Erfahrungen mit zwei Leistungskurs-Abituraufgaben

von Heinz Althoff

Unterrichtliche Voraussetzungen

Die Prüflinge hatten etwas mehr als ein Halbjahr Stochastikunterricht. Dabei wurde das Lehrbuch /Althoff 1985a/ praktisch vollständig durchgearbeitet, wobei die Schüler auch das zugehörige Lösungsheft /Al 1985b/ regelmäßig benutzt haben. An einigen Stellen erfolgten noch inhaltliche Ergänzungen, meistens ausgehend von Schülervorträgen.

Als Hilfsmittel standen den Prüflingen zur Bearbeitung dieser beiden Aufgaben die lose Formelsammlung aus /Althoff, 1985a/, die Tabellen /Barth, 1981/ und ein nicht programmierbarer Taschenrechner zur Verfügung.

Die beiden folgenden Stochastikaufgaben bildeten zusammen mit einer Analysisaufgabe die Abiturklausur 1995 in meinem Leistungskurs am Helmholtz-Gymnasium Bielefeld. Die Arbeitszeit betrug 5 Zeitstunden.

Aufgabenstellungen

Aufgabe 1:

Die Krankheit K wird entweder durch den Erreger E_1 oder durch den (von E_1 schwer unterscheidbaren) Erreger E_2 verursacht. Das Medikament M_1 heilt die Krankheit K bei Vorliegen des Erregers E_1 bei 30% der Erkrankten und bei Vorliegen des Erregers E_2 bei 90% der Erkrankten. Das Medikament M_2 heilt K bei Vorliegen des Erregers E_1 bei 60% und bei Vorliegen des Erregers E_2 bei 40% der Erkrankten.

a) Wir gehen in den Teilaufgaben a) und b) von der Annahme aus, daß K in 40% der Fälle von E_1 und in 60% der Fälle von E_2 verursacht wird.

a_1) Ein Patient ist an K erkrankt; es ist nicht bekannt, ob die Krankheit durch E_1 oder durch E_2 verursacht wurde. Mit welcher Wahrscheinlichkeit wird er bei einer Behandlung mit M_1 geheilt?

a_2) Ein Patient wird mit Hilfe von M_2 geheilt. Mit welcher Wahrscheinlichkeit hatte er dann den Erreger E_1?

b) Das Medikament M_1 heilt erfahrungsgemäß die Krankheit K bei etwa zwei Drittel aller Patienten. Berechnen Sie damit die Wahrscheinlichkeit der folgenden Ereignisse:

A: Von 10 an K erkrankten und mit M_1 behandelten Patienten werden mindestens 7 geheilt.

B: Von 200 an K erkrankten und mit M_1 behandelten Patienten werden mindestens 70% geheilt.

C: Unter 4 an K erkrankten und mit M_1 behandelten Patienten sind genau 2, die den Erreger E_1 haben und geheilt werden.

c) Wir wollen jetzt die in a) gemachte Annahme, daß der Anteil p derjenigen Patienten, deren Krankheit von E_1 verursacht wurde, etwa 40% beträgt, überprüfen. Dazu werden 1.000 an K erkrankte Personen mit dem Medikament M_1 behandelt; es stellt sich heraus, daß 657 von ihnen geheilt werden. Ermitteln Sie aus den bekannten Daten (zweckmäßig unter Benutzung eines Baumdiagramms) einen geeigneten Schätzwert für p.

d) Ein Arzt mißtraut sowohl dem in a) angenommenen als auch dem in c) ermittelten Schätzwert für p. Er will ein Verfahren benutzen, bei dem die Heilwahrscheinlichkeit für die Krankheit K gar nicht von p abhängt.
Dazu verordnet er einem (zufällig ausgewählten) Anteil r der an K erkrankten Patienten das Medikament M_1 und den restlichen an K erkrankten Patienten das Medikament M_2.

d_1) Geben Sie die Wahrscheinlichkeit P(H), mit der ein an K erkrankter Patient geheilt wird, in Abhängigkeit von p und r an.

d_2) Schreiben Sie P(H) dann in der Form $P(H) = f_1(r) \cdot p + f_2(r)$, wobei $f_1(r)$ und $f_2(r)$ Terme (nur) mit der Variablen r sind.

d_3) Welchen Wert muß jetzt $f_1(r)$ annehmen, damit P(H) von p unabhängig ist?

d_4) Welche Heilwahrscheinlichkeit P(H) erreicht der Arzt mit seinem Verfahren?

