

MAGKS Doktorandenkurs Wissenschaftstheorie

1. Logik und Erkenntnistheorie ✓
2. Unrealistische Annahmen und empirische Prüfung ✓
3. Der Bayesianismus
 - 3.1 Wahrscheinlichkeiten
 - 3.2 Geschichte des Induktionsproblems
 - 3.3 Bayesianisches Lernen
 - 3.4 Kritik am Bayesianismus

106

3.1 Wahrscheinlichkeiten

- 3.1.1 Ein ökonomisches Experiment
- 3.1.2 Axiome der Wahrscheinlichkeitstheorie
- 3.1.3 Wahrscheinlichkeiten und Aussagen
- 3.1.4 Interpretationen der Wahrscheinlichkeitstheorie

107

3.1.1 Ein ökonomisches Experiment

„Endowment effect“ (Kahneman, Knetsch & Thaler 1991)

- N VP, 2 Versuchsbedingungen, „Between-subjects“-Design:
 $N/2$ Verkäufer, $N/2$ Wähler (im folgenden: $N = 20$)
- Messung der Zahlungsbereitschaft (WTP) durch den Becker-DeGroot-Marschak-Mechanismus (anreizkompatibel)

Ziel des Experiments: Überprüfung der ökonomischen Theorie menschlichen Verhaltens

Aus dieser Theorie folgt deduktiv, dass die Vpn eine exogen gegebene WTP haben.

Problem: Wie kann man das testen? Jede Person ...

- ... ist entweder Verkäufer oder Wähler;
- ... hat eine eigene WTP.

108

Lösung: Man leitet eine überprüfbare Hypothese (Nullhypothese) ab. Dazu braucht man eine Hilfshypothese, mit der Wahrscheinlichkeiten eingeführt werden.

Hilfshypothese H_H : Alle Aufteilungen der V_{pn} auf die beiden Gruppen sind gleichwahrscheinlich.

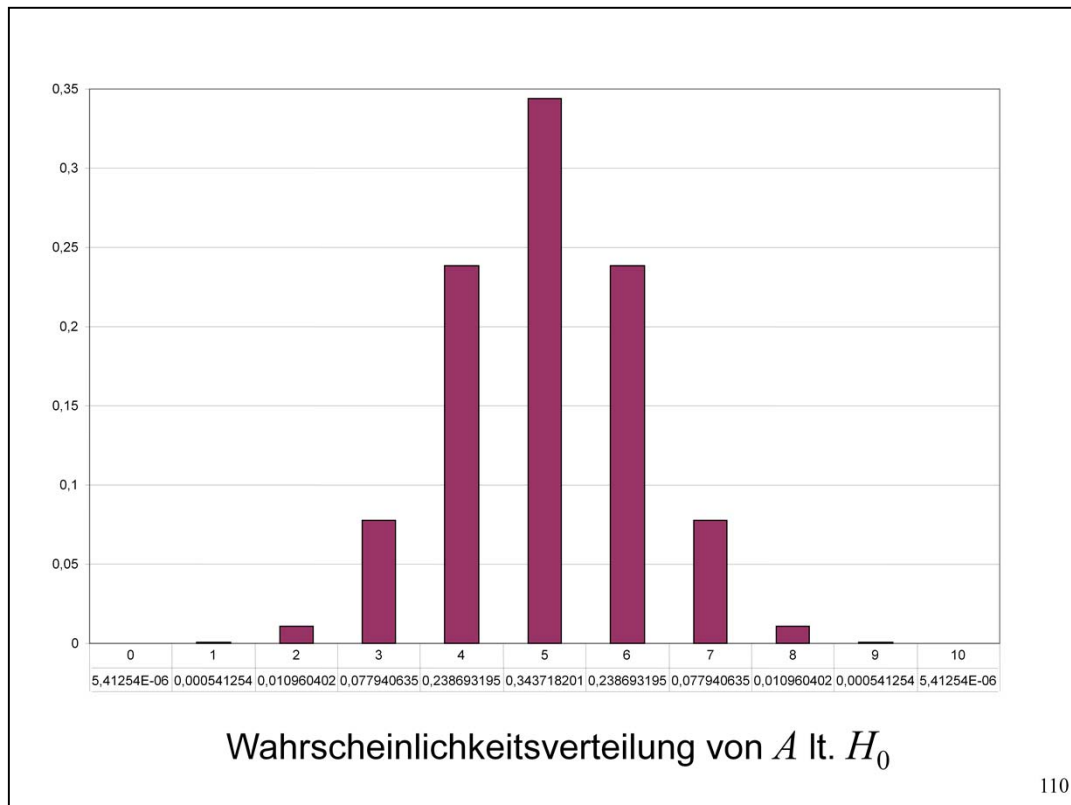
Teststatistik: Sei m der beobachtete Median der WTP über alle V_p . Benutze die Zahl A der Verkäufer mit $WTP \geq m$ als Teststatistik.

