Einfache Paradoxien der beschreibenden Statistik

Von JÖRG MEYER, Hameln

Zusammenfassung: *Es wird eine größere Reihe von Sachverhalten vorgestellt, die der Beschreibenden Statistik zugerechnet werden können und die dem gesunden Menschenverstand paradox oder doch sehr verblüffend erscheinen, jedenfalls auf den ersten Blick. Die Ausgangsbeispiele sind oft sehr einfache Situationen, die aber auf praktisch wichtige Fälle übertragen werden können.*

Manchmal sehen wir etwas, geben aber nicht acht. Seneca.

Die Genese der gesamten elementaren Stochastik wird geradezu beherrscht von Auseinandersetzungen über Paradoxien. Winter.

0. Einleitung

Es gibt verschiedene Klassen von Paradoxien (Ansätze einer Klassifikation liefert Poundstone 1991, S. 16-19). Die hier behandelten sind solche, die dem gesunden Menschenverstand widersprechen. Bei hinreichend naher sachgemäßer Betrachtung entpuppen sie sich als Banalitäten.

Dies soll in diesem Aufsatz anhand einiger einfacher Beispiele aus der beschreibenden Statistik illustriert werden. Dabei wird oft der umgekehrte Weg durchlaufen: Banalitäten werden schief beleuchtet, so daß sie paradox erscheinen.

Die Anwendungen sind mitunter überraschend. Sie sind ausgesprochen oberflächlich und zeigen so etwas von der Buntheit der Welt.

Zu den hier behandelten Paradoxien gehören auch die von Blyth und Simpson, die trotz ihrer Einfachheit erst 1972 bzw. 1951 beschrieben wurden. Sie werden in gestufter Form innerhalb eines gemeinsamen Kontextes dargestellt.

Selbstverständlich ist es nicht sinnvoll, alle Kapitel dieses Aufsatzes geballt und zusammenhängend im Unterricht zu behandeln. Vielmehr möge man sie als *Materialsammlung* ansehen, aus der dosiert und kontextabhängig Beispiele zur Illu-

Stochastik in der Schule **15**(1995), Nr. 2, S. 27-50

stration bzw. zur Motivation weiterführender Fragen ausgewählt werden können. Insbesondere sind die hier präsentierten Beispiele weder unterrichtsmäßig aufbereitet noch in ihrer theoretischen Substanz hinreichend diskutiert. Das Interesse konzentriert sich stattdessen auf den Zusammenhang der dargestellten Oberflächenphänomene.

1. Was heißt „Mehrheit"?

Wie oft kann ich mich nicht entscheiden! Soll ich dieses Papier nun in „Stochastik in der Schule" (Möglichkeit A) oder in einer anderen Zeitschrift (Möglichkeit B) veröffentlichen? Für beide Entscheidungen mögen gleich gewichtige und gleich gute Gründe sprechen.

In solchen Fällen überlasse ich gern einer *Münze* die Entscheidung. Dabei ist es natürlich sicherer, die Münze dreimal zu werfen und dann nach Mehrheit zu entscheiden. Vielleicht sollte ich lieber drei Dreierserien werfen? Die naheliegende Iteration breche an dieser Stelle ab; das Ergebnis der Münzwürfe ist:

1. Serie: A A B (also ist A Sieger der 1. Serie)
2. Serie: A B A (also ist A Sieger der 2. Serie)
3. Serie: B B B (also ist B Sieger der 3. Serie)

A hat demnach 2 von 3 Serien gewonnen, und B hat 5 von 9 Würfen gewonnen. Wer ist nun Sieger?

Man sieht: Die Mehrheit von Mehrheiten kann eine Minderheit darstellen!

Anders formuliert: Eine geschickte Partitionierung der Wirklichkeit kann zu falschen Aussagen führen!
Man muß nur die jeweiligen Mehrheiten iterativ ermitteln.

Anwendung auf Wahlen:

Beim Mehrheitswahlrecht handelt es sich um eine iterative Mehrheitsermittlung: In jedem Wahlkreis wird ein Sieger festgestellt; die Menge der Sieger bildet dann das Parlament (Beispiel: England). Nun kann es vorkommen (und in England war das mitunter auch schon der Fall), daß eine Partei zwar in der Mehrheit der Wahlkreise siegte, ohne aber die Mehrheit der (Ur-)wählerstimmen auf sich vereinigen zu können.

Natürlich sind *Möglichkeiten zur Manipulation* bei iterierten Wahlen reichlich vorhanden.
Dies läßt sich mit Hilfe des Anfangsbeispiels illustrieren: Die wählende Bevölkerung bestehe aus 9 Personen und sei in 3 Wahlkreise zu je 3 Personen eingeteilt; die Parteien heißen A und B.

Das Wahlverhalten sei:

1. Wahlkreis: A A B (also ist A Sieger im 1. Wahlkreis)
2. Wahlkreis: A B A (also ist A Sieger im 2. Wahlkreis)
3. Wahlkreis: B B B (also ist B Sieger im 3. Wahlkreis)

Bei dieser Wahlkreisaufteilung ist A Gesamtsieger.

Bemerkung: Der zu zeigende Effekt kommt natürlich auch bei großen (realitätsnäheren) Zahlen vor. Die hier gewählten kleinen (realitätsfernen) Zahlen dienen aber der größeren Durchsichtigkeit. Diese Bemerkung gilt auch für alle folgenden Beispiele.

Iteriert man die Partitionierung, so reichen schon $(2/3)^3$, also knapp 30% aller „Urstimmen" zum Sieg:

A A B	A		
A B A	A	A	
B B B	B		
A A B	A		
A B A	A	A	A
B B B	B		
B B B	B		
B B B	B	B	
B B B	B		

Vergröbert man dagegen die Wahlkreise von 3 auf einen, vermindert man also die Iterationstiefe, so ist B Gesamtsieger.