Aufgabe 2:

Eine Seuche S hat in einem ostasiatischen Land eine Verbreitung von 3%.

a) Mit welcher Wahrscheinlichkeit befinden sich dort in einer Zufallsstichprobe von 200 Personen

a_1) höchstens 5,

a_2) 3% bis 4% an S erkrankte Personen?

b) In einer Zufallsstichprobe vom Umfang n soll mit einer Mindestwahrscheinlichkeit von 98% wenigstens eine an S erkrankte Person vorkommen. Wie groß muß dann n sein?

c) Ein altes Medikament gegen die Seuche hat die Heilungswahrscheinlichkeit p = 10%. Eine Arzneimittelfabrik bringt das neue Medikament „Seuchin" auf den Markt und behauptet, daß dieses eine höhere Heilungswahrscheinlichkeit hat als das alte Medikament. Diese Hypothese soll nun anhand einer Stichprobe vom Umfang n = 100 auf dem Signifikanzniveau 1% abgesichert werden.

c_1) Warum wählt man für den Test H_0: p = 0,10 und $\overline{H_0}$: p > 0,10?

c_2) Ermitteln Sie für die vorgegebenen Daten die optimale Entscheidungsregel.

c_3) Skizzieren Sie den Graphen der zugehörigen Gütefunktion g_1 mit $g_1(p) = 1 - B(100; p; 18)$ für $0{,}1 \le p \le 1$.

c_4) Was gibt $g_1(0{,}15)$ an, was $1 - g_1(0{,}15)$? Formulieren Sie den Sachverhalt (auch) inhaltsbezogen auf das vorliegende Sachproblem.

d) Zum Stichprobenumfang n = 300 gehört auf dem gleichen Signifikanzniveau 1% die Gütefunktion g_2 mit $g_2(p) = 1 - B(300; p; 42)$ für $0{,}1 \le p \le 1$ und nebenstehender Wertetabelle.

p	$g_2(p)$
0,11	0,04
0,12	0,12
0,14	0,47
0,16	0,81
0,18	0,96
0,20	0,99

d_1) Zeigen Sie, wie man die zugehörige Entscheidungsregel ermitteln kann.

d_2) Berechnen Sie $g_2(0{,}15)$. Welche sachbezogenen Konsequenzen hat die Tatsache, daß $g_2(0{,}15) > g_1(0{,}15)$ ist?

d_3) Inwiefern unterscheidet sich der Graph von g_2 deutlich vom Graphen von g_1? Welche inhaltliche Bedeutung hat dies für den konkret vorliegenden Test, insbesondere auch bezüglich des Fehlers 2. Art?

Erwartete Schülerleistungen und Bewertungsvorschlag

Wenn man in Nordrhein-Westfalen die beiden Abiturvorschläge bei der Schulaufsicht einreicht, muß man neben den Lösungen auch einen „Erwartungshorizont" hinzufügen (eine Hilfspunktewertung für die einzelnen Aufgabenteile wird seit einigen Jahren nicht mehr verlangt). Für die obigen beiden Aufgaben habe ich geschrieben:

Die Aufgabe 1, zu deren Formulierung wesentliche Ideen aus der Aufgabe 92|2 im Zentralabitur BW stammen, enthält (im Gegensatz zur Aufgabe 2) nur Probleme aus der Wahrscheinlichkeitsrechnung. Die Hauptaufgabe der Prüflinge besteht (wie bei allen Aufgaben der Stochastik) darin, die gegebenen Angaben zu interpretieren und daraus neue Erkenntnisse zu ziehen unter Anwendung bekannter Verfahren. Bei a) und b) dürfte das den Prüflingen keine größeren Schwierigkeiten bereiten (bis auf den letzten Teil von b). Dagegen wird von den Prüflingen in c) und d) relativ viel Selbständigkeit erwartet, da solche Probleme im Unterricht direkt nicht behandelt wurden, so daß beide Teilaufgaben in den Anforderungsbereich III hineinragen. Bei d) schien es mir sogar erforderlich zu sein, den Lösungsweg in Teilaufgaben aufzuteilen, die zumindest gedanklich unabhängig voneinander bearbeitet werden können.

Die Aufgabe 2, zu deren Zusammenstellung wichtige Anregungen der Aufgabe 5 auf Seite 235 des Lehrbuchs „Stochastik" im Verlag Cornelsen entnommen wurden, enthält im Gegensatz zur Aufgabe 1 mehr Probleme aus der beurteilenden Statistik als aus der Wahrscheinlichkeitsrechnung. Die Teilaufgaben a) und b) erfordern Standardverfahren aus der Wahrscheinlichkeitsrechnung und dürften den meisten Prüflingen keine größeren Schwierigkeiten bereiten, obwohl in b) eine Exponentialungleichung zu lösen ist. Die Teilaufgaben c) und d) zum Hypothesentest stellen (auch vom Umfang her) den wesentlichen Teil der Aufgabe dar. Insbesondere in den letzten Teilen dieser beiden Teilaufgaben geht es um selbständige Interpretationen inhaltsbezogener Ergebnisse, was wohl beim Anforderungsbereich III einzuordnen ist.

