Statistische Methodik im Wandel der Zeit

A.o. Univ.-Prof. Dr. Marcus Hudec, Universität Wien – Fakultät für Informatik

Jubiläen bilden traditionell einen willkommenen Anlass zu einem Rückblick über historische Entwicklungen. Dabei steht man vor der Herausforderung in knapper Redezeit essentielle Entwicklungen und Weichenstellungen im Verlaufe einer langjährigen Entwicklungsgeschichte aufzuzeigen und im Kontext der heutigen Zeit transparent zu machen. Das Ansinnen vor dem Hintergrund 150 Jahre amtliche Statistik in Wien auch nur ansatzweise einen vollständigen Überblick über den Wandel in der statistischen Methodik in dieser Zeitspanne wiedergeben zu wollen, muss von Grund auf zum Scheitern verurteilt sein. In diesem Sinne wollen wir in diesem Vortrag kursorische Streiflichter auf einige wenige aus der subjektiven Sicht des Referenten bedeutsame, die Arbeitsweisen der amtlichen Statistik massiv verändernde methodische Innovationen und Paradigmenwechsel werfen und auch einen Ausblick auf künftige Entwicklungen wagen.

Historische Wurzeln

Die statistische Methodenlehre entspringt aus zwei vollkommen unterschiedlichen Quellen. Zum einen führte die Beschäftigung mit Fragen der Chancen bei Glücksspielen zur Entwicklung der Wahrscheinlichkeitstheorie, welche die Grundlage der modernen Inferenzstatistik bildet, zum anderen führte die Notwendigkeit der quantitativen Beschreibung von Staaten von alters her zur Entwicklung deskriptiver Methoden der Statistik. Dieser zweite Ansatz ist zumindest für die Entstehung der amtlichen Statistik naturgemäß von primärer Bedeutung und kann ihrerseits durch zwei Hauptströme charakterisiert werden. Die unter dem Namen „politische Arithmetik“ subsummierten Ansätze der Bevölkerungsstatistik, die auf den Arbeiten von John Graunt (1620–1674), Edmond Halley (1656–1742) und Johann Peter Süßmilch (1707–1767) basiert, und der Staatsbeschreibung wie sie etwa von Gottfried Achenwalls (1719–1772) zur Zustandsbeschreibung (status › statistique) des Staates entwickelt wurde.

Entwicklung stichprobenbasierter Statistik

Für die historischen Vorläufer der amtlichen Statistik war die Vollerhebung von Populationen eine conditio sine qua non. Der Siegeszug der Statistik im 20. Jahrhundert, die sich in einer immer weiteren Nachfrage nach verlässlichen statistischen Befunden für Entscheidungsträger widerspiegelt, ist eng verknüpft mit der Verschmelzung der beiden ursprünglich isolierten Wurzeln der Statistik, welche einen echten Paradigmenwechsel im Sinne von Kuhn auslöste und der Stichprobentechnik den Einzug in die statistischen Ämter ermöglichte.

Der erste wichtige Proponent für diesen Paradigmenwechsel in den der amtlichen Statistik war der Norweger Anders Nicolai Kiær (1838–1919), der 1895 beim ISI-Meeting in Bern erstmals ein Framework statistischer Methodik vorstellte, in welchem die Stichprobentechnik zu einem zentralen Instrument wurde. Kiærs Hauptansatz war dabei die Vorstellung, dass bei Teilerhebungen mittels Miniaturisierung einer Population im Sinne einer repräsentativen Stichprobe wertvolle Ergebnisse gewonnen werden können. Motivation war dabei, dass der Effizienzgewinn durch Teilerhebungen es ermöglicht, das Spektrum statistischer Untersuchungen deutlich auszuweiten.

Wie häufig in der Wissenschaftsgeschichte waren die Reaktionen der Kollegenschaft auf diese innovative Idee zunächst durchgehend negativ. War doch die herrschende Lehrmeinung – geprägt von dem damals die amtliche Statistik dominierenden, aus der Buchhaltung übernommenen Streben nach Vollständigkeit und Korrektheit der Bücher –, dass Teilerhebungen niemals Vollerhebungen ersetzen könnten. Dennoch vertrat Kiær diesen Ansatz bei den folgenden ISI-Mee-

tings und fand in Caroll D. Wright (Gründer des U.S. Bureau of Labor Statistics) und später auch in Arthur L. Bowley erste Unterstützer.

