Logik in der Statistik

Andrea Wiencierz

7. Oktober 2007

Inhaltsverzeichnis

Kapitel 1

Zur Logik in der Statistik

Abgesehen davon, dass die klassische binäre Logik als Grundprinzip der Mathematik allen statistischen Berechnungen zugrunde liegt, spielt die Logik in einem zentralen Gebiet der Statistik eine ganz besondere Rolle, nämlich in der statistischen Inferenztheorie.

1.1 Inferenz oder induktive Statistik

Das Ziel von Inferenztheorie kann darin gesehen werden, eine Antwort auf die folgende Frage zu finden: Wie kann ich auf der Basis einer begrenzten Anzahl gemachter Beobachtungen allgemeine Aussagen über eine Grundgesamtheit treffen? Dafür gibt es unterschiedliche theoretische Ansätze. Jedes dieser theoretischen Konzepte basiert auf einer eigenen Weltsicht. Weltsicht bedeutet im Kontext der Statistik: jede Inferenztheorie legt einen eigenen Wahrscheinlichkeitsbegriff zugrunde und hat eine eigene Vorstellung davon, wie Beobachtungen zustande kommen. Durch diese beiden Komponenten wird dann wiederum bestimmt, welche Rückschlüsse bzw. Aussagen auf Basis der Beobachtungen möglich sind.

Vom Standpunkt der Logik aus betrachtet geht es bei Inferenz also darum, von Einzelfällen auf allgemeine Aussagen zu schließen, d.h. um induktive Schlüsse. Da ein solcher Schluss aber nicht fehlerfrei gemacht werden kann, wie schon David Hume im 18.Jahrhundert zeigte, haben wir Statistiker ein logisches Problem: das Induktionsproblem. Vor diesem Hintergrund könnte man Inferenztheorien auch als Versuche bezeichnen, die negativen Auswirkungen dieses Problems so weit wie möglich zu begrenzen oder es möglichst gut zu umgehen. Denn schließlich sind zur Herleitung wissenschaftlich fundierter Aussagen, ganz im Sinne der analytischen Wissenschaftstheorie Karl Poppers, nur deduktive Schlussverfahren zulässig [Popper, 1934].

Zum Ende des 20. Jahrhunderts existieren drei wesentliche philosophische Standpunkte zur statistischen Inferenz: der Bayesianische, der Frequentistische sowie der Fisherianische. Während die Bayes-Theorie und die frequentistische Inferenz in vieler Hinsicht als zwei entgegengesetzte Pole bezeichnet werden können, sind Fishers Überlegungen irgendwo zwischen diesen beiden Extremalpunkten zu verorten. Im Folgenden möchte ich diese drei vorherrschenden Inferenzkonzepte noch einmal kurz gegenüberstellen, bevor ich dann zwei etwas unbekanntere Inferenztheorien vorstellen werde, denen eine nicht zu verachtende historische Bedeutung zukommt und deren Grundidee möglicherweise für die Zukunft wieder stärker von Bedeutung sein könnte.

1.2 Drei vorherrschende Inferenztheorien

Zunächst möchte ich versuchen, die drei allseits bekannten Inferenzkonzepte anhand der eben genannten Elemente zu charakterisieren.

1.2.1 Drei Weltsichten

Frequentistische Inferenz:

- *Wahrscheinlichkeitsbegriff*: Frequentistisch/objektiv, d.h. Wahrscheinlichkeit wird definiert als Grenzwert der relativen Häufigkeit einer Merkmalsausprägung bei unendlich häufigen Wiederholungen eines Experiments. Somit stellen die Wahrscheinlichkeiten der Ausprägungen quasi Eigenschaften des Merkmals dar und sind in diesem Sinne objektiv bzw. objektbezogen.

- *Beobachtungen*: Die Beobachtungen werden von der (parametrischen) Merkmalsverteilung bestimmt, deren Verteilungsparameter fest aber unbekannt sind. Man stellt sich die Welt "deterministisch" vor, d.h. man nimmt an, die Geschehnisse in der Welt folgten festen Gesetzmäßigkeiten.

- *Rückschlüsse*: Schlüsse von der Beobachtung auf die Grundgesamtheit sind nur als Aussagen über die Annehmbarkeit von bestimmten Parameterwerten im Lichte der gemachten Beobachtungen möglich. Die frequentistische Inferenz folgt also einer streng deduktiven Logik.

Bayes-Inferenz:

- *Wahrscheinlichkeitsbegriff*: Im Prinzip subjektiv, d.h. Wahrscheinlichkeiten drücken unvollständiges Wissen des Wissenschaftlers aus, sie hängen also vom analysierenden Subjekt ab.

- *Beobachtungen*: Was wir beobachten entspringt dem Zusammenspiel mehrerer Zufallsgrößen, denen jeweils dem Wissensstand entsprechende Wahrscheinlichkeitsverteilungen zugeordnet werden. Alles in der Welt passiert zufällig bzw. ist es letztlich egal, was man darüber denkt, denn ob man einen deterministischen Zusammenhang nur unzureichend kennt oder einen zufälligen exakt formal darstellen kann kommt hier quasi auf das Gleiche hinaus [de Finetti, 1974].

- *Rückschlüsse*: Da die Verteilungsparameter ebenfalls als Zufallsgrößen angenommen werden, können ihnen auch richtige Wahrscheinlichkeitsverteilungen zugeordnet werden, eben die auf die Beobachtung bedingten A-Posteriori-Verteilungen. Auf Grundlage dieser Verteilungen können dann Wahrscheinlichkeitsaussagen über die Parameter gemacht werden.

Fishers Likelihood-Inferenz: Der versuchte Kompromiss

- *Wahrscheinlichkeitsbegriff*: Fisher versteht Wahrscheinlichkeiten eher in einem klassisch objektiven Sinne.

- *Beobachtungen*: Wie in der klassischen Inferenz, geht man hier davon aus, dass die Beobachtungen von der Merkmalsverteilung bestimmt werden, über die eine Verteilungsannahme gemacht wird.

