ABHANDLUNGEN

Eine neue Sicht der Bayesschen Regel

von Wolfgang Riemer, Köln

Der Artikel ist eine stark verkürzte Darstellung von Kapitel 1 des Buches: W. Riemer, "Neue Ideen zur Stochastik", BI-Wissenschaftsverlag, Mannheim 1985. Der interessierte Leser findet dort einen neuen curricularen Ansatz zur Stochastik mit vielen ausgeführten Beispielen.

"Wir haben die Hypothese H auf dem 5 % Signifikanzniveau verworfen, also gilt die Alternative H mit 95 %iger Sicherheit." Solche von Lernenden immer wieder ausgesprochenen Sätze dokumentieren ein tiefes Unverständnis der Testtheorie, welches auf Versäumnisse in früherliegenden Stufen des Stochastikunterrichts schließen läßt:

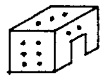
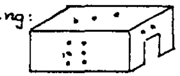
1. Versäumnis:
Wir sprechen auf unteren Curriculumstufen nie von Hypothesen; Wahrscheinlichkeiten werden genau berechnet, es gibt stets nur eine richtige Lösung.

2. Versäumnis:
Wir sprechen (auch auf höheren Curriculumstufen) nie über die subjektiven Wahrscheinlichkeiten, die wir für verschiedene Hypothesen besitzen (im obigen Zitat 95 %).

Hier soll bewußt gemacht werden, wie man durch iterierte Anwendung der Bayesschen Regel schon früh Hypothesen gegeneinander abwägen und die Testtheorie genetisch vorbereiten kann, so daß die oben erwähnten Fehlvorstellungen nicht auftreten können.

I. Hypothesen (ein Beispiel)

Man sägt (aus Profilleisten) teilweise symmetrische Objekte, hier "kurze und lange Zinkenwürfel". Nur aus der Geometrie der Objekte leiten Schüler vor jeglichem Experiment Vermutungen über das Verhalten der Würfel ab (erste Zeile H_1, quantifiziert durch H_2 und H_3):

Tab. 1			Kurz: 						lang: 						Glaubwürdigkeit vor / nach Versuch: für 'kurz':	
Hypothese von:			1	2	3	4	5	6	1	2	3	4	5	6		
allen	H_1	erwartete Reihenfolge	d	c	b	a	c	d	c	d	b	a	d	c	++	+++
Ingo	H_2	erwartete relat. Häufigkeit (%)	4	16	25	35	16	4	10	5	20	50	5	10	0	–
Nele	H_3	" " " (%)	3	8	35	43	8	3	6	5	34	44	5	6	0	+
		relat. Häufigkeit (n = 250) (%)	5,6	14	22,8	33,6	20	4	7,2	5	28,4	44	5	10,5		
vielen	H_4	verbesserte Erwartung (%)	4,5	17,5	22	34	17,5	4,5	9	5	28	44	5	9		+++

Die Zahlen drücken Erwartungen über relative Häufigkeiten aus:
Wir nennen sie Wahrscheinlichkeitshypothesen. Ein Experiment zwingt uns, Neles Hypothese H_3 (für kurz) zu verwerfen, sie sagt die relativen Häufigkeiten sehr viel schlechter voraus als die Hypothese von Ingo. Ingos Hypothese ist recht brauchbar, wir können sie akzeptieren oder unter Verwendung der experimentellen Ergebnisse verbessern (H_4). In der Tat werden wir bei der (folgenden) Anwendung der Bayesschen Regel annehmen, daß die Würfel durch H_4 beschrieben werden.

Wir halten fest:

- Mit Wahrscheinlichkeiten (Hypothesen) machen wir Vorhersagen.
- Manche Hypothesen halten wir für wahrscheinlicher als andere, manche müssen wir ganz verwerfen (in der Tabelle +, -, ++ ...)
- Wahrscheinlichkeiten sind von relativen Häufigkeiten prinzipiell verschieden: alle Wahrscheinlichkeitsverteilungen werden bzgl. der Zahlen 1-6 und 2-5 symmetrisch sein, für die relativen Häufigkeitsverteilungen wird das nur in Ausnahmefällen gelten (vgl. Tabelle 1).