Eine wesentliche Bereicherung der Ideen von Kiær lieferte 1903 der eher unbekannte französische Statistiker Lucien March, der in der Diskussion zu Kiærs Paper beim ISI-Meeting 1903 in Berlin erstmals Ideen zur randomisierten Stichprobenziehung einbrachte. Mit diesem Vorschlag, der wissenschaftshistorisch als Meilenstein zu werten ist, kam es erstmals zum Brückenschluss zwischen der mathematisch fundierten Wahrscheinlichkeitstheorie und der amtlichen Statistik. Die Verschmelzung dieser beiden Ideenwelten erwies sich für die Entwicklung der Methodologie der gesamten Statistik im 20. Jahrhundert als extrem fruchtbar.

Die Idee das Kalkül des Zufalls für repräsentative Schlüsse über Grundgesamtheiten heranzuziehen, war zum damaligen Zeitpunkt insofern revolutionär, als das Prinzip des Zufalls damals bei Weitem nicht jenen Raum in der Wissenschaftstheorie einnahm wie heute. Erst viel später verhalfen die bahnbrechenden Erkenntnisse der Physik (Quantenmechanik, Thermodynamik) sowie der modernen Evolutionsforschung zu einer breiten Akzeptanz stochastischer Gesetzmäßigkeiten. Dass der Gedanke randomisierter Stichproben zu Beginn des 20. Jahrhunderts bei den arrivierten Statistikern der damaligen Zeit auf Ablehnung stieß, ist nicht verwunderlich, bedenkt man, dass selbst Einstein noch 1926 in einem Brief an Max Born seine Skepsis gegenüber stochastischen Erklärungsmodellen in der Physik mit der berühmten Aussage *„Gott würfelt nicht“* Ausdruck verlieh.

Die unseres Wissens historisch erste empirische Anwendung einer randomisierten Stichprobe war die sozialstatistische Studie über Armut in Reading im Jahre 1912, welche der britische Statistiker Arthur L. Bowley durchgeführt hat.

Bowleys Arbeiten führten letztlich dazu, dass sich das Internationale Statistische Institut 1924 dazu durchrang, eine Kommission einzusetzen, die sich ernsthaft mit dem Anwendungspotential von Stichprobentechniken auseinandersetzen sollte. Das beim ISI-Meeting in Rom 1925 veröffentlichte Ergebnis führte zu einer Resolution, welche die Akzeptanz von Teilerhebungen beinhaltete, wobei sowohl bewusste als auch randomisierte Auswahlverfahren inkludiert waren.

Im folgenden Jahrzehnt wurden Teilerhebungen auf Basis bewusster Auswahlverfahren in Dänemark (Jensen) und Italien (Gini) durchgeführt. Neben Bowley vertrat vor allem Margaret Hogg in den USA die Ansicht, dass nur randomisierte Stichproben einen Test auf Repräsentativität ermöglichen.

Der Siegeszug randomisierter Stichprobentechniken

Eine neue Ära läutete die bahnbrechende Arbeit von Jerzy Neyman „On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection” 1934 ein.

Neyman konnte in seinem Beitrag an Hand von praktischen Beispielen und mittels theoretischer Überlegungen schlüssig darlegen, dass durch Randomisierung im Vergleich zur bewussten Auswahl die Herausforderungen, mit denen Statistiker bei der Durchführung von Teilerhebungen konfrontiert waren, wesentlich besser bewältigt werden konnten. Für die weitere Entwicklung statistischer Methodik von entscheidender Bedeutung waren zwei theoretische Ergebnisse Neymans. Zum einen entwickelte er theoretische Konzepte, welche die Technik der Schätzung von Konfidenzintervallen begründete. Zum anderen durchbrach er mit seinen Ergebnissen die vorherrschende Meinung repräsentativer Stichproben, welche davon ausging, dass valide Stichprobenergebnisse nur dann möglich sind, wenn alle Elemente der Grundgesamtheit mit gleicher Inklusionswahrscheinlichkeit in die Stichprobe gelangen können. Seine Ergebnisse über geschichtete Stichprobenverfahren bilde-

Definition of Statistics: The science of producing unreliable facts from reliable figures.

EVAN ESAR

...while a man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.

SHERLOCK HOLMES

ten die Grundlage für die Entwicklung designbasierter Stichprobenverfahren.

Die Diffusion von Neymans Ideen wurde dadurch befruchtet, als W. Edwards Demming Neyman nach Washington DC einlud, wo er einen großen Einfluss auf die Entwicklung der amtlichen Statistik in den USA und somit indirekt weltweit ausübte. In einem historischen Rückblick begründen Duncan & Shelton in ihrem Buch „Revolution in United States Government Statistics" (1978) den großen Einfluss Neymans auf die Entwicklung der amtlichen Statistik mit *„Neyman's ability to relate theory to practice from personal experience in the economic and social fields ..."*.