- *Rückschlüsse*: Neben der Verteilungsannahme geht ausschließlich die Information aus der Beobachtung in Aussagen über den Parameter ein. Die hier konstruierte Paramterverteilung, nämlich die Likelihood-Funktion, wird dann eher klassisch als Plausibilitätsgrad für bestimmte Parameterwerte vor dem Hintergrund der gemachten Beobachtung interpretiert.

Die Likelihood-Inferenz ist nicht die einzige Inferenztheorie, die auf Fisher zurückgeht. Er entwickelte ebenfalls das Konzept der sogenannten Fiduzialwahrscheinlichkeiten, welches ich im weiteren Verlaufe des Vortrags auch noch vorstellen werde. Die grundlegende Weltsicht Fishers bleibt auch darin erkennbar.

1.2.2 Eine Gegenüberstellung

Nun können die drei verschiedenen Philosophien nach Bradley Efron noch anhand von drei weiteren Gesichtspunkten gegenübergestellt werden:

BAYES	FISHER	FREQUENTIST
1. Individual (personal decisions)	***	Universal (world of science)
2. Coherent (correct)	*************	Optimal (accurate)
3. Synthetic (combination)	****	Analytic (separation)
4. Optimistic (aggressive)	*****	Pessimistic (defensive)

Abbildung 1.1: Tabelle aus [Efron, 1998]

Kohärenz vs. Optimalität: Bayesianer betonen die Kohärenz ihrer Ergebnisse, welche sich aus den Konsistenzanforderungen ihres Entscheidungsmodells ergibt, z.B. gibt es verschiedene Möglichkeiten einen bayesoptimalen Schätzer zu berechnen. Optimalität im frequentistischen Sinne ist hingegen häufig unkohärent. Fisher wollte hier sowohl Optimalität und Kohärenz erreichen und in der Tat erfüllen die ML-Schätzungen mit ihrer Äquivarianzeigenschaft auch beide Anforderungen. In der Praxis allerdings war Fisher durchaus bereit, die Forderung nach logischer Konsistenz für

eine einfachere Lösung eines konkreten Problems zu opfern.
Synthese vs. Analyse: Bayesianisches Entscheiden basiert auf Sammeln von Informationen aus allen möglichen Quellen, sowie der Synthese all dieser Informationen in der finalen Inferenz. Frequentisten hingegen tendieren dazu, die Probleme in kleine getrennte Stücke aufzuteilen, die dann einzeln analysiert werden, d.h. es werden optimale Lösungen für die separaten Teile bestimmt. Fisher ist hier näher an der Bayesianischen Position, da er die Verwendung aller Informationen grundsätzlich als wichtiges Kennzeichen korrekter Inferenz versteht.
Optimismus vs. Pessimismus: Bayesianer scheinen risikofreudiger bei ihren Analysen zu sein (eher optimistisch), Frequentisten hingegen gehen z.B. bei Minimax vom schlimmsten Fall aus (eher pessimistisch). Fisher sucht auch hierbei eine Mittelposition einzunehmen.

Kapitel 2

Fishers Fiduzialargument

Ronald A. Fisher (1890-1962), den man zurecht als **den** Statistiker des 20. Jahrhunderts bezeichnen kann, hat die statistische Welt um eine Vielzahl grundlegender Konzepte und Werkzeuge bereichert, die insbesondere für die Lösung praktischer statistischer Probleme von großer Bedeutung sind. Zu nennen wären hier z.B. Suffizienz, Fisher-Information oder die Maximum-Likelihood-Methode. Fisher befasste sich aber nicht ausschließlich mit praktischen Fragen der Statistik, sondern war auch auf der Suche nach einer allumfassenden Inferenztheorie.
Fisher glaubte, dass es eine spezielle Logik induktiver Inferenz geben müsste, die die richtigen Antworten zu jedem statistischen Problem liefern kann, ebenso, wie die uns bekannte Logik deduktive Probleme zu lösen vermag. Wenn ein Statistiker dann nach dieser speziellen Logik vorginge, müsste er z.B. auch ohne die subjektive Annahme von A-Priori-Verteilungen in der Lage sein a posteriori Wahrscheinlichkeitsaussagen über die Parameter von Merkmalsverteilungen zu machen. Fishers Idee dieser induktiven Inferenzlogik bestand hauptsächlich darin, gegebene Inferenzprobleme logisch bis auf ganz einfache Fragestellungen zu reduzieren, deren Antworten offensichtlich sind bzw. über deren Antworten ein allgemeiner Konsens herrscht. In diesem Geiste steht auch das von Fisher erdachte Konzept der Fiduzial-Inferenz.

2.1 Konzept

Fisher wollte eine Methode entwickeln, quasi A-Posteriori-Verteilungen für unbekannte Parameter zu bestimmen, ohne dabei auf subjektive A-Priori-Verteilungen zurückgreifen zu müssen. Das Grundkonzept seines Fiduzialarguments enthält drei Komponenten:

- eine suffiziente Statistik
- eine Pivotgröße
- das Fiduzialargument

Eine suffiziente Statistik ist eine Funktion der Stichprobe, welche die gesamte Information über den zu schätzenden Parameter enthält. Dies bedeutet, dass die bedingte Verteilung der Stichprobe, gegeben die suffiziente Statistik, unabhängig vom unbekannten Parameter ist. Unter dem

