mathbf{P}(A|B^c) = \mathbf{P}(C|B^c)\mathbf{P}(A|B^c \cap C) + \mathbf{P}(C^c|B^c)\mathbf{P}(A|B^c \cap C^c)$$
- nicht möglich, wenn B und C unabhängige Ereignisse sind
- kann auch in realen Datensätzen beobachtet werden

Beispiel 1.2.1 in Pearl, Glymour, Jewell, 2021

- Genesung von 700 Patienten mit einer bestimmten Krankheit, die ein spezielles Medikament nehmen konnten oder nicht, 350 wählten das Medikament, 350 nicht

	Medikament (B)	kein Medikament (B^c)
Mann (C)	81 von 87 genesen (A)	234 von 270 genesen (A)
Frau (C^c)	192 von 263 genesen (A)	55 von 80 genesen (A)
insgesamt	273 von 350 genesen (A)	289 von 350 genesen (A)

- $\mathbf{P}(A) \approx 0.80\,,\ \mathbf{P}(B) \approx 0.5\,,\ \mathbf{P}(C) \approx 0.51$
- $\mathbf{P}(A|B) \approx 0.78 < \mathbf{P}(A|B^c) \approx 0.83$
- $\mathbf{P}(A|B \cap C) \approx 0.93 > \mathbf{P}(A|B^c \cap C) \approx 0.87\,,$

  $\mathbf{P}(A|B \cap C^c) \approx 0.73 > \mathbf{P}(A|B^c \cap C^c) \approx 0.69$
- $\mathbf{P}(C|B) \approx 0.25\,, \quad \mathbf{P}(C^c|B) \approx 0.75$

  $\mathbf{P}(C|B^c) \approx 0.77\,, \quad \mathbf{P}(C^c|B^c) \approx 0.23$

Im Beispiel:

- bedingte Version des Satzes der totalen Wkt. für $\mathbf{P}(A|B)\,,\mathbf{P}(A|B^c)$:

$$0.78 \approx 0.25 \cdot 0.93 + 0.75 \cdot 0.73$$
$$0.83 \approx 0.77 \cdot 0.87 + 0.23 \cdot 0.69$$

- Struktur: (mit $\alpha, \beta, a_1, a_2, b_1, b_2 \in (0,1)$, $a_1 > b_1$, $a_2 > b_2$)

$$w_1 = \alpha a_1 + (1-\alpha) a_2$$
$$w_2 = \beta b_1 + (1-\beta) b_2$$

- notwendige Bedingung für $w_1 < w_2$:

$$\min\{a_1, a_2\} < \max\{b_1, b_2\}$$

  (Überlappung der (konvexen) Intervalle mit a_1, a_2 bzw. b_1, b_2 als Intervallgrenzen)

- hier sind allerdings a_1, a_2, b_1, b_2 nicht fest vorgegeben, wenn ein solches Ereignis C gesucht wird, ebenso α, β

Fragen, Überlegungen dazu im Buch, S.2-3

"The data seem to say that if we know the patient's gender – male or female – we can prescribe the drug, but if the gender is unknown we should not! Obviously, that conclusion is ridiculous. If the drug helps men and women, it must help anyone; our lack of knowledge of the patient's gender cannot make the drug harmful.

Given the results of this study, then, should a doctor prescribe the drug for a woman? A man? A patient of unknown gender? Or consider a policy maker who is evaluating the drug's overall effectiveness on the population. Should he/she use the recovery rate for the general population? Or should he/she use the recovery rates for the gendered subpopulations?

The answer is nowhere to be found in simple statistics. In order to decide whether the drug will harm or help a patient, we first have to understand the story behind the data – the causal mechanism that led to, or generated, the results we see.