Lösung zu Aufgabe 1:

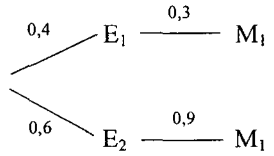
Es sollen bedeuten:

$E_1(E_2)$: Die Krankheit K wird durch den Erreger $E_1(E_2)$ verursacht.

$M_1(M_2)$: Das Medikament $M_1(M_2)$ heilt bei seiner Anwendung K.

a) Die gegebenen Daten werden in zwei Baumdiagrammen zusammengefaßt (die Grundmenge bilden die an K erkrankten Personen), die gesuchten Wahrscheinlichkeiten werden damit dann nach dem bekannten Verfahren berechnet:

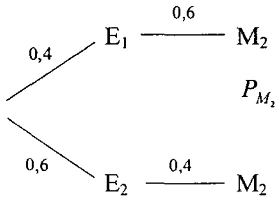
a_1)



$P(M_1) = 0{,}4 \cdot 0{,}3 + 0{,}6 \cdot 0{,}9 = 0{,}66$

Der Patient wird bei Anwendung des Medikaments M_1 mit der Wahrscheinlichkeit 0,66 geheilt.

a_2)



$$P_{M_2}(E_1) = \frac{P(E_1 \cap M_2)}{P(M_2)} = \frac{0{,}4 \cdot 0{,}6}{0{,}4 \cdot 0{,}6 + 0{,}6 \cdot 0{,}4} = 0{,}50$$

Bei einem mit Hilfe des Medikaments M_2 geheilten Patienten wurde die Krankheit mit 50% Wahrscheinlichkeit vom Erreger E_1 verursacht.

b) Die Ereignisse A, B und C können mit Hilfe einer binomialverteilten Zufallsgröße X_n (Anzahl der geheilten Patienten unter n mit M_1 behandelten) beschrieben werden, da die Patienten unabhängig voneinander geheilt werden. Man erhält dann

$$P(A) = P(X_{10} \geq 7) = 1 - B\left(10; \frac{2}{3}; 6\right) = 0{,}56$$

$$P(B) = P(X_{200} \geq 140) = 1 - B\left(200; \frac{2}{3}; 139\right) = 0{,}18$$

$$P(C) = P((X_4 = 2) \cap E_1) = b(4; 0; 12; 2) = \binom{4}{2} \cdot 0{,}12^2 \cdot 0{,}88^2 = 0{,}07$$

c) Im Ansatz von a_1) sind 0,4 durch p, 0,6 durch 1 - p und $P(M_1)$ durch $\frac{657}{1000} =$ 0,657 zu ersetzen. Aus der Gleichung kann man dann durch Auflösung nach p einen Schätzwert für p ermitteln:

$$p \cdot 0{,}3 + (1 - p) \cdot 0{,}9 = 0{,}657 \Leftrightarrow -0{,}6p = -0{,}243 \Leftrightarrow p = \frac{0{,}243}{0{,}6} = 0{,}405$$

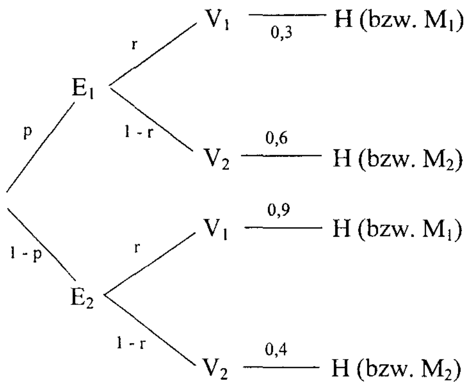
Der in a) benutzte Schätzwert 0,4 paßt also gut zum Versuchsergebnis.

d) Folgende weitere Bezeichnungen werden verwendet:

$V_1(V_2)$: Der Arzt verordnet dem Patienten das Medikament $M_1(M_2)$.

H: Der Patient wird geheilt.