Begriff Pivot versteht man hier eine Funktion, die sowohl von der Statistik als auch vom unbekannten, festen Parameterwert abhängt, dessen Verteilung aber von beiden unabhängig ist und im vorliegenden Modell eindeutig bestimmt wird, also bekannt ist. Das Fiduzialargument besagt dann schließlich, dass von der Verteilung einer solchen Pivotgröße eine Verteilung des unbekannten Parameters abgeleitet werden kann, die sogenannte Fiduzialverteilung.
Zur Illustration kann das einfache Beispiel einer $N(\mu, 1)$ - verteilten Zufallsgröße X herangezogen werden. Als Pivotgröße kann hier die Variable $U = X - \mu$ gewählt werden, deren Verteilung bekannterweise die Standardnormalverteilung ist, $X - \mu = U \sim N(0, 1)$, unabhängig vom wahren Wert von μ. Wenn nun $X = 1.805$ beobachtet wird, so folgt aus der Gleichung $U = X - \mu \Leftrightarrow \mu = X - U$ als Fiduzialverteilung für μ eine $N(1.805, 1)$ - Verteilung.
Die auf diese Weise gewonnene Parameterverteilung wird allerdings nicht als klassische Wahrscheinlichkeitsverteilung verstanden, da Fisher den Parameter nicht als zufällig betrachtet. Vielmehr wird die Fiduzialverteilung als Maß der Gewissheit über den festen aber unbekannten Parameterwert interpretiert. Die Anwendung des Fiduzial-Konzeptes ist noch an eine Reihe weiterer Bedingungen geknüpft, auf die ich aber erst an späterer Stelle eingehen möchte.

2.2 Wirkung

Fishers Aufsatz *“Inverse Probability”*, in dem er sein Fiduzial-Konzept einführte, stieß bei seiner Veröffentlichung 1930 auf ein geteiltes Echo und war Auslöser für viele wissenschaftliche Kontroversen. So ging beispielsweise aus der steten Auseinandersetzung zwischen Jerzy Neyman und Fisher die Neyman-Pearson-Theorie der Konfidenzintervalle hervor. Letztere machen ähnliche Aussagen über den unbekannten Parameter, wie die Inferenzaussagen, die man mit Hilfe der Fiduzialwahrscheinlichkeiten treffen kann. Jedoch wurde dieses Konzept von Fisher vehement zurückgewiesen, wie auch umgekehrt Neyman das Fiduzialargument heftig kritisierte.

2.2.1 Fiduzialwahrscheinlichkeiten vs. Konfidenzaussagen

Latein: “fiducia” = Vertrauen, Zuversicht vs. “confidentia” = Selbstvertrauen, Dreistigkeit

Nach Fisher sind Fiduzialwahrscheinlichkeiten und Konfidenzaussagen zwei konzeptionell völlig verschiedene Dinge, was er mit der unterschiedlichen Interpretierbarkeit dieser Inferenzaussagen begründet. So können nach Beobachtung der Daten die Fiduzialverteilungen prädiktiv interpretiert werden, d.h. man kann anhand dieser Verteilungen voraussagen, wie gewiss ein bestimmtes Ereignis in einem zukünftigen Experiment eintreten würde. Konfidenzwahrscheinlichkeiten hingegen lassen nur eine postdiktive Interpretation zu, d.h. sie können nur die Frage beantworten, wie annehmbar meine Beobachtung vor dem Hintergrund des vorausgesetzten Wahrscheinlichkeitsmodelles ist. Nehmen wir wieder die $N(\mu, 1)$ - verteilte Zufallsgröße X, so lässt sich über das zweiseitige 95%-Konfidenzintervall $[X - 1.96, X + 1.96]$ nach Beobachtung von X nur sagen, dass falls μ nicht in dem Intervall liegen sollte, man ein Ereignis beobachtet hat, dessen Wahrscheinlichkeit lediglich 5% beträgt. Ein 95%-Konfidenzintervall enthält also alle möglichen

Parameterwerte, die mit 5% Fehlerwahrscheinlichkeit nicht ausgeschlossen werden. “Ein 95%-Fiduzialintervall behauptet darüber hinausgehend, dass die Wetten 19 zu 1 stehen, dass wenn der Parameterwert ermittelt würde, er tatsächlich in diesem Intervall läge.” [Dempster, 1964] Trotz dieser begrifflichen Abgrenzung gibt es große Ähnlichkeiten zwischen beiden Konzepten und in vielen Fällen stimmen die Fiduzialintervalle sogar mit den Konfidenzintervalle überein (z.B. im eindimensionalen stetigen Fall). Fisher betonte dennoch stets den Unterschied zwischen prädiktiven und postdiktiven Interpretationen, wobei er vor allem die weit verbreitete prädiktive Interpretation von Konfidenzaussagen als ungerechtfertigt ansah.

2.2.2 Kritikpunkte

Natürlich gab es auch heftige Kritik an Fishers Fiduzialkonzept, welches einige sogar als Fishers “Biggest Blunder”, also als seinen “peinlichsten Irrtum”, bezeichnen [Efron, 1998]. Die Kritiker richten sich einerseits gegen die Verwendung von Pivotgrößen an sich und andererseits gegen die vorhin erwähnten Voraussetzungen für die Anwendung der Fiduzial-Inferenz, auf die ich an dieser Stelle also zurückkomme.
Der Erfolg dieser Inferenz-Methode hängt prinzipiell stark von der Pivotgröße ab. Somit ist es für die Anwendung im konkreten Falle entscheidend, ob überhaupt eine geeignete Pivotgröße gefunden werden kann. Ebenso ist fraglich, ob die Pivotgrößen immer eindeutig konstruiert werden können, so wie Fisher sich das vorstellte. Insbesondere in mehrdimensionalen Problemstellungen besteht die Schwierigkeit, dass keine eindeutige Fiduzialverteilung bestimmt werden kann. Dies hängt eben damit zuammen, dass sich bei mehrdimensionalen Parametern verschiedene Pivotgrößen konstruieren lassen, die dann jeweils auch zu unterschiedlichen Fiduzialverteilungen führen. Des weiteren setzte Fisher unter anderem stetige Variablen oder suffiziente Statistiken voraus. Kritiker stellten aber heraus, dass diese Bedingungen in keinem logisch zwingenden Zusammenhang mit dem Kern des Konzeptes stehen. Dieser kann in der Idee gesehen werden, die Fiduzialverteilung des Parameters von einer bekannten, durch die Beobachtungen nicht veränderlichen Verteilung abzuleiten, eben von der Verteilung der Pivotgröße. In diesem Zusammenhang zeigte Arthur P. Dempster beispielsweise die Unnötigkeit der Suffizienzanforderung [Dempster, 1964]. Weiterhin fanden andere Statistiker heraus, dass die übrigen genannten Bedingungen ebenfalls überflüssig sind und enwickelten Adaptionen des Fiduzialkonzeptes für diskrete Variablen oder sogar ganz ohne Pivotgrößen [Heike, C. Târcolea, Demetrescu, A. Târcolea, 2003]. Die logische Inkonsistenz stellte demnach den stärksten Kritikpunkt an diesem Inferenzkonzept dar, der aber gepaart mit dem Charme der zentralen Idee einigen Statistikern gerade als Motivation zur Verbesserung oder Weiterentwicklung des Grundkonzeptes diente.
Somit konnte Fisher mit seinem Fiduzialargument zwar nicht sein ambitionierte Ziel einer allumfassenden Inferenztheorie erreichen, doch bot er damit einigen anderen Statistikern einen durchaus fruchtbaren Anknüpfungspunkt für die Entwicklung ihrer eigenen Theorien, so z.B. Arthur P. Dempster.