Verändert man hingegen die Wahlkreisaufteilung, indem man von der horizontalen

A A B	A
A B A	A
B B B	B

zur vertikalen Gliederung übergeht, so ist ebenfalls B Gesamtsieger:

$$\begin{array}{c} A\ A\ B \\ A\ B\ A \\ B\ B\ B \\ \hline A\ B\ B \end{array}$$

Daß eine Partei die Mehrheit der Wahlkreise erreichen kann, aber nicht die Mehrheit der Wählerstimmen, daß also die Mehrheit von Mehrheiten eine Minderheit darstellen kann, ist auch ein Grund dafür, daß das alte lateinische Erfolgsrezept „Divide et impera!" (Teile und herrsche!) so effektiv sein kann: Hetze Deine Gegner gegeneinander auf, und wenn das nicht zu Deinem Sieg führt, dann führe iterierte Wahlen ein! Solche iterierten Wahlen waren z.B. das Grundelement des „demokratischen Zentralismus" in „Rätedemokratien". Sie konnten zu einem völlig verfälschten Ergebnis führen!

Aber auch in Vereinen und Parteien werden Vorstandswahlen im allgemeinen iterativ durchgeführt. Nun sind solche Organisationen auf das Wohlwollen ihrer Mitglieder angewiesen, was Manipulationen verhindern sollte. Trotzdem kann es vorkommen, daß die Führung einer Partei andere politische Ziele verfolgt als die Mehrheit der Parteimitglieder und daß diese Differenz auch durch (iterierte) Wahlen sanktioniert worden ist.

Daß Iterationen grundsätzlich den Kern der Verfälschung in sich tragen, zeigt auch das folgende *krasse Beispiel*:

Eine Klasse bestehe aus genau einem Mädchen, eine Parallelklasse aus genau 10 Jungen. Eine Person dieser beiden Klassen soll ausgelost werden. Wie groß ist die Wahrscheinlichkeit, daß die ausgeloste Person ein Mädchen ist?

Lost man unter allen Schülern, so beträgt die Wahrscheinlichkeit 1/11.

Lost man erst die Klasse und dann aus dieser Klasse die Person, so beträgt die Wahrscheinlichkeit 1/2.

Ein Beispiel mit solch krassen Zahlen erscheint natürlich als absurd. Bei weniger krassen Zahlen dagegen hält jeder die beiden Losverfahren für äquivalent.

Zusammenfassung: *Wer meistens lokal gewinnt, kann trotzdem global verlieren.*

Das heißt: Eine Partei kann die Mehrheit der Wahlkreise gewinnen, ohne in der Bevölkerung eine Mehrheit zu haben. Dies ist die 1. Interpretation von „lokal", „global" und „besser".

2. Was heißt „besser"?

Mein Problem, ob ich dieses Papier nun in „Stochastik in der Schule" (Möglichkeit A) oder in einer anderen Zeitschrift (Möglichkeit B) veröffentlichen soll, ist nach den bisherigen Erkenntnissen durch (iterierte) Münzwürfe nicht zu lösen.

Statt einer Münze nehme ich nun einen *Würfel*. Kann nicht er mit seinen 6 Seiten die Präferenz der Pythia für A oder für B viel besser darstellen als eine Münze mit ihren nur 2 Seiten?

Beiden Möglichkeiten wird je ein Würfel zugeordnet. Wie bei Schulnoten soll eine niedrigere Augenzahl als besser als eine höhere gelten, *also sei* $1 > 2 > 3 > 4 > 5 > 6$. Der Würfel mit der besseren Augenzahl soll gewinnen.

Vorsichtshalber würfle ich je dreimal (eine derartige Vorgehensweise hatte ja schon im letzten Abschnitt zu mathematisch interessanteren Effekten geführt); das Ergebnis ist:

A: 3 6 2 (Durchschnitt: 3,7)
B: 4 3 3 (Durchschnitt: 3,3)

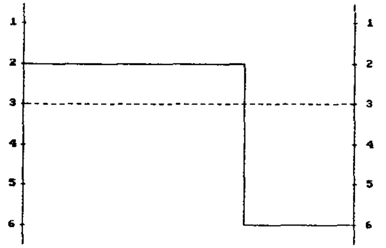
Meistens ist A besser als B, obwohl B mit 10/3 die bessere durchschnittliche Augenzahl als A mit 11/3 hat!

Man sieht: Wer meistens besser ist, kann im Schnitt schlechter sein!

Es kommt bei dem Spiel nicht auf die absoluten Ergebnisse von A und B an, sondern nur auf die Relation der Ergebnisse. Übersichtlicher ist der folgende, zum obigen äquivalente Spielverlauf:

A: 2 2 6 (Durchschnitt: 3,3)
B: 3 3 3 (Durchschnitt: 3,0)

Er läßt sich leicht veranschaulichen:



Meistens verläuft der strichlierte Graph unterhalb des durchgezogenen Graphen. Trotzdem ist der Flächeninhalt unterhalb des strichlierten Graphen größer als der unterhalb des durchgezogenen Graphen.

Anwendungen:

a. Schulnoten:
Ist A oder B der bessere Schüler?

b. Wetten:
Soll man auf A oder auf B wetten (der Wettgewinn richte sich nach dem Augenunterschied)?
Bei nur einer Wette ist eine Wette auf A sinnvoll, da meistens A gewinnt. Bei vielen Wetten sollte man auf B setzen, da dann der Erwartungswert des Gewinns positiv ist.

