

## Korollar statt Gegenentwurf? Mayo und der Bayesianismus\*

Sebastian Lutz

### 1. Einleitung

Grundlage der bayesianistischen Schlussweise ist die Annahme, dass Hypothesen Wahrscheinlichkeiten zugeordnet werden können und diese mit Hilfe des Bayes'schen Satzes (1) neuer Evidenz angepasst werden. Als Gegenentwurf präsentiert Deborah G. Mayo in *Error and the Growth of Experimental Knowledge*<sup>1</sup> eine Schlussweise basierend auf Neyman und Pearsons Fehlerstatistik. Ich will in diesem Beitrag zeigen, dass Mayos grundlegende Annahme über die Anwendung von Fehlerstatistik, das Fehlerargument (*argument from error*), ein Korollar der bayesianistischen Schlussweise ist. Des Weiteren argumentiere ich dafür, dass Mayos Ansatz nicht ampliativ ist.

In Abschnitt 2 stelle ich den Bayes'schen Satz und seine Probleme bei der Anwendung auf Hypothesen vor, Abschnitt 3 enthält die Formalisierung von Mayos Ansatz und Abschnitt 4 dessen Herleitung aus der bayesianistischen Schlussweise unter Berücksichtigung der Probleme. In Abschnitt 5 gehe ich auf Mayos Bewertung größerer Theorien ein.

### 2. Die bayesianistische Schlussweise

Der Bayes'sche Satz ergibt sich mathematisch unkontrovers direkt aus der (definierten oder postulierten) bedingten Wahrscheinlichkeit  $P(\cdot|\cdot)$ :

$$(1a) \quad P(H|e) = \frac{P(H \wedge e)}{P(e)} = \frac{P(H \wedge e) \cdot P(H)/P(H)}{P((e \wedge H) \vee (e \wedge \neg H))} = \frac{P(e|H) \cdot P(H)}{P(e \wedge H) + P(e \wedge \neg H)} = \\ = \frac{P(e|H) \cdot P(H)}{P(e|H) \cdot P(H) + P(e|\neg H) \cdot P(\neg H)}$$

Wird jedoch  $H$  als wissenschaftliche Hypothese interpretiert und  $e$  als eine Menge von Ereignissen (z. B. Testergebnisse), wird der Bayes'sche Satz so problematisch wie interessant. Die linke Seite von Gleichung (1a) ist dann die Wahrscheinlichkeit der Hypothese  $H$  angesichts der Ereignisse  $e$ , d. h. der Bayes'sche Satz gibt eine Methode an, die Wahrscheinlichkeit von Hypothesen im Licht neuer Ereignisse zu berechnen. Die Probleme ergeben sich aber direkt aus dem letzten Term in Gleichung (1a). Während  $P(e|H)$  noch einfach die Wahrscheinlichkeit für die Ereignisse  $e$  ist, falls  $H$

---

\* Ich möchte Prof. C. Kenneth Waters und den Teilnehmern seines Seminars an der University of Minnesota über wissenschaftliche Forschung für hilfreiche Diskussionen danken, unter ihnen besonders Toben Lafrancois für Unterstützung bei der Lektüre von Mayos Buch.

<sup>1</sup> Seitenzahlen beziehen sich auf dieses Werk, in den Zitaten sind Hervorhebungen im Original.

gilt, steht der Term  $P(H)$  für die *a priori*-Wahrscheinlichkeit der Hypothese  $H$  ohne Berücksichtigung von  $e$ , was eine recht ungreifbare, bestenfalls schätzbare Größe ist. Problematischer aber ist die Größe  $P(e|\neg H)$ . Sie gibt die Wahrscheinlichkeit für  $e$  an, falls  $H$  nicht gilt. Um zumindest festhalten zu können, was das überhaupt bedeutet, kann in (1a)  $\neg H$  aufgeteilt werden in disjunkte Hypothesen  $\{H_i\}_{i \in I \setminus \{k\}}$  mit einer Indexmenge  $I$ , so dass, mit  $H_k := H$ ,  $\{H_i\}_{i \in I}$  die Menge aller einander ausschließender Hypothesen ist. Diese Aufteilung führt zu

$$(1b) \quad P(H_k|e) = \frac{P(e|H_k) \cdot P(H_k)}{P(e|H_k) \cdot P(H_k) + \sum_{i \neq k} P(e|H_i) \cdot P(H_i)}$$

Gleichung (1b) setzt für die Berechnung der *a posteriori*-Wahrscheinlichkeit  $P(H_k|e)$  voraus, dass *alle möglichen* konkurrierenden Hypothesen bekannt sind. Das ist offensichtlich nur in wenigen Fällen möglich und geht dann einher mit einer starken Einschränkung der Reichweite der Hypothesen, z. B. auf Aussagen über die Wahrscheinlichkeit der Ereignisse.