Der Sachverhalt kann dann folgendermaßen in einem Baum dargestellt werden (wobei die Reihenfolge der $E_{1/2}$ und $V_{1/2}$ auch vertauscht werden kann, da sie unabhängig voneinander sind)



Es ergibt sich nun:

d_1) $P(H) = p \cdot r \cdot 0{,}3 + p \cdot (1 - r) \cdot 0{,}6 + (1 - p) \cdot r \cdot 0{,}9 + (1 - p) \cdot (1 - r) \cdot 0{,}4$

d_2) $P(H) = p(0{,}3r + 0{,}6 - 0{,}6r - 0{,}9r - 0{,}4 + 0{,}4r) + (0{,}9r + 0{,}4 - 0{,}4r)$

d_3) $r = 0{,}25 \Rightarrow P(H) = 0{,}5 \cdot 0{,}25 + 0{,}4 = 0{,}53$

<u>Lösung zu Aufgabe 2:</u>

a) X : Anzahl der an S erkrankten Personen in der Stichprobe vom Umfang n = 200. X ist binomialverteilt, da es sich um eine kleine Stichprobe aus einer großen Gesamtheit handelt.

p = 3% ist die Wahrscheinlichkeit, mit der eine zufällig ausgewählte Person in dem Land an der Seuche S erkrankt ist.

a_1) $P(X \leq 5) = B(200; 0{,}03; 5) = 0{,}44$

a_2) $P(6 \leq X \leq 8) = B(200; 0{,}03; 8) - B(200; 0{,}03; 5) = 0{,}41$

b) $P(X \geq 1) \geq 0{,}98 \Leftrightarrow 1 - b\,(n;\,0{,}03;\,0) \geq 0{,}98 \Leftrightarrow \binom{n}{0} \cdot 0{,}03^0 \cdot 0{,}97^n \leq 0{,}02$

$\Leftrightarrow n \cdot \lg 0{,}97 \leq \lg 0{,}02 \Leftrightarrow n \geq \frac{\lg 0{,}02}{\lg 0{,}97} = 128{,}4 \Leftrightarrow n \geq 129$

Die Stichprobe muß aus mindestens 129 Personen bestehen.

c_1) Da die Hypothese $p > 0{,}10$ abgesichert werden soll, muß das Gegenteil als Nullhypothese genommen werden. Da $p < 0{,}10$ hier ohne Bedeutung ist (sowohl inhaltlich als auch rechnerisch, da die Wahrscheinlichkeit für einen Fehler 1. Art bei $p = 0{,}10$ am größten ist), ist es sinnvoll, $H_0 : p = 0{,}10$ als Nullhypothese zu nehmen (statt $p \leq 0{,}10$).

c_2) Eine sinnvolle *Form* der Entscheidungsregel ist

Verwirf $H_0 \Leftrightarrow Y > a$

Dabei ist Y (Anzahl der geheilten Personen unter den n mit „Seuchin“" behandelten) aus den gleichen Gründen wie X binomialverteilt. Gesucht ist die kleinste Zahl $a \in N$ mit $P_{p=0,1}(Y > a) \leq 0{,}01$. Es gilt nun:

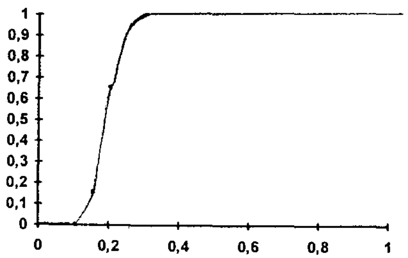
$P_{p=0,1}(Y > a) \leq 0{,}01 \Leftrightarrow 1 - B(100;\,0;1;\,a) \leq 0{,}01 \Leftrightarrow$

$B(100;\,0{,}1;\,a) \geq 0{,}99 \Leftrightarrow a \geq 18$ (evtl. $a \geq 17$ bei anderer Tabelle)

Die optimale Entscheidungsregel auf dem Signifikanzniveau 1% lautet also:

Glaube der Aussage der Arzneimittelfabrik genau dann, wenn von 100 Patienten, die mit dem neuen Medikament behandelt wurden, mehr als 18 geheilt werden.

c_3)



p	$g_1(p)$
0,10	0,005
0,15	0,16
0,20	0,64
0,25	0,94
0,30	0,995
0,35	1,00

c_4) $g_1(0{,}15)$ gibt an, mit welcher Wahrscheinlichkeit man bei diesem Test „Seuchin“ eine höhere Heilungswahrscheinlichkeit zutraut als dem alten Medikament, falls in Wirklichkeit „Seuchin“ eine Heilungswahrscheinlichkeit von 15% hat. $1 - g_1(0,\,15)$ gibt an, mit welcher Wahrscheinlichkeit man bei diesem Test „Seuchin“ irrtümlich keine höhere Heilungswahrscheinlichkeit als dem alten Medikament zutraut, obwohl es in Wirklichkeit eine Heilungswahrscheinlichkeit von 15% hat (und damit deutlich besser ist als das alte Medikament).