Brückenschlag zur theoretischen Inferenzstatistik

Wenn wir bislang hauptsächlich über die historische Entwicklung der stichprobenbasierten Methodik im Bereich der amtlichen Statistik reflektiert haben, so geschah dies primär deshalb, da sich erst durch die Arbeiten Neymans eine Verschränkung zwischen den Problemstellungen empirischer Surveys und der theoretischen Inferenzstatistik entwickelt hat.

In den 30er-Jahren des vorigen Jahrhunderts entwickelten Kapazitäten wie Fisher, Hotelling, Wald, E. Pearson und eben Neyman in einem Wechselspiel zwischen praxisrelevanten Fragestellungen und mathematischer Abstraktion die theoretischen Grundkonzepte der modernen, allgemeinen Inferenzstatistik, welche in ihren Grundzügen bis heute die allgemeine Basis der statistischen Methodenlehre bildet. Ausgehend von praktischen Problemen entwickelten diese genialen Vordenker ein einheitliches mathematisches Framework für optimale Entscheidungen unter Unsicherheit.

Bis hin zum Beginn der 60er-Jahre erlebte diese Ära der theoretischen mathematischen Statistik ihre Blütezeit und dominierte die statistische Methodenlehre. Die wohl wichtigsten Beiträge im Bereich der Stichprobentheorie mit großer Relevanz für die Arbeit statistischer Ämter bildeten die Arbeiten von Cochran, der 1942 wichtige theoretische Ergebnisse zur Verhältnis- und Regressionsschätzung in finiten Grundgesamtheiten veröffentlichte, die Ergebnisse von Hansen und Hurwitz, die 1943 das Prinzip der größenproportionalen Stichprobenziehung begründeten sowie der Beitrag von Horvitz und Thompson, die 1952 ein allgemein anwendbares Schätzprinzip für designbasierte Stichproben entwickelten.

Statistische Modellierung

Die nächste wichtige Epoche der statistischen Methodik läutete John W. Tukey (1962) mit seiner berühmten Arbeit „The Future of Data Analysis" ein. Charakteristisch für diese Ära ist die Abkehr von reiner mathematischer Theorie hin zur statistischen Modellierung realer Problemstellungen im Rahmen der angewandten Statistik.

Ein wichtiger Meilenstein dieser Zeit ist die explizite Abgrenzung zwischen der klassisch konfirmatorischen Analytik und der explorativen Datenanalyse. Von großer Bedeutung für die Anwendbarkeit statistischer Modelle auf neue Fragestellungen sind insbesondere die diversen Weiterentwicklungen von Regressionsmodellen (Logistische Regression, Log-lineare Modelle, Generalized Linear Models, Generalized Additive Models, Proportional Hazard Regression Models) sowie nichtparametrische und robuste Techniken anzuführen. Aber auch die Entwicklung von Re-Sampling Techniken und hier insbesondere das von Bradley Efron entwickelte Bootstrap-Verfahren stellen herausragende methodische Entwicklungen dar. Als besonders prägende Persönlichkeiten sind neben Tukey und Efron an dieser Stelle die beiden britischen Statistiker David R. Cox und John Nelder zu nennen.

Den theoretischen Konnex zwischen diesen Modellen der angewandten Statistik und der Survey-Statistik bildeten die Arbeiten von Royall (1970), der mit dem Prinzip des „Prediction Approach" eine neue theoretische Fundierung bereits bekannter

Ergebnisse basierend auf dem Konzept einer unendlich großen Superpopulation gab. Das Konzept des „Model Based Inference“ eröffnet der amtlichen Statistik für die Analyse von erhobenen Stichprobendaten die Nutzung eines weiten Methodenspektrums, welches ursprünglich in einem ganz anderen Kontext entwickelt wurde.

Ein weiterer wichtiger Beitrag zur Methodik der amtlichen Statistik lieferte Rubin, der 1976 mit seinem Beitrag „Inference and Missing Data“ eine theoretische Fundierung für den Umgang mit fehlenden Werten gab, auf deren Basis in der Folge systematische Techniken zur Imputation entwickelt wurden.

Einen anderen für die Methodik amtlicher Statistik relevanten Aspekt bildet die Entwicklung der Theorie der „small area estimates“ in den 80er-Jahren, welche in einem Übersichtsartikel von Gosh und Rao „Small Area Estimation: An Appraisal“ 1994 exzellent dokumentiert wird.