Kapitel 3

Arthur P. Dempster

Arthur P. Dempster ist heute etwa 75 Jahre alt und noch immer als Forschungsprofessor für theoretische Statistik in Harvard aktiv. Seine ersten bedeutenden Artikel veröffentlichte er wenige Jahre nach Fishers Tod. In diesen ersten Artikeln von Mitte der 1960er Jahre setzt sich Dempster kritisch mit Fishers Theorie der Fiduzialwahrscheinlichkeiten auseinander. Bekannt wird er aber später vor allem durch seine 1977 veröffentlichte Einführung des EM-Algorithmus ("Expectation Maximization" - Algorithmus). Dies ist ein Algorithmus zur ML-Schätzung im Falle fehlender Daten, ein heute beliebtes praktisches Instrument, was bei Dempsters sonst eher theoretischen Ausrichtung doch ein wenig überrascht. Begeistert von Fishers grundlegender Idee und motiviert durch die Schwachstellen dessen Konzeptes, machte es sich Dempster aber zunächst zur Aufgabe, das Fiduzialargument von seiner inkonsistenten Basis zu befreien und von grundlegenderen Prinzipien ausgehend neu zu formulieren.

3.1 Dempsters Inferenzkonzept

Das von Dempster entwickelte Inferenzmodell haben wir bereits als Bestandteil der Dempster-Shafer-Theorie zu Beginn des Seminars kennengelernt. Ich möchte es an dieser Stelle noch einmal aus dem wahrscheinlichkeitstheoretischen Blickwinkel heraus erläutern und dann insbesondere auf die dahinterstehende Theorie eingehen.

3.1.1 Das Grundmodell - statische Komponente / vor Beobachtung

Man nehme zwei Räume $\mathbf{X}$ und $\mathbf{S}$ sowie eine mehrwertige Abbildung:

$$\Gamma : \mathbf{X} \to \mathbf{S}, \qquad x \mapsto \Gamma(x) \subset \mathbf{S}.$$

Ferner sei auf dem Raum $\mathbf{X}$ ein klassisches Wahrscheinlichkeitsmaß μ definiert. In Dempsters Inferenzmodell steht der Raum $\mathbf{X}$ in der Regel für die Grundgesamtheit der Untersuchungseinheiten, während der Raum $\mathbf{S}$ die möglichen Ausprägungen oder Kategorien des interessierenden Merkmals enthält. Als klassisches Wahrscheinlichkeitsmaß μ kann eine Gleichverteilung auf $\mathbf{X}$

angenommen werden, da bei einer reinen Zufallsstichprobe jedes Individuum die gleiche Wahrscheinlichkeit hat, ausgewählt zu werden. Wird die Sichprobe allerdings anders gezogen, kann μ entsprechend auch eine andere Verteilung sein.

Nun möchte man schon vor der tatsächlichen Beobachtung von Daten Wahrscheinlichkeitsaussagen über beliebige Teilmengen $\mathrm{T} \subset \mathbf{S}$ treffen. Wäre Γ eine eindeutige Abbildung, so würden sich die Wahrscheinlichkeiten für Teilmengen von $\mathbf{S}$ einfach als Summe der Wahrscheinlichkeiten der zugehörigen Elemente $x \in \mathbf{X}$ mit $\Gamma(x) \in \mathrm{T}$ ergeben (vgl. dazu Zufallsgrößen und die von ihnen induzierten Wahrscheinlichkeiten, die sich als Summe der Wahrscheinlichkeiten der zugehörigen Elemente der Grundgesamtheit ergeben). Da wir es hier aber mit einer mehrwertigen Abbildung zu tun haben, d.h. einer Untersuchungseinheit aus der Population eine ganze Teilmenge von Elementen des Bildraumes $\mathbf{S}$ zugeordnet werden kann, ist die Rückführung einer Teilmenge $\mathrm{T} \subset \mathbf{S}$ auf die konstituierenden Elemente des Grundraumes $\mathbf{X}$ nicht mehr eindeutig möglich. Alternativ können aber eine "Mindestmenge" T_* und eine "Maximalmenge" T^* der von T betroffenen Populationseinheiten bestimmt werden, die dann zu den bekannten Intervallwahrscheinlichkeiten führen. Man definiert also:

$$T_* := \{x \in \mathbf{X} \mid \Gamma(x) \subset \mathrm{T},\ \Gamma(x) \neq \emptyset\},$$

$$T^* := \{x \in \mathbf{X} \mid \Gamma(x) \cap \mathrm{T} \neq \emptyset\}.$$

Daraus ergeben sich die obere und untere Wahrscheinlichkeit für eine Ereignismenge $\mathrm{T} \subset \mathbf{S}$:

$$P_*(\mathrm{T}) = \frac{\mu(T_*)}{\mu(S^*)}, \quad P^*(\mathrm{T}) = \frac{\mu(T^*)}{\mu(S^*)},$$