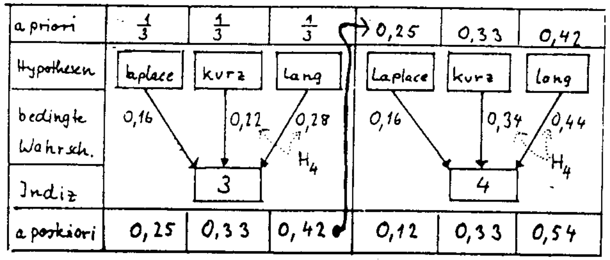
Die ausschließliche Verwendung "klassischer Zufallsobjekte" wie Münze, Würfel, Glücksrad und Reißnagel halte ich für fatal:
Entweder endet man wegen totaler Symmetrie (Münze, Würfel, Glücksrad) sofort bei Laplace-Wahrscheinlichkeiten, es verschwindet der Gedanke, daß es verschiedene Hypothesen gibt, oder man endet wegen der vollständigen Asymmetrie (Reißnagel) bei der Definition von Wahrscheinlichkeit als Grenzwert relativer Häufigkeiten.

Es verschwindet der Gedanke, daß Wahrscheinlichkeiten Vorhersagen machen und von relativen Häufigkeiten prinzipiell verschieden sind.
(In der Tat beobachtet man, daß Kinder von "relativen Wahrscheinlichkeiten" sprechen. Die Konzepte "relative Häufigkeit" und "Wahrscheinlichkeit" als Erwartung für relative Häufigkeit werden nur schwer voneinander getrennt. Die Verwendung teilweise symmetrischer Objekte ist also didaktisch sehr hilfreich.)

II. Iterierte Anwendung der Bayesschen Regel

Das Aufstellen von Hypothesen ist nicht Selbstzweck, wir können mit ihnen arbeiten. In der Tat werden wir jetzt mit der Hypothese H_4 aus Tab. 1 arbeiten:

Ein Partner wählt zufällig zwischen einem unserer Zinkenwürfel und einem Laplacewürfel. Er teilt uns die gewürfelten Ergebnisse mit, wir haben zu entscheiden, um welches Objekt es sich handelt. Anfangs halten wir wegen der zufälligen Wahl jedes der drei Objekte für gleich wahrscheinlich. (Jeweils subjektive Wahrscheinlichkeit $\frac{1}{3}$, vgl. die erste Zeile in Tab. 3). Dann wird uns das Indiz "3" mitgeteilt. Es spricht am meisten für den langen Zinkenwürfel, bei dem dieses Indiz mit 28 % (gegenüber 16,6 % beim Laplacewürfel und 22 % beim kurzen Zinkenwürfel) am häufigsten auftritt. Die Bayessche Regel, "verborgen" in einem umgekehrten Baumdiagramm (vgl. die o.a. Literatur), vergleicht diese Wahrscheinlichkeiten miteinander und liefert die a posteriori Wahrscheinlichkeiten, mit der die "3" von den einzelnen Objekten stammt:



a posteriori Wk. für "lang" nach Indiz "3"

$$\frac{\frac{1}{3}\cdot 0,28}{\frac{1}{3}\cdot 0,16+\frac{1}{3}\cdot 0,22+\frac{1}{3}0,28}\approx 0,42$$

a posteriori Wk. für "lang" nach Indiz "4"

$$\frac{0,42\cdot 0,44}{0,25\cdot 0,16+0,33\cdot 0,34+0,42\cdot 0,44}\approx 0,54$$

Tab. 2

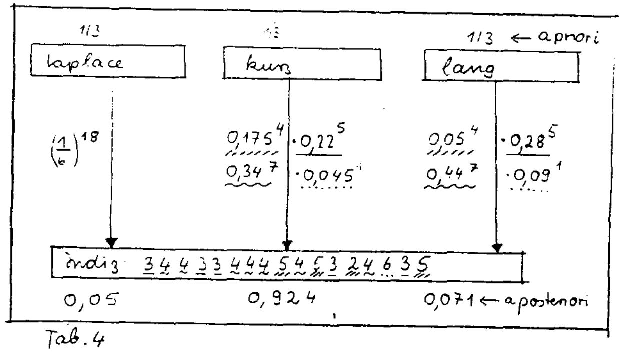
Die a posteriori Verteilung ist gleichzeitig die a priori Verteilung für die nächste Revision, bei der das Indiz "4" erschien...

Im Zuge weiterer Revisionen ergibt sich die nachfolgende Tabelle. Nach 18 Versuchen sind wir mit 92,4 % ziemlich sicher, daß der kurze Zinkenwürfel vorliegt. Hätten wir uns schon nach 8 Versuchen mit 88 % für den langen Zinkenwürfel entschieden, wäre uns ein Fehler unterlaufen. (Fehler erster und zweiter Art).