Nullhypothese H_0 : A ist verteilt nach einer speziellen hypergeometrischen Verteilung (s. nächste Folie).

Mediantest (einseitig): Betrachte H_0 als widerlegt, wenn A unwahrscheinlich hoch ist.

	$WTP \geq m$	$WTP < m$	Σ
Verkäufer	A	B	10
Käufer	C	D	10
Σ	10	10	20

109



Wichtig: Der statistische Test basiert auf H_H .

H_H : Alle Aufteilungen der Vpn auf die beiden Gruppen sind gleichwahrscheinlich.

Fragen

Was bedeutet H_H ?

Wann ist es vernünftig anzunehmen, dass H_H wahr ist?

Was kann man aus den Ergebnissen des Experiments (beobachteter Wert von A) schließen?

Ist es sinnvoll, einen solchen Test wie den Mediantest zu verwenden?

Die Antworten auf diese Fragen hängen davon ab, wie man das Wort „Wahrscheinlichkeit“ interpretiert. Es gibt mehrere Interpretationen, die auch als „philosophische Wahrscheinlichkeitstheorien“ bezeichnet werden.

111

3.1.2 Axiome der Wahrscheinlichkeitstheorie

Menge von Elementarereignissen: Ω

Menge von Ereignissen: Algebra Σ über Ω

Eine Menge Σ von Teilmengen von Ω heißt Algebra über Ω gdw $\Sigma \neq \emptyset$, $E_1 \cup E_2 \in \Sigma$ für alle $E_1, E_2 \in \Sigma$ und $\bar{E} \in \Sigma$ für alle $E \in \Sigma$ (nicht-leer und abgeschlossen unter allen endlichen Mengenoperationen).

Axiome (vereinfacht, für endliches Ω): Ein Wahrscheinlichkeitsmaß auf (Ω, Σ) ist eine reelle Funktion p auf Σ mit

- i) $p(E) \geq 0$,
- ii) $p(\Omega) = 1$ und
- iii) $p(E_1 \cup E_2) = p(E_1) + p(E_2)$ if $E_1 \cap E_2 = \emptyset$.

Wichtig: Diese Axiome sagen nicht, was eine Wahrscheinlichkeit ist. Die Axiome werden durch viele Mengen und Größen erfüllt.

Beispiel: Ω sei eine Menge von 100 Leuten, Σ die Menge aller Teilmengen von Ω und $p(E) = |E|/100$.

112

Bedingte (konditionale) Wahrscheinlichkeiten

Seien E_1 und E_2 zwei Ereignisse. Wenn das Ereignis $E_1 \cap E_2$ eintritt, heißt das, dass E_1 und E_2 eintreten. Die Wahrscheinlichkeit, dass E_2 eintritt, unter der Bedingung, dass E_1 eintritt, ist

$$p(E_2|E_1) := p(E_1 \cap E_2)/p(E_1) \quad \text{falls } p(E_1) \neq 0.$$

Aus dieser Def. folgt unter der Annahme $p(E_1), p(E_2) \neq 0$:

$$p(E_2|E_1)p(E_1) \equiv p(E_1|E_2)p(E_2)$$

Daraus ergibt sich **Bayes' Theorem**:

$$p(E_2|E_1) \equiv p(E_1|E_2)p(E_2)/p(E_1) \quad \text{falls } p(E_1), p(E_2) \neq 0$$

3.1.3 Wahrscheinlichkeiten und Aussagen

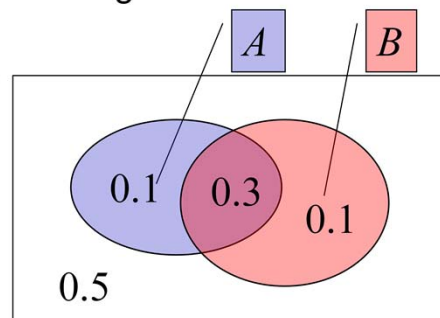
Wir betrachten wieder die Menge der möglichen Welten Ω .