wobei S^* alle Populationsindividuen umfasst, deren Bild $\Gamma(x)$ in $\mathbf{S}$ nicht die leere Menge ist, d.h. $S^* = \mathbf{X} \setminus \{x \in \mathbf{X} \mid \Gamma(x) = \emptyset\}$. Hier ist logischerweise $S^* = S_*$. Die Intervallwahrscheinlichkeiten haben die bekannten Eigenschaften:

$$0 \leq P_*(\mathrm{T}) \leq P^*(\mathrm{T}) \leq 1,$$

$$P^*(\mathrm{T}) = 1 - P_*(\bar{\mathrm{T}}).$$

Im eindimensionalen Falle können mit diesen Wahrscheinlichkeiten sogar eine obere und eine untere Verteilungsfunktion, sowie ein oberer und ein unterer Erwartungswert definiert werden:

$$H_*(t) = P_*(\mathrm{T} \leq t), \qquad H^*(t) = P^*(\mathrm{T} \leq t),$$

$$E_*(\mathrm{T}) = \int_{-\infty}^{\infty} t\, dH^*(t), \qquad E^*(\mathrm{T}) = \int_{-\infty}^{\infty} t\, dH_*(t).$$

Im Folgenden möchte ich den unscharfen Maßtransport von $\mathbf{X}$ nach $\mathbf{S}$ anhand eines Beispiels noch etwas veranschaulichen.

3.1.2 Beispiel: Erster Teil

Da wir im Zusammenhang mit Dempsters Modell schon wiederholt politische Parteien und das Wahlverhalten als Beispiel herangezogen haben, möchte ich das auch hier tun.
Stellen wir uns vor, dass morgen Bundestagswahlen stattfinden, wir aber schon heute eine Vorstellung davon bekommen möchten, wie die Parteien dabei abschneiden werden. Wir konzentrieren uns hier auf die Zweitstimme, da diese direkt für die Partei und nicht für einen bestimmten Kandidaten abgegeben wird. Sei **X** der Raum der Wahlberechtigten (ca. 62 Mio. waren es im Jahre 2005) und stellen wir uns weiterhin vor, es stünden nur vier Parteien zur Wahl, nämlich **S**={ SPD, CDU, B90/G, FDP }. Da einige Wähler noch unentschlossen sind, ergeben sich folgende theoretisch mögliche Teilmengen $\Gamma(x) \subset \mathbf{S}$:

SPD	SPD, CDU	SPD, B90/G	CDU, B90/G, FDP
CDU	CDU, B90/G	SPD, FDP	SPD, B90/G, FDP
B90/G	CDU, FDP	SPD, CDU, B90/G	$\emptyset$
FDP	FDP, B90/G	SPD, CDU, FDP	**S**

Möchte man nun die Wahrscheinlichkeit für das Ereignis T $\subset$ **S** angeben, dass ein zufällig ausgewählter Wahlberechtigter die Partei B90/Grüne wählen würde, so ergeben sich folgende Mengen und Intervallwahrscheinlichkeiten:

$$T_* := \{x \in \mathbf{X} \mid \Gamma(x) = B90/G\},$$

T_* umfasst also alle Wahlberechtigten, die fest entschlossen sind, für die B90/Grünen zu stimmen.

$$T^* := \{x \in \mathbf{X} \mid \Gamma(x) \in \{B90/G;\ CDU, B90/G;\ SPD, B90/G;\ FDP, B90/G;\ SPD, CDU, B90/G;$$

$$CDU, B90/G, FDP;\ SPD, B90/G, FDP;\ \mathbf{S}\}\},$$

d.h. T^* enthält all jene, die den B90/Grünen sicher oder vielleicht ihre Stimme geben würden.
Die dazugehörigen obere und untere Wahrscheinlichkeit sind dann wieder:

$$P_*(\mathrm{T}) = \frac{\mu(T_*)}{\mu(S^*)}, \qquad P^*(\mathrm{T}) = \frac{\mu(T^*)}{\mu(S^*)}.$$

S^* umfasst hier all diejenigen Wahlberechtigten, die von Ihrem Wahlrecht tatsächlich Gebrauch machen und auch keine ungültige Stimme abgeben wollen. Da nun μ die Gleichverteilung über den Raum der Wahlberechtigten ist, reduziert sich die Berechnung dieser Wahrscheinlichkeiten auf Zählen der Mengenelemente in T_*, T^* sowie S^* und einfaches Dividieren.

3.1.3 Bedingungs- und Kombinationsregel - dynamische Komponente / nach Beobachtung

Nachdem man dann Merkmalswerte an zufällig ausgewählten Untersuchungseinheiten beobachtet hat, werden die Wahrscheinlichkeitsaussagen entsprechend auf die Beobachtung bedingt. Die beobachteten Werte werden nämlich als zusätzliche Information über die Merkmalsverteilung aufgefasst, die als solche bei den Wahrscheinlichkeitsaussagen berücksichtigt werden sollten. Die Menge der beobachteten Merkmalsausprägungen wird hier mit $\mathrm{R} \subseteq \mathbf{S}$ bezeichnet. Dempsters Bedingungsregel hatten wir unter dem Namen "D-Conditioning"-Regel ebenfalls schon kennengelernt. Die Bedingung geschieht durch eine Beschränkung der Zuordnungsvorschrift Γ auf R, d.h. $\Gamma'(x) = \Gamma(x) \cap \mathrm{R}$, sodass nur noch jene Merkmalsausprägungen berücksichtigt werden, die auch tatsächlich beobachtet wurden. Es ergeben sich damit die bedingten oberen und unteren Wahrscheinlichkeiten:

$$P_*(\mathrm{T} \mid \mathrm{R}) = 1 - P^*(\bar{\mathrm{T}} \mid \mathrm{R}) = 1 - \frac{\mu(\bar{T}R^*)}{\mu(R^*)}, \qquad P^*(\mathrm{T} \mid \mathrm{R}) = \frac{P^*(\mathrm{T} \cap \mathrm{R})}{P^*(\mathrm{R})} = \frac{\mu(TR^*)}{\mu(R^*)}.$$