Diese beiden Probleme haben viel Kritik auf die bayesianistische Schlussweise gezogen, unter anderem von Deborah Mayo, die die Probleme mit ihrem Ansatz der Fehlerstatistik umgehen will.

### 3. Formalisierung des Fehlerarguments

Mayos Ansatz benutzt die von Neyman und Pearson entwickelte statistische Methode, um Aussagen über Hypothesen machen zu können. Ausgangspunkt dafür sind Experimente:

Experimental activities do offer especially powerful grounds for arriving at data and distinguishing real effects from artifacts, but what are these grounds and why are they so powerful? [...] What is the structure of the argument for arriving at this knowledge? My answer is [...]: [I]t follows the pattern of *an argument from error* or *learning from error*. (S. 64)

Um also aus den Experimenten mit Hilfe der Neyman/Pearson-Methode Erkenntnis zu gewinnen, wird nach Mayo das Fehlerargument benötigt. Dessen formale Version<sup>2</sup> ist das

*Severity requirement*: Passing a test  $T$  (with  $e$ ) counts as a good test of or good evidence for  $H$  just to the extent that  $H$  fits  $e$  and  $T$  is a *severe test* of  $H$ . (S. 180)

Als Interpretation von ‚fit‘ gibt Mayo an: „ $e$  fits  $H$  to the extent that  $P(e|H)$  is high“,<sup>3</sup> wobei sich  $e$  auf den Test  $T$  mit seinem Ergebnis beziehen muss, denn Messdaten sind nutzlos ohne die Angabe des Tests  $T$ , der sie hervorgebracht hat. Wie sich zeigen wird,

<sup>2</sup> S. 64 Fußnote

<sup>3</sup> S. 179 Fußnote

ist die Strenge (*severity*) eines Tests  $T$  für  $H$  auch abhängig vom Testergebnis, so dass sich als Abkürzung für die Strenge  $SEV(H, e)$  anbietet, wobei  $e$  wieder für die Kombination von Test  $T$  und seinem Ergebnis steht.

Mit der weiteren Abkürzung  $EVID(H, e)$  für die Güte der Evidenz von  $e$  für  $H$  ergibt sich folgende Form des *severity requirement*:

1.  $EVID(H, e)$  ist monoton steigend in  $P(e|H)$  und
2.  $EVID(H, e)$  ist monoton steigend in  $SEV(H, e)$ .

Die kontinuierliche Formulierung klingt vielleicht schwächer als das *severity requirement*, ist in diesem aber implizit, da nicht angegeben wird, welche Schwellenwerte für  $P(e|H)$  und  $SEV(H, e)$  (von Mayo ebenso als kontinuierlich auf dem Intervall  $[0,1]$  angenommen)<sup>4</sup> gelten. Natürlich kann aber ein Schwellenwert nach pragmatischen Kriterien eingeführt werden. Die Strenge des Tests ist gegeben durch das

*Severity criterion* [...]: There is a very low probability that test procedure  $T$  would yield so good a fit, if  $H$  is false. (S. 180)

Da ‚fit‘ über das Testergebnis definiert wurde, ist also die Strenge eines Tests abhängig vom Test  $T$  selbst und seinem Ergebnis, d. h. von  $e$ . Die Strenge eines Testes ist damit hoch, wenn ein hoher Wert für  $P(e|\neg H)$  unwahrscheinlich, d.h.  $P(e|\neg H)$  niedrig ist. Also ist  $SEV(H, e)$  monoton fallend in  $P(e|\neg H)$  und daher monoton steigend in  $1 - P(e|\neg H)$ . Damit ist  $SEV(H, e)$  im *severity requirement* eliminierbar, da ‚monoton steigend‘ transitiv ist:  $EVID(H, e)$  ist monoton steigend in  $1 - P(e|\neg H)$  oder eben monoton fallend in  $P(e|\neg H)$ . Das vollständige *severity requirement* lautet damit

1.  $EVID(H, e)$  ist monoton steigend in  $P(e|H)$  und
2.  $EVID(H, e)$  ist monoton fallend in  $P(e|\neg H)$ .