Das Beispiel zeigt demnach, daß eine Maximierung des Erwartungswerts (auf lange Sicht) etwas anderes ist als die Maximierung der Wahrscheinlichkeit, bei einer Gelegenheit das Optimum zu erreichen.

c. Aktien und Lose:
Otto kauft Aktien. Macht er das nur einmal und nur für kurze Zeit, so muß er anders beraten werden, als wenn er es häufig tut.

Bemerkung: Was heißt in diesem Zusammenhang "häufig"? Bei dem Würfelexperiment läßt sich das berechnen, bei Aktienkäufen dagegen nicht.

Erna überlegt sich, ob sie ein Los zu 10 DM kaufen soll. Sie habe eine Chance von 1:100, den einzigen Gewinn von 2000 DM zu erzielen (Diese Lotterie ist

also etwas merkwürdig). Der Erwartungswert für Erna beträgt $\frac{1}{100} \cdot 2000 - 10 = 10$; lohnt es sich also?

d. Medikamententests:
Medikament B hat konstante Wirkung. Medikament A ist meistens besser, in den anderen Fällen aber viel schlechter. Was soll man verschreiben?

e. Politikerauswahl:
Welchen Politiker soll man bevorzugen? Den Kompromißkandidaten, der keinen vom Stuhle reißt, mit dem aber alle einigermaßen leben können? Oder den Polarisierer mit großer fanatischer Gegnerschaft und etwas größerer fanatischer Anhängerschaft? (Von der Beantwortung dieser Frage hängt die Gestaltung des Wahlrechts ab.)

f. Wahlen mit Mehrheitswahlrecht:
Man habe 3 Wahlkreise. In den meisten Wahlkreisen ist A besser als B:

1. Wahlkreis:	A A B	(A ist besser als B)
2. Wahlkreis:	A B A	(A ist besser als B)
3. Wahlkreis:	B B B	(B ist viel besser als A)

Trotzdem ist insgesamt B besser als A.

g. Kumulation bei Wahlen:
Kumulation ist ein Fall der Mehrstimmgebung, der oftmals als „Verfeinerung“ angesehen wird.
Betrachten wir den einfachsten Fall: 3 Stimmberechtigte stimmen über 2 Kandidaten (A und B) ab.

Bei einer kumulativen Mehrstimmgebung habe jeder Wähler 3 Stimmen; das Ergebnis sei:

	A	B
1. Wähler:	2	1
2. Wähler:	2	1
3. Wähler:	0	3

B bekommt die meisten Stimmen, ist also gewählt.

Hat hingegen jeder Wähler nur eine Stimme (Einzelstimmgebung), so bekommt A 2 und B 1 Stimme; A ist also gewählt.

Einzelstimmgebung und kumulative Mehrstimmgebung können also zu ganz unterschiedlichen Resultaten führen! Insbesondere ist die Mehrstimmgebung keine „Verfeinerung" der Einzelstimmgebung.

h. Ranking vs. Scoring:
Insbesondere gilt: Die Umfrage „Wen halten Sie für besser, A oder B?" ist verschieden von der Umfrage „Benoten Sie A und B jeweils mit den Schulnoten 1 bis 6!" und kann zu einem ganz anderen Ergebnis führen!

Zusammenfassung: *Wer meistens lokal gewinnt, kann trotzdem global verlieren.*

Das heißt: Man kann die meisten Spiele gewinnen und trotzdem im Schnitt schlechter sein als sein Gegner. Dies ist die 2. Interpretation von „lokal", „global" und „besser".

3. „Besser" ist nicht transitiv

Würfel A war der Möglichkeit zugeordnet, dieses Papier in „Stochastik in der Schule" zu veröffentlichen, Würfel B der Möglichkeit, es sonstwo zu tun. Das letzte Wurfprotokoll war:

A: 2 2 6
B: 3 3 3

Nun ist meine Frau der Ansicht, dies Papier sei für eine Publikation noch nicht ausgereift genug. Sie hat deshalb heimlich im Hintergrund mitgewürfelt; ihr Würfel C ist somit der Möglichkeit zugeordnet, dies, was Sie hier sehen, gar nicht zu veröffentlichen. Zusammen mit meinen Würfeln ergibt sich das folgende Wurfprotokoll:

A: 2 2 6
B: 3 3 3
C: 4 1 4

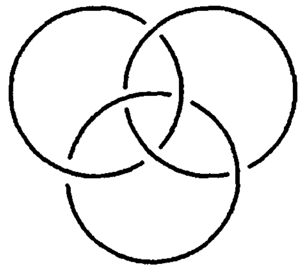
Ergebnis: Meistens ist A besser als B, meistens ist B besser als C, meistens ist aber auch C besser als A!

Man sieht: Die Relation „ist meistens besser als" ist i.a. nicht transitiv!

Nichttransitive Relationen kommen zwar häufig vor, sind aber trotzdem oft verblüffend. Dies wird durch die folgenden Beispiele illustriert.

Anwendungen:

a. Die Borromäischen Ringe bestehen aus drei Ringen A, B, C, so daß A über B und B über C sowie C über A liegt:



b. Die folgende Schlußkette beruht auf einer vermeintlichen Transitivität:
„Die meisten Freiburger sind Deutsche.
Die meisten Deutschen leben nicht in Freiburg.
Also:
Die meisten Freiburger leben nicht in Freiburg."

c. Mädchen haben die drei (als gleichrangig angesehenen Eigenschaften), mehr oder weniger hübsch, reich oder nett zu sein. Ein Mädchen wird für besser als ein anderes gehalten, wenn es bei mindestens zwei Eigenschaften höhere Werte hat.