Die zugehörigen Mengen von Populationsindividuen sehen folgendermaßen aus:

$$\bar{T}R^* = \{x \in \mathbf{X} \mid \Gamma(x) \cap \mathrm{R} \cap \bar{\mathrm{T}} \neq \emptyset\}, \qquad TR^* = \{x \in \mathbf{X} \mid \Gamma(x) \cap \mathrm{R} \cap \mathrm{T} \neq \emptyset\},$$

$$R^* = \{x \in \mathbf{X} \mid \Gamma(x) \cap \mathrm{R} \neq \emptyset\}.$$

Wie man sieht bedingt Dempster hier mit der oberen Wahrscheinlichkeit der Beobachtung $P^*(\mathrm{R})$, was allerdings nicht unumstritten ist. Würde man dagegen mit der unteren Wahrscheinlichkeit $P_*(\mathrm{R})$ bedingen, so würde die bedingte obere Wahrscheinlichkeit von T größer und die untere kleiner. Das Intervall wäre also breiter und damit eigentlich vorsichtiger. Warum Dempster dennoch die obere Wahrscheinlichkeit nimmt, wird von ihm leider nicht explizit begründet.
Hat man nun schließlich mehrere unabhängige Informationsquellen, wie z.B. zwei nicht-überlappende Zufallsstichproben aus der Grundgesamtheit, so kann man deren Informationen synthetisieren, um noch bessere Wahrscheinlichkeitsaussagen machen zu können. Die von Dempster dazu verwendete Kombinationsregel kann als Verallgemeinerung der Bedingungsregel bezeichnet werden. Dabei wird dann einfach ein weiteres Mal, eben auf die Beobachtung der zweiten Stichprobe, bedingt, d.h. $\Gamma''(x) = \Gamma_1'(x) \cap \Gamma_2'(x)$. Im Prinzip stellt bereits eine Stichprobe vom Umfang n eine n-fache Kombination der Informationen aus den einzelnen Beobachtungen dar. Auf die gleiche Art und Weise kann man dann auch andere objektive Information über die Merkmalsverteilung und ihre Parameter in das Modell integrieren. Da sich die Verteilungsparameter ja in den Beobachtungen niederschlagen, kann ich auf die Teilmenge von Merkmalsauspägungen bedingen, die im Einklang mit dem Vorwissen über die Verteilungsparameter stehen bzw. die aufgrund des Wissensstandes nicht ausgeschlossen werden können. Dabei gilt für die Kombination von Wissen aus verschiedenen Informationsquellen die Eigenschaft, dass die Kombination einer Quelle $\Gamma_1' = \Gamma_1 \cap \mathrm{R}_1$ mit einer weniger informativen, d.h. $\mathrm{R}_2 \subset \mathrm{R}_1$, keinen Erkenntnisgewinn bringt. In diesem Falle sind die kombinierten bedingten Wahrscheinlichkeiten identisch mit den auf die erste Information bedingten.

3.1.4 Beispiel: Zweiter Teil

Wenn wir in unserem Beispiel also 100 Leute am Vorabend der Wahl befragen, wie sie am kommenden Tag wählen würden und niemand würde in Erwägung ziehen, der FDP seine Stimme zu geben. So bedingen wir unsere Intervallwahrscheinlichkeiten nach der oben stehenden Bedingungsregel von Dempster auf diese Beobachtung $\mathbf{S} \supseteq \mathrm{R} = \{SPD, CDU, B90/G\}$. Dies bedeutet, dass alle Zuordnungen $\Gamma(x)$, welche die FDP enthalten, jetzt als theoretisch "unmöglich" wegfallen. Die Tabelle der möglichen Beobachtungen reduziert sich auf die folgende:

SPD	SPD, CDU	SPD, CDU, B90/G
CDU	CDU, B90/G	
B90/G	SPD, B90/G	

Damit ergeben sich die auf die Beobachtung R bedingten Wahrscheinlichkeiten für unser interessierendes Ereignis $\mathrm{T} \subset \mathbf{S}$, dass jemand die Grünen wählt, entsprechend der Formeln:

$$P_*(\mathrm{T} \mid \mathrm{R}) = 1 - P^*(\bar{\mathrm{T}} \mid \mathrm{R}) = 1 - \frac{\mu(\bar{T}R^*)}{\mu(R^*)}, \qquad P^*(\mathrm{T} \mid \mathrm{R}) = \frac{P^*(\mathrm{T} \cap \mathrm{R})}{P^*(\mathrm{R})} = \frac{\mu(TR^*)}{\mu(R^*)}.$$

Die zugehörigen Mengen von betroffenen Wahlberechtigten sind hier folgende:

$$\bar{T}R^* = \{x \in \mathbf{X} \mid \Gamma(x) \cap \{SPD, CDU, B90/G\} \cap \{SPD, CDU\} \neq \emptyset\},$$

$$TR^* = \{x \in \mathbf{X} \mid \Gamma(x) \cap \{SPD, CDU, B90/G\} \cap \{B90/G\} \neq \emptyset\},$$

$$R^* = \{x \in \mathbf{X} \mid \Gamma(x) \cap \{SPD, CDU, B90/G\} \neq \emptyset\}.$$

Nun befrage ich in einem zweiten Schritt noch weitere 50 Leute nach ihren möglichen Wahlentscheidungen, von denen die meisten zwischen CDU und den B90/Grünen schwanken, einige aber auch zwischen FDP und SPD. Kombiniere ich nun die Informationen aus beiden unabhängigen Stichproben, so bleibt als Menge der möglichen Merkmalsausprägungen:

$$\mathrm{R}_{kombi} = \mathrm{R}_1 \cap \mathrm{R}_2 = \{SPD, CDU, B90/G\} \cap \{SPD, CDU, B90/G, FDP\} = \mathrm{R}_1.$$

Hier zeigt sich die oben beschriebene Eigenschaft, dass sich die oberen und unteren Wahrscheinlichkeiten sich nicht verändern, wenn man eine Informationsquelle R_1 mit einer weiteren R_2 kombiniert, die keine zusätzliche Information in Form von auszuschließenden Merkmalswerten enthält.