Diese Aussagen folgen, wie ich zeigen werde, aus der bayesianistischen Schlussweise.

#### 4. Herleitung des Fehlerarguments

Im bayesianistischen Rahmen ist  $e$  *ceteris paribus* um so bessere Evidenz für  $H$ , je höher  $P(H|e)$  liegt, dass heißt

$EVID(H, e)$  ist streng monoton steigend in  $P(H|e)$ .

---

<sup>4</sup> S. 182

Damit reicht es, wieder aufgrund der Transitivität von ‚monoton steigend‘, aus der bayesianistischen Schlussweise die folgenden Aussagen abzuleiten:

1.  $P(H|e)$  ist monoton steigend in  $P(e|H)$  und
2.  $P(H|e)$  ist monoton fallend in  $P(e|\neg H)$ .

Bei den beiden Beweisen muss folgendes beachtet werden: Erstens müssen natürlich die Axiome der Wahrscheinlichkeitstheorie erfüllt sein, d. h. unter anderem müssen die Wahrscheinlichkeiten normiert sein. Zweitens gelten die beiden Abhängigkeiten nur *ceteris paribus*, so dass also alle Werte konstant gehalten werden müssen, die von den beiden betrachteten Größen unabhängig sind.

Um die Normierung der Wahrscheinlichkeit zu gewährleisten, ersetze ich in (1a)  $P(\neg H)$  durch den äquivalenten Term  $1 - P(H)$ . Aussage 1 ergibt sich dann durch partielle Ableitung nach  $P(e|H)$  über die Kettenregel:

$$\begin{aligned}
 \frac{\partial P(H|e)}{\partial P(e|H)} &= \frac{\partial}{\partial P(e|H)} \frac{P(e|H) \cdot P(H)}{P(e|H) \cdot P(H) + P(e|\neg H) \cdot [1 - P(H)]} = \\
 (2) \quad &= \frac{P(H) \cdot [P(e|H) \cdot P(H) + P(e|\neg H) - P(e|\neg H) \cdot P(H)] - P(e|H) \cdot P(H)^2}{\{P(e|H) \cdot P(H) + P(e|\neg H) \cdot [1 - P(H)]\}^2} = \\
 &= \frac{P(H) \cdot P(e|\neg H) \cdot [1 - P(H)]}{\{P(e|H) \cdot P(H) + P(e|\neg H) \cdot [1 - P(H)]\}^2} \geq 0
 \end{aligned}$$

Die letzte Relation gilt, weil der Nenner des Bruches als Quadrat nicht negativ sein kann und da  $P(H) \leq 1$  gilt. Die Gleichheit gilt nur für die trivialen Fälle, dass einer der Terme in Zähler des Bruches 0 ist.

Aussage 2 ergibt sich auf analoge Weise:

$$(3) \quad \frac{\partial P(H|e)}{\partial P(e|\neg H)} = \frac{-P(e|H) \cdot P(H) \cdot [1 - P(H)]}{\{P(e|H) \cdot P(H) + P(e|\neg H) \cdot [1 - P(H)]\}^2} \leq 0$$

Damit ist Mayos Fehlerargument aus der bayesianistischen Schlussweise abgeleitet: Die Steigung von  $P(H|e)$  in Abhängigkeit von  $P(e|H)$  ist positiv und in Abhängigkeit von  $P(e|\neg H)$  negativ. Es bleibt aber die Frage, ob nicht wegen der in Abschnitt 2 genannten Probleme die bayesianistische Schlussweise verworfen und nur eine ihrer Implikationen, eben das Fehlerargument, beibehalten werden sollte. In dieser Hinsicht sind aber die Probleme nur bedingt relevant, da es ausreicht, dass die Wahrscheinlichkeiten im Bayes'schen Satz (in welcher Interpretation auch immer) wohldefiniert sind: Der Beweis kommt ohne irgendwelche Angaben über *a priori*-Wahrscheinlichkeiten oder konkurrierende Hypothesen aus und ist damit unangreifbar für Argumente, die auf die Unbestimmbarkeit oder Beliebigkeit dieser Elemente der bayesianistischen Schlussweise zielen. Damit ist die bayesianistische Schlussweise auch unter der Annahme der Unbe-

stimmmbarkeit oder Beliebbarkeit von *a priori* Wahrscheinlichkeiten respektive konkurrierender Hypothesen noch mindestens so stark wie Mayos Fehlerargument.