Agnes ist hübsch, arm und mittelnett.
Berta ist mittelhübsch, reich und unwirsch.
Claudia ist häßlich, mittelreich und nett.

Dann ist Agnes besser als Berta, Berta besser als Claudia, Claudia besser als Agnes.

Man hätte auch sagen können: Agnes ist hübscher und netter als Berta, Berta ist hübscher und reicher als Claudia, Claudia ist reicher und netter als Agnes. Es scheint wesentlich zu sein, daß die Mädchen immer bezüglich anderer Eigenschaften verglichen werden.

d. Die Nichttransitivität von „ist meistens besser als" hat Konsequenzen für die Ermittlung von Gewinnern beim „k.o.-Verfahren": Von 3 Kandidaten (A, B, C) soll der Beste folgendermaßen ermittelt werden.

Spielregel: Erst zwei gegeneinander (welche, wird durch Los ermittelt), dann der Sieger gegen den Dritten.

Es sei jeweils meistens A besser als B, B besser als C, C besser als A.

Fall 1: Erst A gegen B (meistens ist A Sieger), dann A gegen C.
Meistens siegt C.
Fall 2: Erst B gegen C (meistens ist B Sieger), dann B gegen A.
Meistens siegt A.
Fall 3: Erst C gegen A (meistens ist C Sieger), dann C gegen B.
Meistens siegt B.

Man hätte sich die Zweikämpfe auch sparen können und den Sieger gleich durch Losen ermitteln können.

e. Man habe drei gezinkte Würfel A, B und C mit folgenden *Wahrscheinlichkeitsverteilungen*:

Das Würfeln mit B ergibt nur Dreien.
Das Würfeln mit A ergibt in 2/3 aller Fälle eine Zwei, sonst eine Sechs.
Das Würfeln mit C ergibt in 2/3 aller Fälle eine Vier, sonst eine Eins.

Dann gilt:

In 2/3 aller Fälle ist A besser als B.
In 2/3 aller Fälle ist B besser als C.
In 5/9 aller Fälle ist C besser als A.

Nun betrachte man das folgende *Spiel*: Die Personen P1 und P2 wählen je einen der Würfel A, B, C. Sie würfeln mit ihrem gewählten Würfel; die Person mit der besseren Augenzahl gewinnt.
P1 kenne die obige Wahrscheinlichkeitsverteilung der 3 Würfel, P2 nicht. Dann kann P1 großmütig seinem Konkurrenten anbieten, sich als erster einen Würfel zu wählen; er selbst findet dann immer einen meistens noch besseren.

f. Kümmert man sich nicht um die jeweiligen Augenzahlen, sondern nur darum, welcher Würfel besser ist als ein anderer, so läßt sich der Spielverlauf auch darstellen als

A B C
B C A
C A B

(*Lesebeispiel*: Beim ersten Wurf (erste Spalte) hat A die beste und C die schlechteste Augenzahl; beim 3. Wurf ist C optimal und B pessimal.)

Dies Protokoll läßt sich auch als *Präferenztabelle* interpretieren, in diesem Fall wird das Paradoxon nach dem Marquis de Condorcet (1743-1794) benannt.

Person P1 (1. Spalte) bevorzugt A gegenüber B und B gegenüber C; die weiteren Spalten gehören zu den Personen P2 und P3.

Nun wird sukzessive gewählt: Erst wird über A und B abgestimmt, dann über den entstehenden Sieger und C.

Beim 1. Wahlgang (A gegen B) stimmen P1 und P3 für A, also siegt A.
Beim 2. Wahlgang (Sieger A gegen C) stimmt nur P1 für A, also siegt C.

Mit diesem Ergebnis ist P1 natürlich nicht einverstanden. Bei der nächsten Abstimmung kann er darauf dringen, daß nicht zuerst über A und B abgestimmt wird. Sollte er dies nicht durchsetzen können, bleibt für ihn die Möglichkeit, beim 1. Wahlgang nicht für seinen Favoriten A, sondern für B zu stimmen. Dann siegt B im 1. und auch im 2. Wahlgang.

Bemerkung: Eine Fülle weiterer taktischer Möglichkeiten findet sich in dem Buch „Mathematik in der Praxis", S. 127 ff.

Aus der Sicht von P1 ist also die Strategie „Um das Schlimmste zu verhindern, darf man nicht immer für das Beste eintreten" erfolgreich.

Zusammenfassung: *Wer meistens lokal gewinnt, kann trotzdem global verlieren.*

Das heißt: Zur Abwehr eines globalen Gesamtverlustes sind mitunter lokale Zugeständnisse sinnvoll. Dies ist die 3. Interpretation von „lokal", „global" und „besser".

g. Das einfachste nichttransitive Beispiel eines Spielverlaufs ist

A: 1 3 2
B: 2 1 3
C: 3 2 1

Addiert man in der vorletzten Spalte 3 und in der letzten 6, so ergibt sich:

A: 1 6 8
B: 2 4 9
C: 3 5 7

Dies läßt sich auch folgendermaßen interpretieren: Gegeben seien 9 Spieler mit den Spielerstärken 1 bis 9 (wieder in der Schulnotenreihenfolge: 1 sei der beste und 9 der schlechteste Spieler). Diese 9 Spieler werden zu 3 Mannschaften zusammengesetzt: Mannschaft A bestehe aus 1, 6 und 8, Mannschaft B aus 2, 4 und 9, und Mannschaft C entspreche der letzten Zeile im Schema. In insgesamt 27 Spielen spiele jeder Spieler jeder Mannschaft je einmal gegen jeden Spieler aus den beiden anderen Mannschaften. Wenn nun immer der Bessere gewinnt, so erhält man:

Mannschaft A gewinnt 5 von 9 Spielen gegen Mannschaft B,
Mannschaft B gewinnt 5 von 9 Spielen gegen Mannschaft C,
Mannschaft C gewinnt 5 von 9 Spielen gegen Mannschaft A.

