

Objektivistische oder subjektivistische Statistik?

Zur Überfälligkeit einer Grundsatzdiskussion

von Raphael Diepgen, Bochum

Was denn nun? Einerseits behauptet Buth (1991, S. 21) jüngst in dieser Zeitschrift, beim "offiziellen" statistischen Hypothesentesten lasse sich "nur die ziemlich belanglose Wahrscheinlichkeit" berechnen, "mit der man einen Irrtum riskiert, wenn man sich gegen eine Hypothese ausspricht, die in Wirklichkeit zutrifft". Anstatt nun konsequenterweise die offiziellen Verfahren der Stochastik, hier also den Signifikanztest im Sinne von Neyman und Pearson, wegen Belanglosigkeit aus dem Stochastikunterricht zu verbannen, bekümmert Buth andererseits die Schwierigkeiten beim Verständnis dieser Verfahren im Unterricht, die - so Buth bedauernd - leider kaum durch die Bayessche Perspektive erleichtert würden, wie von Riemer mehrfach versprochen. Da fragt sich der kopfschüttelnde Leser: Was interessiert es Buth, Verfahren verständlich zu machen, die er - nicht ohne Grund - für belanglos hält? Riemer (1991 a, b) nun geht in seiner Antwort, höflich und diplomatisch, wie er ist, auf diesen Widerspruch nicht ein, sondern präsentiert - als sei nichts gewesen - ein schönes, unterhaltsames und schülerfreundliches Konzept für einen propädeutischen Unterricht, der verblüffenderweise in Berechnung genau jener Wahrscheinlichkeiten mündet, die Buth zuvor als belanglos kritisiert hat.

Man gewinnt den Eindruck, es solle für den Bereich der Statistikdidaktik das Aufbrechen jenes Konfliktes vermieden werden, der für die mathematische und angewandte Statistik seit langem prägend ist, der Konflikt nämlich zwischen einerseits der "objektivistischen" Schule, die die Wahrscheinlichkeiten nur als klassische oder statistische kennt und daher inferenzstatistische Verfahren nur verstehen kann im Sinne Neymans und Pearsons als *im Vorhinein* festgelegte Randomisierungsstrategien mit berechenbaren hypothetischen Erfolgs- oder Fehlerwahrscheinlichkeiten, und andererseits der "subjektivistischen" Schule bayesscher Provenienz, für die es Wahrscheinlichkeiten auch gibt als Maße subjektiver Glaubensstärke, die es gemäß dem ehrwürdigen Theorem von Thomas Bayes jeweils *nach* Konfrontation mit der Empirie fortlaufend zu modifizieren gilt. Dieser Konflikt läßt sich verstehen als ein Konflikt über zulässige Interpretationen des Kolmogoroffschen Axiomensystems der Wahrscheinlichkeitstheorie, nämlich darüber, ob auch subjektive Wahrscheinlichkeiten ein Maß im Kolmogoroffschen Sinne

und daher legitimer Gegenstand wahrscheinlichkeitstheoretischer Argumentation sind. Zugleich aber läßt sich dieser Konflikt verstehen als Konflikt über die Größe des Anwendungsbereichs für mathematische Stochastik: Für die Bayesianer ist dieser Anwendungsbereich viel breiter als für die Objektivisten. Denn die *im Vorhinein* definierten objektivistischen Strategien *mit Randomisierung* - wie den Signifikanztest oder das Konfidenzintervall durchzuspielen, dies ist offensichtlich nur in eher wenigen Forschungs- und Praxisbereichen, nämlich solchen mit hoher Planbarkeit, möglich, etwa in der industriellen Qualitätskontrolle, der pharmazeutischen Forschung, bei technischen Prüfanstalten. Es müssen dafür schließlich Stichproben vorab fixierter Größe wirklich nach Zufall gezogen werden, es müssen die Frage- und Problemstellungen von Anfang an klar sein und während der Datenerhebung unverändert bleiben, es darf nichts Unvorhergesehenes geschehen usw. Wer beispielsweise empirische Sozialforschung, Demoskopie oder Marktforschung aus der Praxis kennt, der weiß, wie unerreichbar - und unsinnig - solche Bedingungen zumeist sind. Angesichts der Enge des Anwendungsbereiches objektivistischer Verfahren fragt sich natürlich, was sie im allgemeinbildenden Curriculum eines anwendungsorientierten Mathematikunterrichts noch zu suchen haben, und dann, ob nicht tatsächlich eine konsequente Umorientierung des Unterrichtes hin zu Bayes sinnvoll wäre gerade aus dem Wunsch nach breiter Anwendbarkeit heraus.