Ω_A : Teilmenge möglicher Welten,
in den A wahr ist

$$\Omega_{\neg A} = \bar{\Omega}_A$$

$$\Omega_{A \vee B} = \Omega_A \cup \Omega_B$$

$$\Omega_{A \wedge B} = \Omega_A \cap \Omega_B$$



Sei p ein Wahrscheinlichkeitsmaß auf (Ω, Σ) . Die Elemente von Σ seien die Teilmengen Ω_A, Ω_B , etc.

Wahrscheinlichkeiten der Aussagen: $P(A) := p(\Omega_A)$, etc.

Bed. W. von B geg. A : $P(B|A) := P(A \wedge B)/P(A)$ falls $P(A) \neq 0$.

Beispiel: $P(A) = 0.4$. $P(A \wedge B) = 0.3$. $P(B|A) = 0.75$.

114

Aus der Definition der W. für Aussagen ergibt sich:

$P(A) \geq 0$ (vgl. Axiom i der Wahrscheinlichkeitstheorie).

Wahrscheinlichkeit der Negation: $P(\neg A) = 1 - P(A)$

Sei W analytisch wahr („Alle Junggesellen sind unverheiratet.“), so dass $\Omega_W = \Omega$. Dann gilt $P(W) = p(\Omega) = 1$ (vgl. Axiom ii).

Sei F ein Widerspruch („Es existiert ein verheirateter Junggeselle.“), so dass $\Omega_F = \emptyset$. Dann gilt $P(F) = p(\emptyset) = 0$.

Ist $A \wedge B$ ein Widerspruch, gilt $P(A \vee B) = P(A) + P(B)$ (vgl. Axiom iii).

Wenn $A \Rightarrow B$, dann gilt $P(B) \geq P(A)$ (wg. $\Omega_A \subseteq \Omega_B$).

Mit $P(A) \neq 0$, $A \Rightarrow B$ gilt $P(B|A) = P(A \wedge B)/P(A) = 1$.

Mit $P(\neg B) \neq 0$, $A \Rightarrow B$ gilt $P(A|\neg B) = P(A \wedge \neg B)/P(\neg B) = 0$.

115

3.1.4 Interpretationen der Wahrscheinlichkeitstheorie

	objektive W.	subjektive* W.
physische W.	- Häufigkeitsinterpretation - Propensitätsinterpretation	(leer)
epistemische W.	- klass. Th. d. W. - Th. der logischen W. (auch: objektiver Bayesianismus)	(subjektiver) Bayesianismus (in der Entscheidungstheorie auch SEUTheorie)

1. Epistemische Wahrscheinlichkeitsaussagen sind Urteile darüber, wie sicher man ist (oder vernünftigerweise sein sollte), dass eine Aussage wahr ist. Sie sind weder wahr noch falsch. Epist. W. können allen Aussagen zugewiesen werden.
2. Physische W. sind Eigenschaften bestimmter Situationen (genannt „Zufallsexperimente“). Sie können nur Ergebnissen von Zufallsexperimenten bzw. den entsprechenden Aussagen zugewiesen werden. Physische Wahrscheinlichkeitsaussagen sind wahr oder falsch.

116

Theorie der logischen Wahrscheinlichkeiten

- Die W. werden durch Logik und weitere Prinzipien bestimmt.
- Diese Prinzipien sind meist Gleichwahrscheinlichkeitsannahmen: „Indifferenzprinzip“, „Prinzip des unzureichenden Grundes“ (vgl. auch die klassische Theorie der W.).
- Es wird behauptet, dass die Akzeptanz dieser Prinzipien rational zwingend ist. Damit wären die logischen W. durch Logik und Rationalität fixiert und in diesem Sinne objektiv.