Übrigens: Wenn man die Mannschaftsstärke als durchschnittliche Spielerstärke definiert, erhält man für A, B und C jeweils den Wert 5. Mit einer solchen Definition trifft man das Phänomen also nicht.

4. Blyth

Im letzten Abschnitt wurde der Spielverlauf

A B C
B C A
C A B

diskutiert.

4.1. *Verzerrt* man diesen Spielverlauf auf 7 Partien, so lautet er:

A A	B B B	C C
B B	C C C	A A
C C	A A A	B B

Diese 7 Wettkämpfe werden am häufigsten von B gewonnen. Trotzdem wird B meistens von A geschlagen. Um diesen Sachverhalt klarer darstellen zu können, führen wir folgende Begriffe ein:

Ein *Duell* sei ein Wettkampf, an dem zwei Personen teilnehmen (Beispiel: Wettlauf von zwei Personen).
Ein *Triell* sei ein Wettkampf, an dem drei Personen teilnehmen (Beispiel: Wettlauf von drei Personen).

Man sieht: Wer die häufigsten Trielle gewinnt, kann trotzdem in Duellen meistens geschlagen werden.

Anwendungen:

a. Siebenmal laufen A, B und C um die Wette. Am häufigsten ist B Sieger. Dann stellt sich heraus, daß C nachträglich disqualifiziert werden muß, woraufhin A Gesamtsieger wird.

Dies Paradoxon wurde (in anderer Form) wohl erst 1972 von C. R. *Blyth* beschrieben. Der recht späte Zeitpunkt der Entdeckung wurde schon vor gut 200 Jahren von *Lichtenberg* kommentiert:

> *Es ist sonderbar, daß nur außerordentliche Menschen die Entdeckungen machen, die hernach so leicht und simpel erscheinen. Dieses setzt voraus, daß die simpelsten, aber wahren Verhältnisse der Dinge zu bemerken sehr tiefe Kenntnisse nötig sind.* (Sudelbücher J 1529)

b. Wieder gelte die Zuordnung

Würfel A: „Stochastik in der Schule"
Würfel B: sonstwo
Würfel C: gar nicht.

Ich möchte mich zwischen den Alternativen A und B entscheiden und würfle deshalb mit den Würfeln A und B je 7 mal; das Wurfprotokoll ist:

		B B B
A A	A A	A A A
B B	B B	

Die Entscheidung fällt somit für „Stochastik in der Schule".

Nun ist ja meine Frau für die Möglichkeit C; sie hat wieder mit ihrem C-Würfel an meinem Entscheidungsproblem teilgenommen. Zusammen mit meinen Würfeln ergibt sich das folgende Wurfprotokoll:

C C		B B B
A A	A A	A A A
B B	B B	
	C C	C C C

Da jetzt B am häufigsten gesiegt hat, fällt die Entscheidung somit gegen „Stochastik in der Schule".

c. Die Situation läßt sich auch mit der folgenden absurd erscheinenden Konversation illustrieren:

– „Was möchtest Du zum Nachtisch? Es gibt einen Apfel und eine Birne."
= „Ich möchte den Apfel."
– „Es gibt auch eine Carotte."
= „Dann möchte ich lieber die Birne."

4.2. Bei den Beispielen in **4.1** wurden die Trielle am häufigsten von B gewonnen. Betrachtete man sie als Duelle, so wurde aber meistens B von A geschlagen.

Der Zusatz „von A" läßt sich durch eine geeignete Modifikation beseitigen: Der Spielverlauf sei jetzt

A A	B B B	C C
C C	* * *	A A
B B	* * *	B B

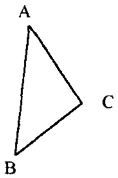
B gewinnt die häufigsten Trielle. Trotzdem wird B in den meisten Duellen geschlagen (von jedem Gegner).

4.3. Dies Beispiel läßt sich noch weiter ausschärfen: Das Wurfprotokoll sei jetzt

A A	B B B	B	C C C
C C	A A A	C	A A A
B B	C C C	A	B B B

Trielle werden am häufigsten von B und am seltensten von A gewonnen. Aber es gilt auch:

Fehlt A, so siegt meistens C über B;
fehlt B, so siegt meistens A über C;
fehlt C, so siegt meistens A über B.



Duelle werden demnach meistens von A gewonnen und meistens von B verloren (jeweils unabhängig vom Gegner).

Also gilt: In Bezug auf Duelle ist A optimal und B pessimal.
In Bezug auf Trielle ist es umgekehrt.

Zusammenfassung: Wer meistens lokal gewinnt, kann trotzdem global verlieren.

Das heißt: Wer Duelle meistens gewinnt, kann trotzdem die meisten Trielle verlieren. Dies ist die 4. Interpretation von „lokal", „global" und „besser".

4.4. Variiert man das Wurfprotokoll abermals, und zwar zu

A A A A A A	B B B	B B B B B	C C C C C	C C
C C C C C C	C C C	A A A A A	A A A A A	B B
B B B B B B	A A A	C C C C C	B B B B B	A A

so gilt wieder wie in **4.3**: In Bezug auf Duelle ist A optimal und B pessimal; in Bezug auf Trielle ist es umgekehrt.

Nun betrachte man folgende Wette (die sich auf Trielle bezieht):

3 Pferde A, B und C nehmen an Wettrennen teil. Man setze auf eines der drei Pferde.

Ist das gesetzte Pferd an 1. Stelle, so bekomme man 1 DM.
Ist das gesetzte Pferd an 2. Stelle, so bekomme man gar nichts.
Ist das gesetzte Pferd an 3. Stelle, so verliere man 1 DM.