d) Für die hier interessanten Werte von p (0,1 bis 0,3) gilt $n \cdot p\,(1 - p) = 300 \cdot p \cdot (1 - p) > 9$, so daß man die binomialen Wahrscheinlichkeiten mit der Laplace-Näherung ermitteln kann.

d_1) $1 - B\,(300;\,0{,}1;\,a) \leq 0{,}01 \Leftrightarrow \Phi\left(\frac{a + 0{,}5 - 30}{\sqrt{27}}\right) \geq 0{,}99 \Leftrightarrow \frac{a - 29{,}5}{\sqrt{27}} \geq 2{,}3264$

$\Leftrightarrow a \geq 2{,}3264 \cdot \sqrt{27} + 29{,}5 \Leftrightarrow a \geq 42$

d_2) $g_2(0{,}15) = 1 - B\,(300;\,0{,}15;\,42) \approx 1 - \Phi\left(\frac{42{,}5 - 45}{\sqrt{45 \cdot 0{,}85}}\right) = \Phi(0{,}404) = 0{,}66$

$g_2(0{,}15) > g_1(0{,}15)$ bedeutet:

Heilt „Seuchin“ in Wirklichkeit 15% der Erkrankten, so wird bei einer Stichprobe vom Umfang 300 (wesentlich) häufiger als bei einer Stichprobe vom Umfang 100 (richtig) bestätigt, daß „Seuchin“ besser ist als das alte Medikament.

d_3) Vergleicht man die Werte von $g_2(p)$ mit denen von $g_1(p)$, so erkennt man, daß der Graph von g_2 im „schrägen“ Teil steiler ist als der Graph von g_1 und deshalb schon bei $p \approx 0{,}2$ (statt bei $p \approx 0{,}3$) fast oben ist. Bei $n = 300$ treten in einem kleineren Bereich für p als bei $n = 100$ größere Wahrscheinlichkeiten für Fehler 2. Art auf. Falls „Seuchin“ also wirklich besser ist als das alte Medikament, so wird dies bei einem Test mit $n = 300$ auf dem vorgegebenen Signifikanzniveau (wesentlich) häufiger abgesichert als bei einem Test mit $n = 100$.

Zu den erreichten Ergebnissen

Hinsichtlich des Schwierigkeitsgrades der Teilaufgaben wurden meine Erwartungen von den Prüflingen im wesentlichen bestätigt. Das zeigt auch die folgende Tabelle, in der ich angegeben habe, wieviel Prozent der Punkte die Prüflinge in den Teilaufgaben *durchschnittlich* erreicht haben.

Aufgabe Nr.	im Durchschnitt erreichte Prüfungsleistung (in %)
1a	88
1b	75
1c	47
1d	30
2a	92
2b	77
2c	76
2d	65

Aufgabe 2, die ich mit 55 Punkten bewertet habe [a) 8, b) 10, c) 20, d) 17], wurde von den drei Aufgaben mit durchschnittlich 75% erreichten Punkten am besten bearbeitet, Aufgabe 1 [insgesamt 45 Punkte, davon a)11, b) 13, c) 9, d) 12] mit durchschnittlich 60% erreichten Punkten am schlechtesten.
In der mit 50 Punkten bewerteten Analysisaufgabe erreichten die Prüflinge im Durchschnitt 67% der Punkte.

Als Problem erwies sich für die Prüflinge der offenbar insgesamt etwas zu große Umfang der Klausur. Keiner aus dem leistungsmäßig überdurchschnittlichen Kurs hat seine Klausur vorzeitig abgegeben, auch die meisten sehr guten Prüflinge haben zum Schluß ein bis zwei Teilaufgaben nur noch teilweise oder gar nicht mehr geschafft. Dies habe ich bei der Bewertung der Leistungen dadurch berücksichtigt, daß ich in einzelnen solchen Fällen (durch Hinweis im Gutachten) die Qualität der vorhandenen Lösungen in den Vordergrund gerückt und die Note „nach oben aufgerundet" habe.

Literatur:

H. Althoff, 1985a: Wahrscheinlichkeitsrechnung und Statistik. Verlag Metzler, Stuttgart 1985.
H. Althoff, 1985b: Wahrscheinlichkeitsrechnung und Statistik, Lösungen. Verlag Metzler, Stuttgart 1985.
F. Barth u.a.: Stochastische Tabellen. Verlag Ehrenwirth, München 1981.

StD Heinz Althoff
Ruschfeldweg 17
33619 Bielefeld