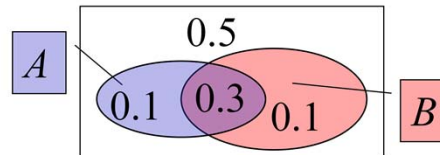
(Subjektiver) Bayesianismus

- W. sind Grade des persönlichen Glaubens.
- Rationalität erfordert, dass diese Glaubensgrade die Axiome der W. erfüllen („Dutch book argument“).
- Rationalität erfordert (fast) keine weiteren Restriktionen: Ablehnung des Indifferenzprinzips etc. Damit ergibt sich ein weiter Spielraum für die Wahl subjektiver W.

117

Idee der probabilistischen Induktion

Die Zuweisung von epistemischen W. an alle Aussagen erlaubt eine Verallgemeinerung der logischen Folgerung, nämlich die probabilistische Stützung: A stützt B gdw $P(B|A) > P(B)$. Eine solche Stützung ist zwischen beliebigen synthetischen Aussagen A und B möglich, wenn sie sich nicht widersprechen und keine Folgerungsbeziehung zwischen ihnen herrscht.



Beispiel: A stützt B wg. $P(B|A) = 0.75 > P(B) = 0.5$. Aussagen können sich gegenseitig stützen (wie hier) oder sich gegenseitig schwächen. Asymmetrie ist unmöglich.

Induktion: Sei T eine Theorie und B eine Beobachtungsaussage. B stützt T induktiv gdw $P(T|B) > P(T)$.

118

Häufigkeitsinterpretation

Zufallsexperiment: beliebige wiederholbare Situation, in der unterschiedlichen Ereignisse eintreten können.

Sei S ein Zufallsexperiment. Sei $h_n(E)$ die Häufigkeit von Ereignis E in den ersten n einer hypothetischen unendlichen Zahl von Durchführungen von S . Die W. von E ist definiert als

$$p(E) := \lim_{n \rightarrow \infty} h_n(E)/n,$$

falls der Limes existiert und jede berechenbare Teilfolge denselben Grenzwert hat (Zufälligkeit).

Anmerkungen zur Häufigkeitsinterpretation

- Besser: $p(E|S)$ statt $p(E)$ (Popper: Axiome für bed. W.).
- Wissenschaftliche Theorien und Hypothesen haben keine W. i.S.d. Interpretation, da über ihre Wahrheit nicht in einem Zufallsexperiment entschieden wird.
- Es gibt deterministische Prozesse (Chaos), die die oben beschriebenen Folgen erzeugen. Die Häufigkeitsinterpretation ist daher mit dem Determinismus vereinbar.

119

Problem des Einzelfalls

Wahrscheinlichkeiten i.S.d. Häufigkeitsinterpretation sagen nichts über einzelne (oder auch endlich viele) Beobachtungen. Sie liefern daher keine vernünftige Basis für Entscheidungen oder Statistik.

Propensitätsinterpretation

Wegen des Problems des Einzelfalls und angeregt durch die Quantentheorie hat Popper eine andere physische Interpretation von Wahrscheinlichkeiten vorgeschlagen. Nach dieser Interpretation sind Wahrscheinlichkeiten Eigenschaften eines einzelnen Zufallsexperiments. Im Gegensatz zur Häufigkeitsinterpretation ist diese neue Interpretation nicht mit dem Determinismus vereinbar.

120

„Principal Principle“ (PP, David Lewis): Kombination von subjektiven und physischen Wahrscheinlichkeiten

Sei H eine Hypothese, die für ein Ereignis E eine physische W. p behauptet (Bsp.: Hilfshypothese im Experiment).

Sei P die subjektive W. einer bestimmten Person.

PP fordert: $P(E|H) = p$.

Also: Unter der Annahme, dass H wahr ist, sollte der Grad des persönlichen Glaubens, dass E eintritt, der von H behaupteten physischen Wahrscheinlichkeit p entsprechen.

PP könnte man auch für logische statt subjektive W. fordern.

Frage: Ist PP für jede physische Interpretation plausibel? Lewis: nur für Propensitäten. Howson und Urbach: nur für Häufigkeitstheorie (s. auch Anscombe-Aumann-Version der SEU-Theorie).