Wenn nun 21 Rennen stattfinden, die gemäß dem obigen Würfelprotokoll verlaufen, so erweist sich insgesamt eine Wette auf C als optimal!

Denn: Setzt man auf A, so hat man nach 21 Rennen einen Gesamtgewinn von 1 DM, setzt man auf B, so hat man einen Gesamtverlust von 2 DM, und setzt man auf C, so hat man einen Gesamtgewinn von 2 DM.

Also gilt: In Duellen ist A optimal, in Triellen ist B optimal, und in Bezug auf eine Wette ist C optimal.

DAS universelle Kriterium gibt es nicht. Bevor man sich auf Wettkampfregeln einigt, sollte man genau seine eigenen Schwächen und die des Gegners kennen.

5. Simpson

Nachdem die bisherigen Würfeleien keine Klarheit gebracht haben, beschränke ich mich wieder auf die Alternativen A und B. Allerdings messe ich nun nicht mehr die Präferenz der Pythia, sondern die Änderung der Präferenz.

In einer ersten Serie würfle ich 5 mal A und 6 mal B;
die zweite Serie ist kürzer: 3 mal A und 4 mal B.

Die Präferenz für B hat sich von 55% auf 57% gesteigert.

Wiederum hat meine Frau im Hintergrund mitgewürfelt.

Ihre erste Serie war	6 mal A und 3 mal B;
ihre zweite lautete	9 mal A und 5 mal B.

Auch bei ihr hat sich die Präferenz für B gesteigert, und zwar von 33% auf 36%.

Damit ist wohl klar: „Stochastik in der Schule“ wird dieses Papier nicht zur Publikation bekommen! Endlich ist die Entscheidung gefallen...Aber nun?...Was ist, wenn man die Würfelresultate von meiner Frau und von mir zusammen betrachtet?

Dann wurden in der ersten Serie	11 mal A und 9 mal B
und in der zweiten Serie	12 mal A und 9 mal B

geworfen. Die Gesamtpräferenz der Pythia für B hat sich verringert, und zwar von 45% auf 43%. Also besteht doch noch eine Möglichkeit, daß Sie, verehrte Leserin von „Stochastik in der Schule“, meinen Beitrag in Ihrem Journal lesen können?

Dies Paradoxon ist nach E. H. Simpson (1951) benannt.

Man sieht: Was lokal wächst, kann global schrumpfen.

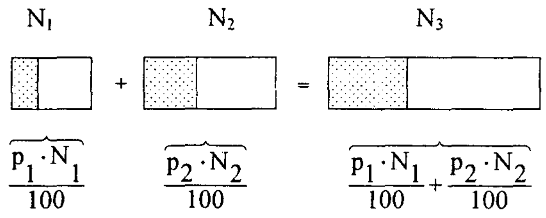
Wie soll man diese Paradoxie verstehen? Vielleicht durch ein extremes Beispiel?

	ich	meine Frau
1. Serie:	A B B B	A
2. Serie:	B	B A A A

Bei mir hat sich die Präferenz der Pythia für B von 75% auf 100% gesteigert, bei meiner Frau von 0% auf 25%. Bei uns beiden zusammen hingegen hat die Präferenz für B von 3/5 auf 2/5 verringert.

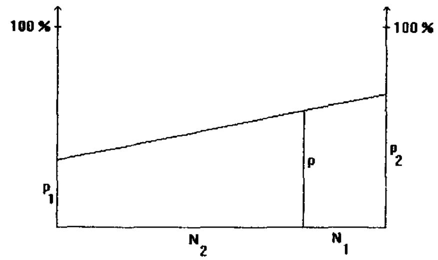
Beim Übergang zur „Zusammenveranlagung“ handelt es sich strukturell um dasselbe Problem wie:

N_1 Liter einer p_1%igen Säure werden mit N_2 Litern einer p_2%igen Säure zusammengeschüttet:

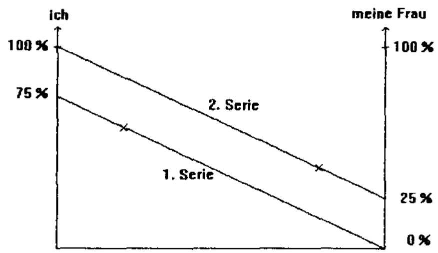


Als Ergebnis bekommt man (N_1 + N_2) Liter p%ige Säure. Wegen $\frac{p_1 \cdot N_1}{100} + \frac{p_2 \cdot N_2}{100} = \frac{p \cdot (N_1 + N_2)}{100}$ ist $p = \frac{p_1 \cdot N_1 + p_2 \cdot N_2}{N_1 + N_2}$.

Diese Formel drückt ein Teilverhältnis aus und läßt sich daher gut visualisieren:



Mithin läßt sich auch das Simpson-Paradoxon gut veranschaulichen; auf den Achsen werden jeweils die B-Prozente des letzten Beispiels angetragen:



Bei meiner Frau und mir hat sich die Präferenz für B jeweils um 25% gesteigert. Betrachtet man uns zusammen, so entsprechen die Kreuze unseren Gemeinschaftswerten; die Gesamtpräferenz für B ist also gefallen. - Diese geometrische Interpretation folgt A. Tan (1986).

Dies alleine führt aber im Unterricht noch keineswegs zum angestrebten Durchblick; ein komisches Gefühl bleibt trotz des schönen Diagramms bei den Schülerinnen und Schülern bestehen.