3.2 Dempsters Inferenztheorie

Nachdem ich die technischen Bestandteile von Dempsters Modell dargestellt habe, möchte ich nun auf seine Inferenztheorie, d.h. die dahinter stehende Philosophie eingehen. Prinzipiell erkennt Dempster die subjektiven Aspekte wissenschaftlichen Arbeitens an und wendet sich gegen einen Pseudo-Objektivismus, der die Augen vor diesen verschließt. Die Ergebnisse einer wissenschaftlichen Arbeit hängen nämlich wesentlich von der Wahl eines geeigneten Modells ab, was aber ein individueller und häufig informell ablaufender Entscheidungsprozess ist. Da im Sinne der Wissenschaftlichkeit dennoch ein möglichst hoher Grad an Objektivität bzw. intersubjektiver Einigkeit anzustreben ist, plädiert er dafür, alle Arbeitsschritte möglichst transparent zu machen und die subjektiven Komponenten so weit wie möglich zu reduzieren. Dempster hat daher ein Inferenzmodell entwickelt, das ausschließlich objektive Informationen verarbeitet, die aber sowohl die Form von Verteilungsannahmen, A-Priori-Verteilungen als auch von Beobachtungen haben können, insofern sie objektiv begründbar sind. Mit Hilfe der zu Anfang eingeführten Charakteristika lässt sich Dempsters Weltsicht wiefolgt darstellen.

- *Wahrscheinlichkeitsbegriff*: Eher subjektiv, d.h. Wahrscheinlichkeiten drücken unvollständiges Wissen aus.

- *Beobachtungen*: Beobachtet werden Ausprägungen eines Merkmals, die den Individuen einer Population zugeordnet werden können. Dabei geht man zwar prinzipiell davon aus, dass das Merkmal irgendeine (parametrische) Verteilung in der Population hat, jedoch wird davon nur dann eine Verteilungsannahme für die Stichprobe abgeleitet, wenn man diese auf objektive Informationen stützen kann. Ist dies nicht möglich, so nimmt man an, dass die Verteilungsparameter sich in der Zuordnungsvorschrift Γ und damit in den Beobachtungen niederschlagen.

- *Rückschlüsse*: Es werden ausschließlich objektiv begründbare Informationen verwendet, welche alle zu einer finalen Inferenz synthetisiert werden. Aussagen über die Merkmalsverteilung in der Population oder in zukünftigen Beobachtung sind allerdings nur in Form von oberen und unteren Wahrscheinlichkeiten möglich.

Somit hebt sich Dempsters Modell doch deutlich von Fishers Fiduzialargument ab, was ja ursprünglich als Ausgangspunkt diente. Zwar werden in beiden Modellen Inferenzaussagen grundsätzlich von den Beobachtungen abgeleitet und auf offensichtliche Wahrscheinlichkeiten zurückgeführt, doch führt dies im Falle Fishers zu einer Parameterverteilung, in die neben Verteilungssannahme, Pivotgröße und Beobachtung keine weiteren Informationen aufgenommen werden können. Insofern ist Dempsters Modell weniger eingeschränkt, da nicht zwingend eine Verteilungsannahme getroffen, keine Pivotgröße konstruiert werden muss und jegliche zusätzliche Information über die Verteilung des interessierenden Merkmals integriert werden kann. Weiterhin ist Dempsters Inferenztheorie ein ganzes Stück näher an der Bayes-Inferenz als Fishers Konzept der Fiduzialwahrscheinlichkeiten. Der mathematische Rahmen von Dempsters Modell ist aber so weit gefasst, dass es sowohl Fishers Fiduzialmodell als auch die klassische Bayes-Inferenz als Spezialfälle umfasst.

3.3 Die verallgemeinerte Bayes-Inferenz

Da Dempsters Inferenzmodell die klassische Bayes-Situation als Spezialfall enthält, kann es auch als Verallgemeinerung der Bayes-Inferenz betrachtet werden. Weitere Versuche die Bayes-Theorie zu verallgemeinern hatten wir ja schon an anderer Stelle kennengelernt, nämlich in den Vorträgen zum Imprecise Dirichlet Modell und zur robusten Bayesanalyse. Bayes-Inferenz funktioniert im Prinzip ja so, dass man mit einer globalen Verteilung für alle relevanten Variablen beginnt, dann einige dieser Variablen beobachtet und schließlich die auf diese Beobachtungen bedingte Verteilung der übrigen Variablen berechnet, mit Hilfe derer dann Inferenz-Aussagen getroffen werden. In Dempsters Modell wird dagegen keine gemeinsame Ausgangsverteilung gefordert. Wenn wir allerdings als Ausgangspunkt eine Intervallwahrscheinlichkeit von [0,1] für alle möglichen Merkmalsausprägungen unterstellen, die dann durch Bedingen auf Beobachtungen und ein Vorwissen über die Verteilungsparameter eingeschränkt wird, erhalten wir eine A-Posteriori-Verteilung im bayesianischen Sinne. Demnach kann man die klassische Bayes-Situation auch in Dempsters Inferenzkonzept modellieren. Es kann in mehrer Hinsicht als eine Verallgemeinerung der Bayes-Inferenz angesehen werden. Zunächst können in Dempsters Modell sowohl mit als auch ohne A-Priori-Verteilungen A-Posteriori-Wahrscheinlichkeiten für ein Merkmal oder seine Verteilungsparameter bestimmt werden. Des weiteren ist in Dempsters Modell eine dem Bayes-Lernen ähnliche fortwährende Integration von zusätzlichen Informationen über die Merkmalsverteilung möglich. Die Annahme einer [0,1]-Intervallwahrscheinlichkeit als Ausgangspunkt kann weiterhin als Verallgemeinerung der Laplaceschen Gleichverteilung als A-Priori-Verteilung angesehen werden. Die [0,1]-Verteilung kann in Dempsters Modell das A-Priori-Nichtwissen um die Merkmalsverteilung ausdrücken, hat aber im Gegensatz zur Gleichverteilung in der klassischen Bayes-Inferenz keinen Einfluss auf die A-Posteriori-Wahrscheinlichkeitsaussagen. Schließlich wurde das charakteristische bayesianische Instrument der Bedingung auf die Beobachtung übernommen, weshalb auch Dempster selbst sein Inferenzkonzept eher der bayesianischen Theorierichtung zuordnen würde [Dempster, 1968].