Ja, aber die objektivistischen Verfahren wie der Signifikanztest werden doch tageintagsaus tausendfach angewandt! - höre ich dort einwenden. Irrtum: Es werden - jedenfalls in vielen humanwissenschaftlichen Kontexten - nicht die objektivistischen Verfahren, d.h. also im Vorhinein definierte Strategien mit Randomisierung durchgespielt, sondern es wird häufig nur so getan als ob. Bleiben wir beim Signifikanztest: Der Test wird häufig **nicht** vor der Datenerhebung festgelegt, die Hypothesen werden häufig **nicht** vor der Untersuchung formuliert, die Stichprobe wird häufig **nicht** nach Zufall gezogen, der Stichprobenumfang wird häufig **nicht** vorab fixiert, und überdies soll häufig auch gar **nicht** im Ernst eine endgültige Entscheidung zwischen "Annahme" und "Verwerfung" einer Hypothese getroffen werden, zumal in der handlungsfernen Grundlagenforschung meistens auch noch völlig unklar bleibt, was dieses "Annehmen" oder "Verwerfen" eigentlich bedeutet. (Kann man sich etwa per vorab festgelegtem Willensakt dazu "entscheiden", eine Hypothese für wahr bzw. falsch zu halten?) Und doch wird in den Forschungsberichten - nämlich spätestens in der Angabe von sog. "Signifikanzniveaus" - immer so getan, als ob man dies alles nicht gemacht, als ob man tatsächlich ein im Vorhinein definiertes randomisiertes Entscheidungsverfahren durchgespielt hätte.

Kurzum: Die verbreitete "Anwendung" objektivistischer Verfahren entpuppt sich in vielen, insbesondere humanwissenschaftlichen Feldern häufig nur als eine vermeintliche; ihre ernsthafte Anwendung ist nämlich - jedenfalls in diesen Kontexten - nur selten

möglich und sinnvoll. Tatsächlich dienen die Theorien objektivistischer Statistik dort vielmehr häufig lediglich als ideologischer Überbau für eine ganz anders orientierte Praxis. Dafür hat Gigerenzer (1987, S. 23) - mit Blick auf die empirische Psychologie - in psychoanalytischer Analogie das treffende Bild eines Neyman-Pearson-Überichs und eines Fisherschen Ichs geprägt: Für das Überich des Forschers ist der Signifikanztest im objektivistischen Sinne Neymans und Pearsons ein im Vorhinein festgelegtes randomisiertes Entscheidungsverfahren. Für das praktisch wirksame Ich indessen ist der Signifikanztest den nichtobjektivistischen Plausibilitätsideen Fishers folgend die bloße ex-post-Berechnung und Interpretation von Überschreitungswahrscheinlichkeiten. So geht denn einher die irrationale Praxis des Ichs mit der objektivistischen Rationalität des Überichs. Interessant ist insbesondere die These Gigerenzers (1987, S. 24 f.), daß die Etablierung objektivistischer Als-ob-Statistik im humanwissenschaftlichen Alltag gerade davon lebt, daß die Praktiker und Forscher nicht wissen, was sie tun, und zwar trotz der klaren, objektivistischen Sonntagsreden: "Nevertheless, these clarifications have had almost no effect on research practice. Writers and consumers of null hypothesis testing still believe that the anonymous 'statistic' tells us exactly what we want to know." Dabei spielen gerade die - von Buth wie Riemer beklagte - Bayessche Mißinterpretation des Signifikanzniveaus eine wichtige Rolle: "Here the likelihood $p(D|H_1)$ is erroneously equated with $p(H_1|D)$, the Bayesian posterior probability. This is what we mean by the 'Bayesian interpretation' of 'significance levels'. Of course, such errors are necessary to maintain the illusion that null hypothesis testing can tell us what we want to know, that is, give us information about the validity of hypotheses. It might be asked how such illusions and errors could have persisted since they were introduced in the early years of the inference revolution. We believe that such illusions were *necessary* to maintain the dream of mechanized inductive inference."