121

3.2 Geschichte des Induktionsproblems

Induktion: Schluss von Beobachtungen auf Prognosen oder Hypothesen, aus denen Prognosen folgen

Hume (1711-1776): Induktion ist irrational, Ergebnisse sind

- nicht sicher wahr;
- nicht wahrscheinlicher als andere Vermutungen.

Induktionsproblem: Die wissenschaftliche Methode gilt als induktiv und rational. Laut Hume ist das unmöglich.

Irrationalismus: Jede Prognose/Überzeugung ist genauso (un)vernünftig wie irgendeine andere.

Bayes (1702-1761): Theorie rationalen Lernens auf Basis epistemischer (vielleicht: subjektiver) W., publiziert 1763 durch Price (der meinte, Bayes habe Hume widerlegt)

122

Laplace (1812, 1814): integriert Bayes' Ansatz in die allgemeine Philosophie, „klassische“ Theorie der Wahrscheinlichkeit (bei Laplace: logische W.), Uneindeutigkeiten

Keynes, Jeffreys, Carnap: Theorie der logischen Wahrscheinlichkeit, Uneindeutigkeiten, gilt weitgehend als gescheitert

Fisher, Popper: Gegenbewegung zur Bayes-Laplace-Tradition

- Signifikanztest bzw. Falsifikationismus: Empirische Tests führen zu vorläufiger Widerlegung oder Bewährung.
- Häufigkeitsinterpretation der W. (Popper später: Propensitätsinterpretation), setzt Zufallsexperiment voraus
- Hypothesenw. illegitim (Zufallsexperiment fehlt)
- nötigenfalls: explizite Randomisierung (s. Experiment)

„Pragmatisches“ Induktionsproblem offen: Soll man so handeln, als seien bewährte Hypothesen sicher wahr?

123

Neyman, Pearson, Wald (ab 1930): statist. Entscheidungsth.

- gegebene Hypothesenmenge (statistisches Modell)
- Unterscheidung Fehler 1. und 2. Art
- Maximin-Kriterium (leicht paranoid): offenes Einfallstor für Bayes-Laplace-Tradition

Savage (1954): moderner Bayesianismus

- statistische Entscheidungstheorie mit subjektiven Hypothesenwahrscheinlichkeiten statt Maximin-Kriterium
- radikaler Subjektivismus: alle Wahrscheinlichkeiten subjektiv
- Vorläufer: Ramsey, de Finetti (nicht durchschlagend)
- axiomatische Begründung: Nutzen und subjektive W. aus Präferenzordnung auf der Menge aller Strategien (Strategie: vollständiger bedingter Plan)

124

3.3 Bayesianisches Lernen

3.3.1 Grundlegende Idee

3.3.2 Ein komplexeres Beispiel

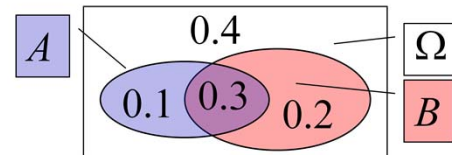
125

3.3.1 Grundlegende Idee

Annahme: Wir beobachten A und interessieren uns für B . Deduktion sei nutzlos: $A \not\Rightarrow B$, $A \not\Rightarrow \neg B$.

Bayesianisches Lernen ("Konditionalisierung")

- Wahrscheinlichkeiten auf Ω liefern $P(A)$, $P(B)$... („Apriori-Wahrscheinlichkeiten“).



- Wird A beobachtet, ergibt sich die neue W. von B („Apriori-W.“) als konditionale W. von B geg. A : $P_{\text{neu}}(B) = P(B|A)$.

Diagramm: Masse auf Ω , W. = Massenanteile, Fläche und totale Masse schrumpfen durch Beobachtung, neue Anteile.