Fortsetzung

For instance, suppose we knew an additional fact: Estrogen has a negative effect on recovery, so women are less likely to recover than men, regardless of the drug. In addition, as we can see from the data, women are significantly more likely to take the drug than men are. So, the reason the drug appears to be harmful overall is that, if we select a drug user at random, that person is more likely to be a woman and hence less likely to recover than a random person who does not take the drug. Put differently, being a woman is a common cause of both drug taking and failure to recover. Therefore, to assess the effectiveness, we need to compare subjects of the same gender, thereby ensuring that any difference in recovery rates between those who take the drug and those who do not is not ascribable to estrogen. This means we should consult the segregated data, which shows us unequivocally that the drug is helpful. This matches our intuition, which tells us that the segregated data is "more specific", hence more informative, than the unsegregated data."

Bemerkungen

- Im Beispiel Genesung von Frauen und Männern:
  - $\mathbf{P}(A|C^c) \approx 0.72\,,\ \mathbf{P}(A|B \cap C^c) \approx 0.73\,,\ \mathbf{P}(A|B^c \cap C^c) \approx 0.69$
  - $\mathbf{P}(A|C) \approx 0.88\,,\ \mathbf{P}(A|B \cap C) \approx 0.93\,,\ \mathbf{P}(A|B^c \cap C) \approx 0.87$
- englisch auch: eine der "statistical fallacies"; "Simpson's paradox", "Simpson's reversal", "Yule-Simpson effect", "amalgamation paradox", "reversal paradox"
- Zusammenhang mit Interaktionen in mehrdimensionalen Kontigenztafeln
- Zusammenhang mit "collapsibility of multidimensional contingency tables"
- Verweis auch auf Auftreten in der sequentiellen Statistik

3. Bemerkungen zu Kausalität und Stochastik/Statistik

- eine Möglichkeit aus COX, Causality: some statistical aspects, 1992
- C, E zufällige Ereignisse ("binary events"), B "a third variable or collection of variables"
- "C is a candidate cause of E" (d.h. eine mögliche Ursache), falls C und E positiv assoziiert sind, d.h. es gilt (Vor.: $0 < \mathbf{P}(C) < 1$)

$$\mathbf{P}(E|C) > \mathbf{P}(E|C^c) \qquad (*)$$

- die Ursache C kann als Scheinursache ("spurious cause") betrachtet werden, falls B die Assoziierung erklärt, d.h. es gilt

$$\mathbf{P}(E|C \cap B) = \mathbf{P}(E|C^c \cap B)$$

- ein Problem: eigentlich sollten [?] Ursache und Wirkung asymmetrisch in Betrachtungen eingehen, dies könnte z.B. durch eine Zuordnung der Ereignisse zu verschiedenen (geordneten) Zeitmomenten realisiert werden

Äquivalenz zur positiven Abhängigkeit

- **Def.** **Geg.:** Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ und $C, E \in \mathcal{F}$ mit $0 < \mathbf{P}(C) < 1$
  - E ist positiv von C abhängig, falls gilt $\mathbf{P}(E|C) > \mathbf{P}(E)$
  - E ist negativ von C abhängig, falls gilt $\mathbf{P}(E|C) < \mathbf{P}(E)$
  - E ist unabhängig von C, falls gilt $\mathbf{P}(E|C) = \mathbf{P}(E)$
- **Beh.** **Geg.:** Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ und $C, E \in \mathcal{F}$ mit $0 < \mathbf{P}(C) < 1 \quad \Rightarrow$
  - E ist positiv von C abhängig $\Leftrightarrow$ C ist eine mögliche Ursache von E im Sinn von Cox, d.h. es gilt $\mathbf{P}(E|C) > \mathbf{P}(E|C^c)$
  - E ist negativ von C abhängig $\Leftrightarrow$ $\mathbf{P}(E|C) < \mathbf{P}(E|C^c)$
  - E ist unabhängig von C $\Leftrightarrow$ $\mathbf{P}(E|C) = \mathbf{P}(E|C^c)$

Zur positiven und negativen Abhängigkeit von Ereignissen

Beh.