Diese - von manchen Praktikern, möglicherweise aber trotz gegenteiliger Lippenbekenntnisse auch von manchen Statistikern - erwünschten Illusionen werden natürlich gefördert durch den üblichen unpräzisen Sprachgebrauch in der Statistikausbildung, der bestimmte Mißinterpretationen nahelegt. Beispiele aus Riemers (1991 b) Unterricht, der in dieser Hinsicht ganz im Fisherschen Rahmen bleibt: "Welche Folgerungen würdest Du ziehen, wenn Deine Trefferhäufigkeit 'oberhalb' ... (bzw.) 'unterhalb' des 90%-Intervalls liegt? Im zweiten Fall ist man vermutlich Inversschmecker" (S. 29). Ja, wieso denn? Was heißt hier "vermutlich"? Oder: "Ist Dein Testergebnis mit der Hypothese vereinbar, daß Du keinen Unterschied schmeckst ...?" (S. 30) Ja, welches Testergebnis könnte damit schon "unvereinbar" sein? Hier werden doch, wenn auch sicherlich nur in abkürzender Absicht, genau jene überaus suggestiven Sprechweisen eingeführt, die die genannten Illusionen erst ermöglichen.

Wie es zu der verbreiteten Pseudoanwendung objektivistischer Verfahren insbesondere in den empirischen Humanwissenschaften kam, kann hier nicht ausführlich erörtert

werden: eine ganze Fülle von interessanten wissenschaftspolitischen, wissenschafts- und professionssoziologischen und psychologischen Interpretationen bieten sich dafür an. Eine kleine Auswahl:

- Inferenzstatistik ist *das* Mittel der Humanwissenschaften, in der Hoffnung auf Reputationsgewinn die Mathematisierung der renommierten Natur- und Technikwissenschaften zu kopieren.
- Die vermeintliche Objektivität objektivistischer Verfahren nährt die für den wissenschaftlichen Wissenschaftsbetrieb ressourcensichernde Illusion subjektunabhängiger und damit unparteilicher Erkenntnis. Bayessche Verfahren könnten diese Funktion kaum erfüllen, da sie unübersehbar abhängig von subjektiven Vorbewertungen sind.
- Die - vermutlich nichtprobabilistisch interpretierten - statistischen "Signifikanzen" schützen den Forschungspraktiker vor dem unangenehmen Erlebnis der Zufälligkeit und Unsicherheit seiner Ergebnisse. Jede Profession, auch die der empirischen Wissenschaftler, entwickelt Mechanismen zur Verdrängung der Unsicherheit, Fraglichkeit und Zweifelhaftheit ihres professionellen Tuns. Die objektivistische Als-ob-Statistik nimmt den empirischen Forschern diese unangenehme Unsicherheit gleichsam ab. Diese beruhigende und entlastende Funktion der "Mechanisierung des induktiven Schließens" (Gigerenzer) kann Inferenzstatistik desto weniger erfüllen, je mehr ihr probabilistischer Charakter, etwa in der Randomisierung, für den Forscher erlebbar wird - wie beispielsweise bei den sequentiellen Verfahren von Walds, die daher kaum angewandt werden.
- Die etablierten und manchmal längst zu ritualisierten Standards erstarrten Verfahren objektivistischer Statistik liefern recht einfach und bequem zu handhabende Kriterien zur Beurteilung von empirischen Arbeiten: Ist der Hypothesentest richtig durchgeführt worden, ist sein Ergebnis signifikant, ja oder nein? Da viele wissenschaftliche Arbeiten an den Hoch-Schulen nicht geschrieben werden, um irgendwelche Probleme zu lösen, sondern vielmehr, um statuslegitimierende Qualifikationen nachzuweisen, um also - von Professoren, Fachkollegen, Gutachtern, Herausgebern, Berufungskommissionen - beurteilt zu werden, sind bequeme und dank Konvention eindeutige Beurteilungskriterien sehr viel wert - und zwar gleichermaßen für den Beurteiler wie den Beurteilten. Hier fungiert - im wahrsten Sinne des Wortes - "Beurteilende" Statistik als ähnlich nützliches Ersatzkriterium wie etwa die äußere Form der Arbeit, die richtige Zitierweise, die orthographische Sorgfalt und die formale Korrektheit. Vor diesem Hintergrund mag übrigens die These mancher Wissenschaftssoziologen verständlich erscheinen, die Etablierung von Statistik sei ein Element im Prozeß der Verkleinbürgerlichung der Humanwissenschaften.