Beispiel: $P(A \cap B) = 0.3$, $P(A \cap \neg B) = 0.1$, $P(\neg A \cap B) = 0.2$, $P(\neg A \cap \neg B) = 0.4 \Rightarrow P(A) = 0.4$, $P(B) = 0.5$

Lernen: $P_{\text{neu}}(B) = P(B|A) = P(A \cap B)/P(A) = 0.75$

126

3.3.2 Ein komplexeres Beispiel

Gegeben: Menge von Hypothesen (physische Wahrscheinlichkeiten, statistisches Modell), epistemische Hypothesenwahrscheinlichkeiten, Geltung von PP

S_t , $t = 1, 2, \dots$: An Tag t geht die Sonne auf.

H_1 : Jeden Tag geht die Sonne auf, d.h. $H_1 \Rightarrow S_t$, $P(S_t|H_1) = 1$.

H_2 : $P(S_t|H_2) = 0.5$

Apriori-Wahrscheinlichkeiten: $P(H_1) = P(H_2) = 0.5$

Beobachtung: $B = S_1 \wedge S_2$

Fragen

Bayesianisches Lernen und Falsifikation?

Apriori-W. $P(S_3)$ für Sonnenaufgang am dritten Tag?

Aposteriori-Wahrscheinlichkeiten $P(H_i|B)$ und $P(S_3|B)$?

127

Falsifikation von H_1 gem. Bayes' Theorem: Es gilt $P(H_1)$, $P(S_t)$, $P(\neg S_t) \neq 0$, $H_1 \Rightarrow S_t$. Dann:

$$P(H_1|\neg S_t) = P(\neg S_t|H_1)P(H_1)/P(\neg S_t) = 0$$

H_2 : nicht falsifizierbar, weil sie keine Beobachtung ausschließt

Satz über die totale Wahrscheinlichkeit liefert Apriori-W.:

$$P(S_3) = P(S_3|H_1)P(H_1) + P(S_3|H_2)P(H_2) = \mathbf{0.75}$$

$$P(B) = P(B|H_1)P(H_1) + P(B|H_2)P(H_2) = 0.625$$

Bayes' Theorem liefert Aposteriori-W. von H_1 , H_2 :

$$P(H_1|B) = P(B|H_1)P(H_1)/P(B) = \mathbf{0.8} > 0.5 = P(H_1)$$

$$P(H_2|B) = P(B|H_2)P(H_2)/P(B) = \mathbf{0.2} < 0.5 = P(H_2)$$

Satz über die totale bedingte Wahrscheinlichkeit liefert Aposteriori-Wahrscheinlichkeit von S_3 :

$$P(S_3|B) = P(S_3|H_1)P(H_1|B) + P(S_3|H_2)P(H_2|B) = \mathbf{0.9} > 0.75_{128}$$

Einfacher: Bayes' Theorem für Verhältnisse von Hypothesenwahrscheinlichkeiten

Likelihood Ratio: $LR(B) := P(B|H_1)/P(B|H_2)$

Bayes' Theorem: $P(H_1|B)/P(H_2|B) = LR(B)P(H_1)/P(H_2)$
(„posterior odds = likelihood ratio $\times$ prior odds“)

Also: $P(B|H_1)/P(B|H_2) = 4$ und $P(H_1|B) + P(H_2|B) = 1$
 $\Rightarrow P(H_1|B) = 0.2, P(H_2|B) = 0.8$

Wichtig: Im Gegensatz etwa zum Signifikanztest spielt für das bayesianische Lernen nur die Beobachtung selbst eine Rolle, kein Bereich möglicher Beobachtungen (Ablehnungsbereich). Entscheidend ist die Likelihood Ratio LR : Wie wahrscheinlich ist die Beobachtung unter den beiden Hypothesen?

$LR > 1$ heißt: Unter H_1 ist die Beobachtung wahrscheinlicher als unter H_2 . Daher erhöht die Beobachtung die W. von H_1 .

129

Lernen mit Wahrscheinlichkeiten: Bildung bedingter W. („Konditionalisieren“, Anwendung von Bayes' Theorem)

Lernen auf Basis gegebener physischer W.: nicht umstritten, rein deduktives Vorgehen (Bsp.: gegebene W. einer AIDS-Erkrankung aus Bevölkerungsstatistik, AIDS-Test mit zwei möglichen Fehlern)

Bayesianismus

- Formal wie Lernen auf Basis gegebener physischer W.
- Epistemische W. für Hypothesen bzw. für alle Aussagen, denen sich keine physische W. zuordnen lässt.