Dem wird man durch *weitere extreme* Beispiele begegnen:

In einem Wahlkreis steigert sich Partei A von einer Wahl zur nächsten von 98% auf 99%; gleichzeitig hat sich aber die Wahlkreisgröße von 1000 Einwohnern auf 100 verringert. In einem zweiten Wahlkreis ist Partei A von 1% auf 2% gewachsen; dort hat sich aber die Wahlkreisgröße von 100 Einwohnern auf 1000 vergrößert.

Hier sieht man, daß absolute Verluste prozentuale Zuwächse sein können:

		1. Wahlkreis	2. Wahlkreis	Zusammen
1. Wahl	Wahlkreisgröße	1000	100	1100
	absolut für Partei A	980	1	981
	relativ für Partei A	98%	1%	89%
2. Wahl	Wahlkreisgröße	100	1000	1100
	absolut für Partei A	99	20	119
	relativ für Partei A	99%	2%	11%

Daß lokales prozentuales Wachstum von globalem prozentualen Wachstum völlig unabhängig ist, läßt sich durch weitere Beispiele klar machen:

a. Ein schlechter Gymnasiast wechselt zur Realschule. Dadurch verbessert sich der Notendurchschnitt an beiden Schulen.

b. Ein Mischling aus Kenia wandert nach Deutschland aus. Dadurch wird die durchschnittliche Hautfarbe der Bevölkerung in beiden Ländern dunkler.

c. Man habe zwei Parteien A und B sowie zwei verschieden große Wahlkreise; der erste bestehe aus 3 und der zweite aus 6 Wählern:

1. Wahlkreis	2. Wahlkreis
B A B	A B A A A A
$h_A \approx 33\%$	$h_A \approx 83\%$

Nun wird die Wahlkreiseinteilung geändert:

1. Wahlkreis	2. Wahlkreis
B A B A B	A A A A
$h_A \approx 40\%$	$h_A \approx 100\%$

Durch die Änderung haben sich die Prozentzahlen für A in beiden Wahlkreisen erhöht, obwohl sich insgesamt die Zahlen überhaupt nicht verändert haben!

Bei diesen Beispielen handelt es sich um *Umverteilung*. Diese läßt sich an einem kleinen Beispiel separat untersuchen:

d. Zwei Fachleiter teilen sich 3 Referendare. Fachleiter A hat einen sehr guten und einen guten Referendar; Fachleiter B hat einen befriedigenden Referendar.

1 2 / 3

Die Durchschnittsnote bei A beträgt $s_A = 1{,}5$; bei B beträgt sie $s_B = 3$.

d.1. Nun wird die Gruppeneinteilung geändert, und zwar

von 1 2 / 3 $\quad s_A = 1{,}5 \quad s_B = 3$

zu 1 / 2 3 $\quad s_A = 1 \quad s_B = 2{,}5$

Die lokalen Notendurchschnitte werden besser, obwohl der globale Notendurchschnitt gleich bleibt. Dieser Sachverhalt ist die *Urform des Simpsonschen Paradoxons*.

Bemerkung: Übrigens hat man wieder nur mit den Zahlen 1, 2 und 3 gespielt. Dieses banale Spielmaterial verbirgt offensichtlich mehr Tiefen und Abgründe, als man zunächst vermuten möchte!

d.2. Wenn auch der globale Schnitt sich ändern soll, muß die Anzahl oder die Größe der Urelemente geändert werden. Hier betrachten wir eine Änderung

von 1 2 / 3 $s_A = 1{,}5$ $s_B = 3$

zu 1 / 2 3,3 $s_A = 1$ $s_B = 2{,}65$

Die lokalen Schnitte werden besser, obwohl der globale Schnitt sich verschlechtert.

d.3. Dies läßt sich nach dem Motto „Wir werden alle schlechter, es darf nur nicht auffallen" noch etwas spektakulärer gestalten: Die Änderung sei

von 1 2 / 3 $s_A = 1{,}5$ $s_B = 3$

zu 1,3 / 2,3 3,3 $s_A = 1{,}3$ $s_B = 2{,}8$

Alle Noten werden schlechter, alle lokalen Schnitte werden besser, der globale Schnitt wird wieder schlechter.

Man sieht: Umstrukturierung kann DAS Zaubermittel der Politik sein! Durch Neuzuschnitt von Regionen, Wahlkreisen, ..., durch Neudefinition von "reich", "arbeitslos", ... kann unabhängig von den wirklichen Gegebenheiten immer eine (regional zu messende) Verbesserung vorgegaukelt werden!

d.4. Gewissermaßen eine Vorstufe zum Simpsonschen Paradoxon ist die folgende Umverteilung

von 1 2 / 2 $s_A = 1{,}5$ $s_B = 2$

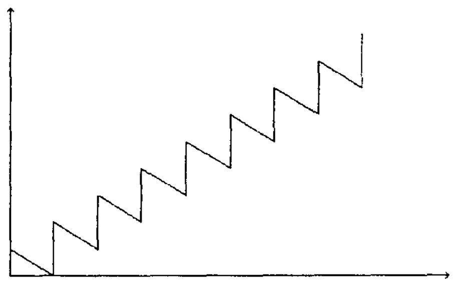
zu 1 / 2 2 $s_A = 1$ $s_B = 2$

Hier verbessert sich nur ein lokaler Schnitt.

Zusammenfassung: *Wer lokal gewinnt, kann global verlieren.*

Das heißt: Eine Größe kann sich in allen Teilbereichen verbessern, obwohl sie sich insgesamt verschlechtert. Dies ist die 5. Interpretation von „lokal", „global" und „besser".

Übungsaufgabe: Geht es bei dem folgenden Graphen aufwärts oder abwärts?