- Statistische Begrifflichkeiten liefern den Humanwissenschaftlern ein Mittel zur Konstruktion einer für den Laien unverständlichen Fachsprache und dienen so zur professionssoziologischen Abgrenzung einer (vermeintlichen?) Expertengruppe. Statistische Ausdrücke mögen hierfür heute eine ähnliche Rolle spielen wie seinerzeit für den Ärztestand die lateinischen Termini.
- Nicht zuletzt dient in vielen Studiengängen die umfangreiche und anspruchsvolle Statistikausbildung als Selektionsinstrument: Statistik ist in den mathematikfernen Wissenschaften wohl die Teildisziplin, an der die meisten Studenten scheitern. Hier dient Statistik dem natürlichen Interesse so mancher Profession, den Zugang zu ihrem Berufsstand zu kanalisieren.

Diese wenigen Andeutungen mögen genügen als Hinweis darauf, daß die verbreitete, aber häufig nur vermeintliche Anwendung der Verfahren objektivistischer Statistik, so sehr sie auch im partikularen Interesse der Profession der Statistiker liegen mag, kaum vernünftiger Grund sein kann, ihnen weiterhin auf der Schule so viel Raum zu geben. Dennoch wird dies wohl auf absehbare Zeit so bleiben, eingedenk der vielen Interessen, die hier involviert sind, und der bekannten Schwerfälligkeit des Tankers "Mathematikunterricht", dem auch für nur kleine Kursänderungen Jahrzehnte kaum genügen. Nun sind Mathematiklehrer zwar an den entfremdenden Zwang gewöhnt, Inhalte zu unterrichten, deren Sinn ihnen selbst nicht so ganz klar ist; dennoch wird man auch ihnen das Recht auf Distanzierung zugestehen. Insofern verdient Buth natürlich Verständnis und Unterstützung für seine Forderung, im Unterricht die objektivistischen Verfahren wenigstens exakt zu interpretieren und so als "ziemlich belanglos" und enttäuschend zu demaskieren. Dazu liegen eine Reihe von Vorschlägen vor (Diepgen u. a., 1989, S. 154ff.), insbesondere der Vorschlag, die objektivistischen Verfahren grundsätzlich in Walds sequentieller Form zu präsentieren (Diepgen u. a., 1989, S. 162ff., Diepgen, 1987). Und in der Tat scheint zweifelhaft, ob der Bayessche Unterbau à la Riemer *hierbei* sonderlich hilft. Solange nämlich - wie heute üblich - Bayes' Theorem im Unterricht behandelt wird nur für den Fall weniger konkurrierender Hypothesen, also bei diskreten Verteilungen, solange kann dem Schüler überhaupt nicht klar werden, daß es zu den objektivistischen Verfahren Signifikanztest und Konfidenzintervall bayessche Alternativen gibt. Denn diese bayesschen Alternativen bedürfen "strenggenommen" unendlicher Hypothesenräume und stetiger Zufallsvariablen. So sind für den Schüler heute bayessche und objektivistische Denkweisen Argumentationsfiguren für gänzlich unterschiedliche Fragestellungen und Situationen, die überhaupt nicht miteinander in Kontrast und Konkurrenz treten. Anders formuliert: Der Schüler kann heutzutage das Signifikanzniveau nur deshalb bayessch fehlinterpretieren, weil er nirgends eine bayessche Alternative zum Signifikanztest kennengelernt hat. Und es mag sogar sein, daß diese Fehlinterpretation besonders naheliegt dann, wenn der Schüler gemäß Riemers Vorschlag Begriffe wie "Hypothese", "Datum" und "likelihood" zunächst im

diskreten bayesschen Kontext erlernt und dann später im objektivistischen Rahmen dort, wo dieselben Begriffe vorkommen, sogleich die vertrauten bayesschen Wahrscheinlichkeiten assoziiert. Diese Problematik läßt sich indessen lösen: Ohne Schwierigkeiten können nämlich schon im Grundkurs bayessche Alternativen zu den klassischen Verfahren besprochen werden dann, wenn die entsprechenden stetigen Verteilungen durch diskrete Verteilungen approximiert werden - was im Zeitalter computergestützter Statistik ohnehin angemessen erscheint (Diepgen u. a., 1989, S. 125).