- **Geg.:** Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ und $C, E \in \mathcal{F}$ mit $0 < \mathbf{P}(C), \mathbf{P}(E) < 1$
- **Dann** sind alle folgenden 20 Beziehungen zueinander äquivalent:
  - $\mathbf{P}(E|C) > \mathbf{P}(E), \quad \mathbf{P}(C|E) > \mathbf{P}(C), \quad \mathbf{P}(C \cap E) > \mathbf{P}(C)\mathbf{P}(E)$
  - $\mathbf{P}(E|C^c) < \mathbf{P}(E), \quad \mathbf{P}(C^c|E) < \mathbf{P}(C^c), \quad \mathbf{P}(C^c \cap E) < \mathbf{P}(C^c)\mathbf{P}(E)$
  - $\mathbf{P}(E^c|C) < \mathbf{P}(E^c), \quad \mathbf{P}(C|E^c) < \mathbf{P}(C), \quad \mathbf{P}(C \cap E^c) < \mathbf{P}(C)\mathbf{P}(E^c)$
  - $\mathbf{P}(E^c|C^c) > \mathbf{P}(E^c), \ \mathbf{P}(C^c|E^c) > \mathbf{P}(C^c), \ \mathbf{P}(C^c \cap E^c) > \mathbf{P}(C^c)\mathbf{P}(E^c)$
  - $\mathbb{C}\mathbf{ov}[\mathbf{1}_C, \mathbf{1}_E] > 0, \quad \mathbb{C}\mathbf{orr}[\mathbf{1}_C, \mathbf{1}_E] > 0$
  - $\mathbb{C}\mathbf{ov}[\mathbf{1}_{C^c}, \mathbf{1}_E] < 0, \quad \mathbb{C}\mathbf{orr}[\mathbf{1}_{C^c}, \mathbf{1}_E] < 0$
  - $\mathbb{C}\mathbf{ov}[\mathbf{1}_C, \mathbf{1}_{E^c}] < 0, \quad \mathbb{C}\mathbf{orr}[\mathbf{1}_C, \mathbf{1}_{E^c}] < 0$
  - $\mathbb{C}\mathbf{ov}[\mathbf{1}_{C^c}, \mathbf{1}_{E^c}] > 0, \quad \mathbb{C}\mathbf{orr}[\mathbf{1}_{C^c}, \mathbf{1}_{E^c}] > 0$
- analoge Aussagen gelten für negativ abhängige und unabhängige Ereignisse C und E

Hinweis auf Arbeit von Chung

- Kai-Lai Chung, On mutually favorable events, The Annals of Mathematical Statistics, 1942, v. 13, no. 3
- Verallgemeinerung des Begriffs der positiven Abhängigkeit ("C is favorable to E", $\mathbf{P}(E|C) > \mathbf{P}(E)$) für zwei Ereignisse auf endliche Ereignissysteme
- analog zum Übergang von paarweiser Unabhängigkeit zur vollständigen Unabhängigkeit
- Ungleichungen für Wahrscheinlichkeiten für das gemeinsame Auftreten solcher Ereignisse

In Zeitreihenanalyse

- Begriffe von kausalen und invertierbaren ARMA$-(p,q)-$Prozessen oder -Modellen

In Pearl, Glymour, Jewell, 2021, S. 5

- For our purposes, the definition of causation is simple, if a little metaphorical: A variable X is a *cause* of a variable Y if Y in any way relies on X for its value. We will expand slightly upon this definition later, but for now, think of causation as a form of listening; X is a cause of Y if Y listens to X and decides its value in response to what it hears.

Eine andere Situation

- **Satz**
  Sei η eine reelle Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ mit stetiger Verteilungsfunktion.

  Dann existieren für jede Folge $(F_n)_{n\in\mathbb{N}}$ von Verteilungsfunktionen deterministische messbare Funktionen $g_n : \mathbb{R} \to \mathbb{R}, n \in \mathbb{N}$, so dass die reellen Zufallsgrößen $\xi_n = g_n(\eta)$ **stochastisch unabhängig** sind und jeweils die Verteilungsfunktionen $F_n, n \in \mathbb{N}$, besitzen.

- **Frage**: Kausalitätseigenschaften in diesem Fall ?