Man kann zeigen: Das gleiche Endergebnis erhält man durch eine einzige Revision nach Bayes durch das zusammengesetzte Indiz 3-4-4-3-3-4-4-4-5-4-5-3-2-4-6-3-5, man berechnet die bedingten Wahrscheinlichkeiten dann mit der Pfadregel:

Tab. 3:

Indiz	Laplace %	kurz %	lang %	
	33,3	33,3	33,3	Vgl. Tab. 2
3	25,0	33,0	42,0	
4	12,2	33,1	54,7	
4	5,5	30,2	64,3	
3	3,6	26,0	70,4	
3	2,3	22,0	75,7	
4	0,9	18,0	80,9	
4	0,4	15,0	84,6	
4	0,1	11,8	88,0	
5	0,4	31,8	67,8	
4	0,2	26,5	73,3	
5	0,3	56,0	43,7	
3	0,2	50,0	49,8	
2	0,3	77,4	22,3	
4	0,1	72,8	27,1	
2	0,2	90,2	9,6	
6	0,6	82,0	17,4	
3	0,4	78,4	21,2	
5	0,5	92,4	7,1	



Tab. 4

Ergebnis:

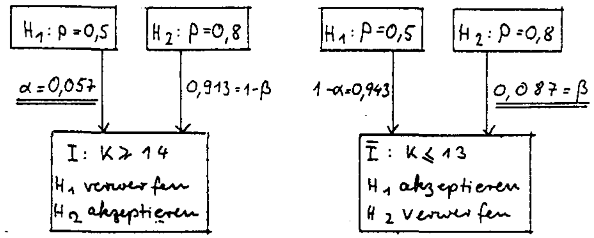
- Durch iterierte Anwendung der Bayesschen Regel können wir quantitativ fundierte Entscheidungen zwischen beliebigen Hypothesen treffen.
- Weil man trotz "komplizierter" Hypothesen mit der Pfadregel auskommt, insbesondere keine Testgrößen benötigt, ist das Verfahren schon für untere Curriculumstufen geeignet (Klasse 8).

III. Testen

Große Verständnisschwierigkeiten bereitet die unterschiedliche Behandlung von H und $\overline{H}$, wenn man sie in der Testtheorie zum ersten Mal kennenlernt. Im Bayesschen Rahmen läßt sich der Gedanke frühzeitig dadurch vorbereiten, daß man die Hypothesen mit sehr verschiedenen a priori Verteilungen bewertet. So betrachten wir in der folgenden Tabelle zwei Bayesianer, die die Hypothesen H_1:p=0,5 und H_2:p=0,8 für die Trefferwahrscheinlichkeit in einer Bernoullikette gegeneinander abwägen. Der erste hält H_1 für sehr viel wahrscheinlicher ($p(H_1)=0,9$), der zweite H_2 ($p(H_2)=0,9$). In der Deutung der Statistiker testet der erste H_2 gegen H_1, der zweite H_1 gegen H_2. Nach 18 Versuchen ist jeder noch sehr von seiner eigenen Hypothese überzeugt, aber man beoabachtet, wie sich mit wachsender Versuchszahl die Hypothese (H_2) durchsetzt, die die relative Trefferhäufigkeit besser vorhersagt. Das ist eine Bayesche Interpretation zum schwachen Gesetz der großen Zahlen.