## 5. Zur Stärke des Fehlerarguments

Ein Argument, dass sich gegen die Bestimmbarkeit konkurrierender Hypothesen wendet, ist für Mayos Fehlerargument mindestens so problematisch wie für die bayesianistische Schlussweise: Ihr *severity criterion* und damit auch das Fehlerargument enthalten die Wahrscheinlichkeit für die Ereignisse  $e$  bei  $\neg H$ ,  $P(e|\neg H)$ . Genau dieser Term aber führte beim Bayes'schen Satz von der Formulierung (1a) zur Formulierung (1b). Wenn also Mayo den Term  $P(e|\neg H)$  als unproblematisch akzeptiert, ist dieses Argument gegen bayesianistische Schlüsse hinfällig. Umgekehrt sind die offensichtlichen Schwierigkeiten mit diesem Term ein Problem für das Fehlerargument. Mayo nimmt bei ihrer Antwort auf eine entsprechende Kritik Earmans Bezug auf einen Test der von der Allgemeinen Relativitätstheorie vorhergesagten Lichtablenkung:

If alternative theories predict the same results, so far as the deflection effects go, [...] then these alternatives are not *rivals* to the particular hypotheses under test. [...] In relation to the hypotheses about the deflection effect, alternatives to the general theory of relativity are on a higher level. The higher-level alternatives are not even being tested by the test at hand. (S. 188f)

Mayo entgeht also dem Problem, indem sie Reichweite der Hypothesen so stark einschränkt, dass  $\neg H$  handhabbar wird: Im Falle der Lichtablenkung ist die Hypothese  $H$ , die getestet wird, nur noch die, dass mit einer gewissen Wahrscheinlichkeit eine Ablenkung gemessen wird gibt.  $\neg H$  ist dann eben, dass die Wahrscheinlichkeit für die Feststellung der gemessenen Ablenkung eine andere ist. Allgemein ist für bestimmte Ereignisse  $e$  die Menge der rivalisierenden Hypothesen also  $\{H_r\}_{r \in [0,1]}$  mit  $H_r := „e$  tritt mit Wahrscheinlichkeit  $r$  auf“<sup>4</sup>. Aber diese Lösung des Problems steht, wie in Abschnitt 2 dargelegt, auch bei der bayesianistischen Schlussweise zur Verfügung. Der Lösungsversuch bringt aber ein neues Problem mit sich: Die Schlussweise ist nicht mehr ampliativ, d. h. wenn die Ereignisse  $e$  eintreten, kann man daraus nur noch schließen, dass  $e$  der Fall ist. Mayo behauptet dennoch, dass mit ihrem Ansatz auch größere Theorien (*higher-level theories*) getestet werden können:

I am not denying the possibility of severe tests of higher-level theoretical hypotheses. When enough is learned from piecemeal studies, severe tests of higher-level theories are possible. [...] The accumulated results from piecemeal studies allow us at some point to say that several related hypotheses are correct[.] (S. 191)

Aber das ist keine Antwort auf das Problem: Es ist klar, dass man aus den isolierten Ereignissen  $e_1, e_2, \dots, e_n$  folgern kann, dass  $e_1, e_2, \dots, e_n$  der Fall ist. Aber das ist immer noch nicht ampliativ. Nach wie vor sind *alle* in  $e_1, e_2, \dots, e_n$  übereinstimmenden Theorien

gleichberechtigt, auch jene, die abgesehen von diesen Ereignissen einander vollkommen entgegengesetzte Voraussagen machen.

## **6. Bemerkungen**

Mayos Fehlerargument ist ein Korollar der bayesianistischen Schlussweise. Das ist positiv sowohl für Anhänger des Fehlerarguments als auch der bayesianistischen Schlussweise: Ein Bayesianist stimmt mit Mayos Ergebnissen überein, soweit sie auf dem Fehlerargument beruhen, und Mayo befürwortet manche bayesianistischen Schlüsse, manche allerdings muss sie für zu stark oder sogar sinnlos halten.

Das Fehlerargument entgeht bestimmten Problemen der bayesianistischen Schlussweise, dafür sind aber auch die Folgerungen, die das Fehlerargument zulässt, wesentlich schwächer. Das ist wohl auch eine triviale Moral dieses Beitrags: Die Stärke der möglichen Folgerungen ist streng monoton steigend in der Stärke der Annahmen.

## **Literatur**

Mayo, Deborah G. (1996). *Error and the Growth of Experimental Knowledge*. Chicago und London: The University of Chicago Press.