Objektiver Bayesianismus: Ausgangspunkt sind logische W., gilt weitgehend als gescheitert, da sich Uneindeutigkeiten nicht plausibel beseitigen lassen.

(Subjektiver) Bayesianismus: Ausgangspunkt sind subj. W., wird im folgenden behandelt.

130

4. Kritik am Bayesianismus

4.1 Bayesianische Induktion: Das Blün-Paradox

4.2 Bayesianische Induktion: Der allgemeine Fall

4.3 Zusammenfassung

Diskutiert wird im folgenden die Behauptung, der Bayesianismus sei eine Lösung des Induktionsproblems, d.h. eine allgemeine und befriedigende Theorie rationalen Lernens.

131

4.1 Bayesianische Induktion: Das Blün-Paradox

Goodman (1955): „new riddle of induction“ oder „grue paradox“ (dt.: Blün-Paradox, hier in einer leichten Abwandlung)

Hypothese H_0 : „Alle Smaragde sind grün.“

Empiristische Sicht: Diese Hypothese ist sehr gut bestätigt, weil soweit alle beobachteten Smaragde grün waren.

Frage: Kann man H_0 für Prognosen verwendbar?

Goodmans Argument: Es gibt unendlich viele ebenso gut bestätigte Hypothesen, die etwas anderes vorhersagen.

- Hypothese H_1 : „Alle Smaragde sind blün.“
(blün: grün bis einschließlich heute, ab morgen blau)
- Hypothesen $H_2, H_3, \dots$: Farbwechsel 1, 2, ... Tage später

Also: Empirische Bestätigung genügt nicht, um Hypothesen rational zu akzeptieren / für Prognosen zu verwenden.

132

Zwischenüberlegung: Wie geht der kritische Rationalismus mit dem Blün-Paradox um?

Naheliegende Lösung: Rückgriff auf Hintergrundwissen

1. Wir suchen nach Gesetzmäßigkeiten. Diese enthalten keine bestimmten Zeitangaben („heute grün, morgen blau“). Die blünartigen Hypothesen sind also keine Gesetzhypothesen.
2. Die Vorstellung, dass es Gesetzmäßigkeiten gibt, hat sich bewährt. Die Suche nach Gesetzmäßigkeiten war, nach allem was wir wissen, sehr erfolgreich.
3. Es gibt Bereiche, in denen wir uns zumindest vorläufig mit Quasi-Gesetzen zufrieden geben müssen. Die Chemie / Mineralogie gehört aber nicht dazu.

Daher: Aus unserem Hintergrundwissen folgt, dass Hypothese H_1 („Alle Smaragde sind blün.“) falsch ist.

133

Bayesianische Lösung für Blün-Paradox (Howson 2000, unveröffentlicht)

Annahme: H_0 („Alle Smaragde sind grün.“) ist wahr, aber niemand weiß das.

Wahl von Apriori-Wahrscheinlichkeiten

- Hypothesenmenge (abzählbar unendlich): $H_0, H_1, H_2, H_3, \dots$
- Anfangsverteilung P (fast) beliebig, natürlich mit $\sum_i P(H_i) = 1$.
- Fast beliebig: $P(H_0) > 0$. „Regularität“: $P(H_i) > 0$ für alle i , damit die Wahrheit immer gelernt werden kann.

2. Konditionalisierung: Beobachtungen falsifizieren $H_1, H_2, H_3, \dots$, eine Hypothese pro Tag. Relative W. aller nichtfalsifizierten Hypothesen bleibt gleich (alle „likelihood ratios“ sind 1).

Nach endlich vielen Tagen: $P(H_0) > P(H_i)$ für alle $i \neq 0$.

Langfristig: $P(H_0)$ konvergiert gegen 1.

134

Ergebnisse des bayesianischen Lernen im Blün-Problem

1. Konvergenz zur Wahrheit: Die subj. W. der wahren Hypothese H_0 konvergiert gegen 1 (falls $P(H_0) > 0$).

2. Verbesserung der Vorhersage: Angenommen, ein Bayesianer muss Punktvorhersagen machen („grün“ oder „blau“). Dann sind seine Vorhersagen nach endlicher Zeit immer korrekt (bzw. bei stochastischen Hypothesen: optimal).