Bleibt schließlich noch die Idee, Stochastikunterricht - jedenfalls in seinen inferenzstatistischen Teilen - langfristig von den objektivistischen Verfahren ganz weg hin zur subjektivistischen Bayesstatistik umzuorientieren. Freilich lauern auch dann unangenehme Fragen, was die Anwendungsrelevanz angeht. Was nämlich Riemer nicht müde wird als didaktischen Vorteil der bayesschen Regel zu loben und für seinen schülerzentrierten Unterricht zu nutzen, nämlich ihre intuitive Einsichtigkeit, dies wird sehr schnell zum Problem: Was benötigen wir mathematisierte Verfahren, die (genau?) das tun, was wir ohnehin intuitiv selbst tun würden - jedenfalls solange wir dieses intuitive Tun nicht an einen Computer delegieren wollen? Wenn der Mensch ohnehin schon intuitiv nach der bayesschen Regel denkt, was braucht er da noch Unterricht über diese bayessche Regel? Was bringt denn eigentlich hier die Explizierung implizierter Denkstrukturen? Freilich, es gibt seit nunmehr zwei Jahrzehnten eine ganze Heerschar von kognitiven Psychologen um Kahnemann und Tversky, die sich um den Beweis bemühen, daß der Mensch immer wieder in relevanten Bereichen durch hinderliche Heuristiken davon abgehalten werde, Erfahrungen so zu verarbeiten, wie es nach Bayes "rational" wäre. Es ist schwer, ein solch verästeltes Forschungsprogramm summarisch zu bewerten; es wäre aber sicherlich verfrüht zu behaupten, daß ihnen dieser Beweis bislang überzeugend gelungen sei. Und völlig offen ist, ob diesen möglichen kognitiven Fehlleistungen überhaupt durch eine formale Schulung vorgebeugt werden könnte. Auch ein Statistikunterricht auf bayesscher Basis dürfte es also nicht leicht haben, sich allein durch relevanten Anwendungsbezug überzeugend zu legitimieren. Immerhin kostet der deutschen Gesellschaft die stochastische Lehre pro Jahr wohl mehr als eine halbe Milliarde DM.

Ein grundsätzlicheres Nachdenken über das Wozu dieses teuren Stochastikunterrichtes scheint da angezeigt: Das Problem Objektivisten versus Subjektivisten wäre dabei nur eines unter vielen anderen. Vielleicht läßt sich Stochastikunterricht auf Dauer auch gar nicht überzeugend legitimieren, solange er sich als Teil des Mathematikunterrichtes versteht. Denn Stochastik als "Kunst des vernünftigen Vermutens" profitiert doch offensichtlich nur begrenzt und eher wenig von ihrer Mathematisierung. Lediglich für wenige Situationen der Ungewißheit verspricht mathematische Argumentation einen nennenswerten Rationalitätsgewinn; nichtmathematische Argumentationsfiguren mögen hier meistens viel wichtiger sein. Vielleicht könnte es daher in weiter Zukunft einen mathematikferneren Stochastikunterricht

geben. in dem es um Ungewißheit. Nichtwissen. Unberechenbarkeit und Unsicherheit geht auch aus philosophischer, psychologischer, soziologischer und politischer Perspektive (vgl. Smithson, 1988).

Literatur

- [1] **Buth, M.** (1991): *Die Behinderung des gesunden Menschenverstandes durch Stochastik*. Stochastik in der Schule 3, 12-22
- [2] **Diepgen, R.** (1987): *Sequentielles Testen - auch didaktisch vielleicht eine gute Alternative*. Stochastik in der Schule 2, 9-25
- [3] **Diepgen, R., Kuypers, W. und Rüdiger, K.** (1989): *Stochastik Grundkurs*. Düsseldorf: Cornelsen Verlag Schwann-Girardet
- [4] **Gigerenzer, G. und Murray, D. J.** (1987): *Cognition as Intuitive Statistics*. Hillsdale, N.J.: Lawrence
- [5] **Riemer, W.** (1991a): *Anmerkungen zu Buth: Die Behinderung des gesunden Menschenverstandes durch Stochastik*. Stochastik in der Schule 3, 23
- [6] **Riemer, W.** (1991b): *Das 'Eins durch Wurzel aus n'-Gesetz. Einführung in statistisches Denken auf der Sekundarstufe I*. Stochastik in der Schule 3, 24-36
- [7] **Smithson, M.** (1989): *Ignorance and Uncertainty. Emerging Paradigms*. New York: Springer

Anmerkung der Herausgeberin:

Wir bitten unsere Leserinnen und Leser, diesen Diskussionsbeitrag zu einem 'heißen' und didaktisch wichtigen Thema zu überdenken und ihn interessierten Kollegen und Kolleginnen, die *Stochastik in der Schule* noch nicht kennen, ebenfalls zu lesen zu geben. Wir hoffen auf rege Rückmeldung z.B. in Form von Leserbriefen, auch um ein - natürlich nicht unbedingt repräsentatives - Bild der Meinungen in der Lehrerschaft dazu zu erhalten.