Der Einfluss moderner Informationstechnologien

Natürlich hat der Siegeszug moderner IT-Technik auch die Arbeitswelt der Statistik nachhaltig verändert. Zunächst ermöglichte die gestiegene Rechenperformance die Anwendung neuer flexiblerer statistischer Modelle. Waren die Methodiker der theoretischen Ära oft zu restriktiven Modellannahmen (z. B. Linearität, Varianzhomogenität) gezwungen, um analytische Lösungen zu erhalten, die mit den damaligen Beschränkungen computationaler Techniken lösbar waren, konnte nunmehr ein wesentlich breiteres Modellspektrum abgedeckt werden. Die Anwendung iterativer Lösungsalgorithmen zur Schätzung von Modellen ohne explizite analytische Lösung sowie Techniken der globalen Optimierung (z. B. Genetische Algorithmen, Simulated Annealing) eröffneten der statistischen Modellierung völlig neue Möglichkeiten.

Der Einsatz des Werkzeugs Computer, um aus Daten wertvolle Erkenntnisse ziehen zu können, führte aber auch zu neuen methodischen Entwicklungen, die oft außerhalb der statistischen Community in interdisziplinären, aber stark von der Informatik beeinflussten Forschungsgebieten (Machine Learning, KDD knowledge discovery from databases, Data Mining) ihren Ursprung haben. Diese Methodologie ersetzt – vereinfacht dargestellt – stochastische Modelle als Nukleus methodischer Überlegungen durch algorithmische Konzepte.

Einer der ersten Statistiker, der die Mächtigkeit dieser neuen datenanalytischen Möglichkeiten früh erkannte, war Leo Breimann, der 1984 mit seinem Werk über Classification and Regression Trees eine neue Ära statistischer Modellierung begründete und 2001 in einem viel beachteten Artikel in Statistical Science „Statistical Modeling: The Two Cultures“ dieses neue Paradigma prägnant charakterisierte:

„There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. … If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.”

Zu den wichtigsten Verfahren dieser stark von der Informatik beeinflussten Methodik zählen Neuronale Netzwerke, Rule Induction Methods oder auch Support Vector Machines.

Einen weiteren wichtigen Aspekt bildet in diesem Kontext auch die Ensemble-Modellierung (z. B. Random Forests und Boosting-Techniken). Die Frage nach einem optimalen Modell wird hier zu Gunsten einer multiplen Modellierung verworfen.

Versuch eines Ausblicks

Die Leistungsfähigkeit moderner Informationstechnologie und die Omnipräsenz von umfassenden Datenbanken in der öffentlichen Verwaltung und der modernen Wirt-

Alles was lediglich wahrscheinlich ist, ist wahrscheinlich falsch.

RENÉ DESCARTES

schaft, werden die künftige Arbeitsweise der amtlichen Statistik signifikant beeinflussen. Die Verfügbarkeit riesiger allerdings nicht primär für statistische Zwecke gesammelter Datenmengen stellt dabei neue Herausforderungen an die Datenmodellierung.

Mittels geeigneter Metadatenkonzepte und Methoden zur Sicherung der Datenqualität können registerbasierte Statistiken die Arbeitswelt der amtlichen Statistik schon in naher Zukunft nachhaltig verändern. Die von Statistik Austria durchgeführte Registerzählung 2011, welche die klassische Volkszählung ablöste, stellt hier einen auch im internationalen Kontext markanten Meilenstein dar. Effizienzdenken und der Wunsch nach Respondenten-Entlastung, die bereits vor rund 100 Jahren der Stichprobentechnik zum Durchbruch verhalfen, sind auch heute wieder Hauptmotivation für die Entwicklung von Techniken zur registerbasierenden Produktion von aussagekräftigen und qualitativ hochwertigen Statistiken.

Einen anderen Aspekt bildet das inhaltliche Streben nach neuen, kohärenten und relevanten Indikatoren-Systemen zur Messung realer Zustände unserer Gesellschaft und unserer Umwelt, was eine Kernaufgabe für die amtliche Statistik darstellt. Zwangsläufig wird die gestiegene Komplexität wirtschafts- und sozial- und umweltpolitischer Fragestellungen auch nach komplexeren, interdisziplinär vernetzten Modellen verlangen, welche ihrerseits neue methodischer Ansätze und Herangehensweisen in der amtlichen Statistik bedingen werden.

Diskussionsrunde: Amtliche Statistik in Österreich – Kontinuität im Wandel
v.l.n.r.: Univ.-Doz. MMag Dr. Weigl, HR Dr. Fürst, A.o. Univ.-Prof. Dr. Hudec

Diskussionsrunde: Statistik in der Verwaltung – Kontinuität im Wandel
v.l.n.r.: SR^{in} Jary, Dipl.-Vw. Himpele, Mag. Dr. Lebhart, Mag. Jörg