6. Zusammenarbeit

Eine Würfelspalte wie

A
C
B

läßt sich auch noch ganz anders deuten. Dazu seien A und B zwei Schüler und C ein Aufgabentyp. Dann soll die Spalte bedeuten, daß Schüler A den Aufgabentyp C beherrscht, Schüler B dagegen nicht.

In einer Klassenarbeit seien nun die beiden Aufgabentypen C und D mit jeweils gleicher Punktzahl zu erwarten. Die beiden Schüler wissen, daß A den einen Typ kann und B den anderen:

A B
C D
B A

Beide Schüler erwarten also eine Vier und beschließen daher, voneinander abzuschreiben. Der Lehrer merkt das nicht, und beide erhalten eine Eins. Zusammenarbeit lohnt sich.

In der nächsten Klassenarbeit sind wieder die Typen C und D zu erwarten, zusätzlich werden aber vorher auch noch die Aufgabentypen E, F, G durchgenommen. Die Schüler lernen zusammen und können beide E und F, aber bei G haben beide Schwierigkeiten. Das Fähigkeitsprofil vor der nächsten Arbeit ist also

A B	A, B	A, B	
C D	E	F	G
B A			A, B

A kann für sich allein 3/5 aller Aufgaben lösen, und dasselbe gilt für B. Beide sind also seit der letzten Arbeit besser geworden. Aber ihre Zusammenarbeit lohnt sich nicht mehr so wie früher, da sie gemeinsam nur noch 80% aller Punkte erreichen werden.

Man sieht: Obwohl sie einzeln besser wurden, haben sie sich zusammen verschlechtert.

Und dies ist wieder eine neue Interpretation der

Zusammenfassung: *Wer lokal gewinnt, kann global verlieren.*

7. Schlußbemerkungen

...weil die ursprüngliche Haltung des Menschen nicht die Erfassung des Konkreten ist, sondern die Bildung eines Abstrakten, das manchmal unangemessen und inkohärent ist. Revuz.

Die behandelten Beispiele haben mit dem Hauptproblem der beschreibenden Statistik zu tun: Wie beschreibt man eine Überfülle von Daten durch wenige Kenngrößen?

Hiermit ist auch einer der zentralen Punkte der Anwendungsorientierung bzw. der Modellbildung angesprochen. Ziel sind sinnvolle Aussagen über die Welt. Nun „besteht" die Welt aus einer Überfülle von Daten, deren Umfang zunächst einmal radikal reduziert werden muß, wenn man überhaupt zu Aussagen kommen will. (Das menschliche Gehirn ist ein informationsverarbeitendes und damit ein informationsreduzierendes Organ; man vergleiche dazu auch das Seneca–Zitat zu Beginn.) Natürlich hofft man, die Welt auf die relevanten Teile reduziert zu haben.

Reduktionen sind mithin für jede Wissenschaft und für jedes Planen von Handlungen notwendig, aber natürlich weder eindeutig (verschiedene Leute halten verschiedene Aspekte für relevant) noch sicher (man kann nie beweisen, mit der erfolgten Reduktion wirklich die relevanten Teile getroffen zu haben). Insbe-

sondere haben Modelle keinerlei absolute Bedeutung; sie sind nicht objektiv begründbar.

So, wie verschiedene Modelle die Welt verschieden reduzieren und damit beschreiben, werden Daten in der beschreibenden Statistik durch wenige Größen beschrieben und damit in ihrem Umfang reduziert, was auf ganz unterschiedliche Art und Weise geschehen kann, je nach dem, welche Aspekte man für relevant hält.

Das wird durch die angeführten Beispiele erläutert: Im Begriff „besser" (bzw. in „gut") steckt eine (durch entsprechende Spielregeln determinierte) Reduktion; eine absolute Bedeutung hat dieser Begriff nicht.

Und auch die folgende Parallele wird sichtbar: So, wie es bei der Welt Standardmodelle gibt, deren Grenzen auf den ersten Blick nicht erkennbar sein mögen, existiert auch eine Alltagsvorstellung von „besser", deren Unangemessenheit in vielen Bereichen durch diesen Aufsatz hoffentlich deutlich geworden ist.

Literatur

Blyth, Colin R.: Some probability paradoxes in choice from among random alternatives. Journ. Amer. Statist. Assoc. **67** (1972), 366-373.

Lichtenberg, G. Chr.: Sudelbücher 2 (Hrsg.: W. Promies). München: Hanser ^{3}1991.

Gardner, Martin: Nontransitive dice and other probability paradoxes. In: Wheels, Life, and other mathematical amusements. New York: W. H. Freeman & Co. 1983.

Gardner, Martin: Nontransitive paradoxes. In: Time travel and other mathematical bewilderments. New York: W. H. Freeman & Co. 1988.

Gardner, Martin: Induction and probability. In: Time travel and other mathematical bewilderments. New York: W. H. Freeman & Co. 1988.

Garfunkel, S. / Steen, L.(Hrsg.): Mathematik in der Praxis. Heidelberg: Spektrum 1989.

Niemi, R. / Riker, W.: The choice of voting systems. Scientific American, June 1976.

Poundstone, W.: Labyrinths of reason. Harmondsworth: Penguin Books 1991.

Revuz, A.: Mathematikunterricht und anwendbare Mathematik. MNU **30**(5) (1977), 257-263.

Seneca: 94. Brief an Lucilius. Hamburg: Meiner 1993.

Tan, A.: A geometric interpretation of Simpson's paradox. Coll. Math. Journ. **17**(4) (1986), 340-341.

Winter, H.: Zur intuitiven Aufklärung probabilistischer Paradoxien. JMD **13** (1992), 23-54.

Jörg Meyer
Schäfertritt 16
31789 Hameln
Tel.: 05151/54236