n	Treffer	relative Trefferhäufigkeit	1. PERSON: 1.HYP 0.500	1. PERSON: 2.HYP 0.800	2. PERSON: 1.HYP 0.500	2. PERSON: 2.HYP 0.800
a priori:			0.900	0.100	0.100	0.900
1	1	1.000	0.849	0.151	0.065	0.935
2	1	1.000	0.779	0.221	0.042	0.958
3	0	0.667	0.898	0.102	0.098	0.902
4	0	0.500	0.956	0.044	0.213	0.787
5	1	0.600	0.932	0.068	0.145	0.855
6	1	0.667	0.896	0.104	0.096	0.904
7	1	0.714	0.843	0.157	0.062	0.938
8	1	0.750	0.770	0.230	0.040	0.960
9	0	0.667	0.893	0.107	0.094	0.906
10	1	0.700	0.840	0.160	0.061	0.939
11	1	0.727	0.766	0.234	0.039	0.961
12	1	0.750	0.672	0.328	0.025	0.975
13	1	0.769	0.561	0.439	0.016	0.984
14	0	0.714	0.762	0.238	0.038	0.962
15	1	0.733	0.666	0.334	0.024	0.976
16	1	0.750	0.555	0.445	0.015	0.985
17	1	0.765	0.438	0.562	0.010	0.990
18	0	0.722	0.661	0.339	0.024	0.976
19	1	0.737	0.549	0.451	0.015	0.985
20	1	0.750	0.433	0.567	0.009	0.991
21	1	0.762	0.323	0.677	0.006	0.994
22	1	0.773	0.229	0.771	0.004	0.996
23	1	0.783	0.157	0.843	0.002	0.998
24	1	0.792	0.104	0.896	0.001	0.999
25	1	0.800	0.068	0.932	0.001	0.999
26	1	0.808	0.043	0.957	0.001	0.999
27	1	0.815	0.028	0.972	0.000	1.000
28	1	0.821	0.017	0.983	0.000	1.000
29	1	0.828	0.011	0.989	0.000	1.000
30	1	0.833	0.007	0.993	0.000	1.000
31	1	0.839	0.004	0.996	0.000	1.000
32	0	0.813	0.011	0.989	0.000	1.000
33	0	0.788	0.026	0.974	0.000	1.000
34	1	0.794	0.017	0.983	0.000	1.000
35	0	0.771	0.041	0.959	0.001	0.999
36	0	0.750	0.096	0.904	0.001	0.999
37	1	0.757	0.062	0.938	0.001	0.999
38	1	0.763	0.040	0.960	0.001	0.999
39	1	0.769	0.025	0.975	0.000	1.000
40	1	0.775	0.016	0.984	0.000	1.000
41	1	0.780	0.010	0.990	0.000	1.000
42	1	0.786	0.006	0.994	0.000	1.000
43	1	0.791	0.004	0.996	0.000	1.000
44	1	0.795	0.002	0.998	0.000	1.000
45	1	0.800	0.002	0.998	0.000	1.000
46	1	0.804	0.001	0.999	0.000	1.000
47	1	0.809	0.001	0.999	0.000	1.000
48	1	0.813	0.000	1.000	0.000	1.000
49	1	0.816	0.000	1.000	0.000	1.000
50	0	0.800	0.001	0.999	0.000	1.000
100		0.810	0.000	1.000	0.000	1.000
150		0.807	0.000	1.000	0.000	1.000
200		0.815	0.000	1.000	0.000	1.000
			Bayesianer, testet H_2 gegen H_1		**Bayesianer, testet H_1 gegen H_2**	

Ein "Testtheoretiker" hat H_1:p=0,5 zugunsten von H_2:p=0,8 mit dem kritischen Bereich I: x > 14 auf dem 5,7 % Signifikanzniveau verwerfen können (n=20). Ein Schüler, der die Bayessche Regel verstanden hat, kann diese Aussage sehr viel fundierter interpretieren als ein unkundiger Klassenkamerad, von dem das eingangs erwähnte Zitat stammt. Er zeichnet die folgenden Baumdiagramme:



	1. Person		2. Person	
H	H_1	H_2	H_1	H_2
p(H)	0,9	0,1	0,1	0,9
p(H|J)	$\boxed{0,36}$	0,64	$\boxed{0,007}$	0,993
p(H|J̄)	0,99	0,01	0,55	0,45
$\alpha p(H_1)+\beta p(H_2)$	0,061		0,084	

Tab. 5

und erkennt, daß aus dem Indiz I: x>14 keine Wahrscheinlichkeitsaussagen über die Tültigkeit von Hypothesen abzuleiten sind. Indizien können höchstens zur Revision schon vorhandener subjektiver Wahrscheinlichkeitsverteilungen benutzt werden. Nehmen wir in der rechtsstehenden Tabelle sehr verschiedene a priori Verteilungen an und revidieren wir nach Bayes, wird deutlich, daß I sehr

für H_2 und gegen H_1 spricht. Auf die Kernfrage nach der Irrtumswahrscheinlichkeit liefert unsere Tabelle sehr eingehende Antwort, wir haben die aus der Bayesschen Regel vertrauten a posteriori Wahrscheinlichkeiten unter Beachtung der Entscheidungsfunktion des Indizes nur neu zu interpretieren:

- Das Signifikanzniveau α ist die bedingte Wahrscheinlichkeit eines Irrtums, wenn wir voraussetzen, daß H_1 gilt. Analog für β .
- Deswegen ist $p(H_1)\alpha + p(H_2)\beta$ die a priori Gesamtwahrscheinlichkeit eines Irrtums.
- $p(H_1|I)$ ist die Wahrscheinlichkeit für H_1, nachdem I eingetreten ist, also H_1 abgelehnt wurde (= Wahrscheinlichkeit, daß die nach I gefällte Entscheidung, H_1 abzulehnen, falsch war). Es ist genau wie $p(H_2|\bar{I})$ eine a posteriori Irrtumswahrscheinlichkeit.

Anschrift des Verfassers:
Dr. Wolfgang Riemer, Kafkastr. 1, 5000 Köln 30