3. Subjektive Sicherheit der Konvergenz zur Wahrheit: Sei S_i die Reihe von Beobachtungen, für die $P(H_i)$ gegen 1 konvergiert. Dann gilt $P(S_i | H_i) = 1$.

4. „Merger of opinions“: Die Aposteriori-W. von Bayesianern mit unterschiedlichen Apriori-W. konvergieren (falls alle $P(H_0) > 0$ setzen).

Bayesianische Folklore: Diese Ergebnisse gelten allgemein unter „schwachen“ Bedingungen.

135

4.2 Bayesianische Induktion: Der allgemeine Fall

Blün-Paradox: Viele falsche Hypothesen haben von Beginn an Wahrscheinlichkeit 0.

Digitale Darstellung: $x_t = 0/1$ bedeutet „Smaragde sind grün / blau an Tag t “.

- Hypothese H_0 : Es wird nur 0 beobachtet, d.h. 000000...
- Hypothese H_6 : 0000001111...

Allgemeines Induktionsproblem: Keinerlei Wissen kann vorausgesetzt werden. Hypothesen wie 01100001001... (beliebige Folgen) müssen ebenfalls betrachtet werden.

Also: Wir betrachten die Menge aller unendlichen 0-1-Folgen als Hypothesenmenge für das Induktionsproblem.

Das ist das allgemeine Induktionsproblem für deterministische Hypothesen, da wir alle Beobachtungen digitalisieren können.

Die Betrachtung von stoch. Hyp. ändert die Folgerungen nicht. ¹³⁶

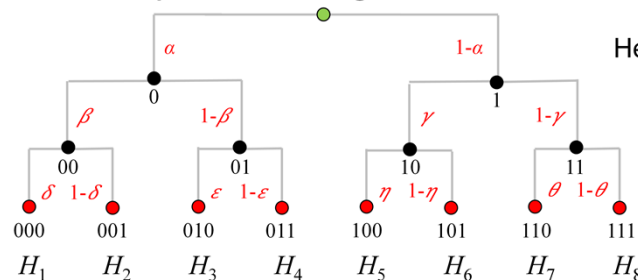
Universelles Vorhersageproblem: $x_t \in \{0,1\}$, $t = 1, \dots, n$

Hypothesen: H_i , $i = 1, \dots, 2^n$ (in Form eines binären Baums)

Vorhersagefkt.: Zuordnung von Wahrscheinlichkeiten zu allen Ästen des Baumes.

Bayesianisches Lernen: Die Vorhersagefkt. muss aus Apriori-W. $P(H_i)$ für die Hypothesen hergeleitet werden.

„Anything-goes theorem“ (Albert 2017): Jede Vorhersagefkt. kann aus Apriori-W. hergeleitet werden, sogar für $n = \infty$.



Herleitung Vorhersagefkt. (Bsp.):

$$\gamma = \frac{P(H_5) + P(H_6)}{\sum_{i=5}^8 P(H_i)}$$

Irrationalismus: Da alle Apriori-W. gleichermaßen (ir)rational sind, gilt das auch für alle Vorhersagefkt.

137

4.3 Zusammenfassung

1. Bayesianische Folklore: Im allgemeinen hat das bayesianische Lernen keine relevanten Konvergenzeigenschaften.
2. Der Bayesianismus war ursprünglich ein Vorschlag zur Lösung des Induktionsproblems mit Hilfe der Wahrscheinlichkeitsrechnung.

„Lösung“ heißt: Vermeidung des Irrationalismus.

Irrationalismus: Jede Überzeugung über die Zukunft ist so (ir)rational wie jede andere.

Das „Anything-goes“-Theorem zeigt, dass der Bayesianismus eine probabilistische Variante des Irrationalismus ist.

3. Man könnte argumentieren, dass der Bayesianismus zwar keine Lösung des Induktionsproblems bietet, aber irgendwelche anderen Vorzüge hat. In der Statistik wird oft für bayesianische Verfahren argumentiert. Diese Argumente muss man für sich betrachten. Aus der Erkenntnistheorie kann man keine Unterstützung für diese Verfahren ableiten.

138