Kapitel 4

Modellierung unsicheren Wissens

Grundlegend geht es bei allen Inferenzkonzepten darum, das Nichtwissen über den Fehler des induktiven Schließens zu modellieren. Könnten wir nämlich einfach direkt von einer Beobachtung auf eine Allgemeinaussage schließen, könnten wir uns die gesamte Inferenztheorie ersparen. Da dies nun aber nicht unmöglich ist, brauchen wir Wahrscheinlichkeiten, um unsere Unsicherheit darüber ausdrücken zu können, ob unsere Beobachtungen die wirklichen Verhältnisse in der Welt widerspiegeln. Die ideale Lösung diese grundlegende Unsicherheit des induktiven Schlusses zu modellieren, wären sogenannte logische Wahrscheinlichkeiten. Dies sind Wahrscheinlichkeiten, die nicht den möglichen Ereignissen zugeordnet werden, sondern den Inferenzaussagen selbst. Das würde im konkreten Falle bedeuten, dass ich mit einer Wahrscheinlichkeit P von einer Beobachtung A auf eine Aussage B schließen kann. Dazu entwickelte Kurt Weichselberger ein Konzept, in dem versucht wird mit Intervallwahrscheinlichkeiten, den Fehler der induktiven Schlüsse zu quantifizieren. Da Weichselberger seine Ideen aber sehr anspruchsvoll formalisiert, möchte ich für eine detaillertere Darstellung seines Konzeptes auf die Literatur verweisen. Man könnte also zum einen sagen, dass in der Statistik prinzipiell unsicheres Wissen modelliert wird. Eine andere Ebene der Modellierung von Unsicherheit ist dann, wie innerhalb eines jeden Inferenzkonzeptes mit unvollständigen oder unscharfen Beobachtungen umgegangen wird.

Welche Inferenz-Theorie nun die grundlegende Unsicherheit, sowie unvollständige oder unscharfe Beobachtungen am besten modelliert, darüber kann man streiten. Bayesianische Verfahren scheinen dafür in Zukunft immer bedeutender zu werden, was mit der besonderen Möglichkeit zusammenhängt, dass Informationen aus verschiedenen Quellen integriert werden können. Vor dem Hintergrund der wachsenden Komplexität wissenschaftlicher Probleme ist dies eine durchaus wünschenswerte Eigenschaft [Efron, 1998]. Dempsters Inferenzkonzept wäre möglicherweise ein sehr gut geeignetes Modell. Es ist formal so offen gefasst, dass es die anderen Inferenz-Theorien als Spezialfälle enthält und sie somit quasi zu einem globalen Konzept vereint. Des weiteren werden nach seiner Theorie auch ungenauen Messungen eigene Wahrscheinlichkeiten zugeordnet, die in den klassischen Inferenz-Modellen nicht berücksichtigt würden. Vielleicht könnte man gar behaupten, dass Dempster gelungen ist, was Fisher nicht schaffte, nämlich eine allumfassende Theorie statistischer Inferenz zu formulieren.

Literaturverzeichnis

[Dempster, 1964] Dempster, Arthur P., 1964, *On the difficulties inherent in Fisher's fiducial argument*, Journal of the American Statistical Association, Vol. 59, No. 305, 56-66.

[Dempster, 1966] Dempster, Arthur P., 1966, *New methods for reasoning towards posterior distributions based on sample data*, Annals of Mathematical Statistics, Vol. 37, 355-374.

[Dempster, 1967a] Dempster, Arthur P., 1967, *Upper and lower probabilities induced by a multivalued mapping*, Annals of Mathematical Statistics, Vol. 38, 325-339.

[Dempster, 1967b] Dempster, Arthur P., 1967, *Upper and lower probability inferences based on a sample from a finite univariate population*, Biometrika, Vol. 54, 515-528.

[Dempster, 1968] Dempster, Arthur P., 1968, *A generalization of Bayesian Inference, Journal of the Royal Statistical Society*, Series B, Vol. 30, 205-247.

[Dempster, 1998] Dempster, Arthur P., 1998, *Logicist Statistics I. Models and Modeling*, Statistical Science, Vol. 13, 248-276.

[Efron, 1998] Efron, Bradley, 1998, *R.A. Fisher in the 21st Century*, Statistical Science, Vol. 13, No. 2, 95-114.

[de Finetti, 1974] de Finetti, 1974, *Theory of Probability*, New York: Wiley, Preface.

[Heike, C. Târcolea, Demetrescu, A. Târcolea, 2003] Heike, H.-D., Târcolea, C., Demetrescu, M. und Târcolea, A. , 2003, *Fiducial inference for discrete and continuous distributions*, BSG Proceedings, Geometry Balkan Press, 69-80.

[Popper, 1934] Popper, Karl R., 1934, *Die wissenschaftliche Methode* in: Miller, David (Hrsg.), 1995, *Lesebuch: ausgewählte Texte zu Erkenntnistheorie, Philosophie der Naturwissenenschaften, Metaphysik, Sozialphilosophie / Karl R. Popper*, Tübingen: Mohr, 118-127.

[Rüger, 1999] Rüger, Bernhard, 1999, *Test- und Schätztheorie, Band 1: Grundlagen*, München: Oldenbourg Verlag, Kap. 2: Inferenzkonzepte.

[Weichselberger, 2001] Weichselberger, Kurt, 2001, *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitstheorie I. Intervallwahrscheinlichkeit als umfassendes Konzept*, Heidelberg: Physica-Verlag, Kap. 1.3: Entstehung der Theorie.