

Schwierigkeiten mit der Nullhypothese

Gunter Stein, Darmstadt

Zusammenfassung: Beim Hypothesentesten werden zwei Hypothesen gegeneinander ausgespielt, und zwar erfolgt die Annahme der Alternativhypothese, wenn das Testergebnis bei Vorliegen der Nullhypothese sehr unwahrscheinlich ist. Dabei verneint die Nullhypothese das, was man eigentlich nachweisen will. Dies in bestimmten Anwendungssituationen zu akzeptieren, kann Schwierigkeiten bereiten. Von sehr anschaulichen und einleuchtenden Anwendungsbezügen abzusehen, und den Effekt, den man eigentlich nachweisen will, vorübergehend für "null und nichtig" zu erklären, darf nicht nur formal irgendwo als Nullhypothese aufgeschrieben werden, sondern muß bei manchen Wahrscheinlichkeitsberechnungen voll inhaltlich akzeptiert und berücksichtigt werden.

Bei der Behandlung der Stochastik in der Oberstufe nach dem Schema *Kombinatorik -> Wahrscheinlichkeitsrechnung -> Zufallsvariable -> Binomialverteilung -> Normalverteilung -> Hypothesentesten*

wird die beurteilende Statistik erst sehr spät und daher oft knapp oder gar nicht behandelt. Aus diesem Grund gibt es Versuche, die beurteilende Statistik möglichst früh im Stochastikunterricht zu thematisieren, indem man sich bei der vorangehenden Stoffbehandlung auf das Nötigste beschränkt. Einen Versuch hierzu findet man bei Engel (1987, S. 32-35), wo schon im 2. Kapitel Grundaufgaben der Statistik thematisiert werden, nachdem zuvor abzählende Kombinatorik und diskrete Wahrscheinlichkeitsräume behandelt werden.

Bei einem so frühen Zeitpunkt innerhalb eines Kurses ist es notwendig, Fragestellungen zu finden, die zum Hypothesentesten weder Binomial- noch Normalverteilung voraussetzen. Solche in diesem Sinn voraussetzungsfreien Testaufgaben sind nicht einfach zu finden. Bei Engel sind es im wesentlichen fünf Aufgaben; bei einer davon geht es um Experten für Geschmacksprüfungen:

5. Ein Experte für Geschmacksprüfungen soll getestet werden, indem ihm zwei Sorten A und B von Wein vorgesetzt werden, die für Laien geschmacklich nicht unterscheidbar sind. Der Experte weigert sich, mehr als 8 Gläser zu kosten, da der Geschmack abstumpft. Wir betrachten 6 Prüfpläne, wobei der Prüfplan dem Experten jeweils mitgeteilt wird.

a) Es werden ihm 4 Paare von Gläsern vorgesetzt. Jedes Paar enthält je ein Glas A- und B-Wein. Er identifiziert alle 4 (nur 3) Paare richtig. Berechne den P-Wert.

b) Es werden ihm je 4 Gläser A- und B-Wein in zufällig permutierter Reihenfolge vorgesetzt. Er identifiziert alle 8 (nur 6) Gläser richtig. Berechne den P-Wert. Kann er genau 7 Gläser richtig identifizieren?

c) Bei jedem der 8 Gläser wird durch Werfen ein L-Münze entschieden, ob es durch A- oder B-Wein gefüllt wird. Er identifiziert 8 (7, 6) Gläser richtig. Berechne den P-Wert.

d) Der Test c ist unfair. Sollte die Münze 8mal 0 oder 1 zeigen, so verliert der Experte den Vorteil des Vergleiches. Daher werden ihm mindestens zwei Gläser jeder Sorte zugebilligt, indem Münzenwürfe mit mindestens 7 Einsen oder Nullen verworfen werden. Berechne den P-Wert bei richtiger Identifizierung von 8 (7) Gläsern.

e) Dem Experten werden drei Gläser jeder Sorte in zufälliger Reihenfolge vorgesetzt, und er identifiziert alle richtig. Berechne den P-Wert.

f) Der Experte wird überredet, 12 Gläser zu prüfen. Es werden ihm 6 Gläser jeder Sorte in zufälliger Reihenfolge vorgesetzt. Er identifiziert 10 richtig. Berechne den P-Wert.

Lösungen:

5. a) $P = \frac{1}{16}$ ($P = \frac{5}{16}$) b) $P = \frac{1}{70}$ ($P = \frac{17}{70}$) Nein, da sich die Sorte des 8. Glases aus den restlichen ergibt.

c) $P = \frac{1}{256}$ ($P = \frac{9}{256}$, $P = \frac{37}{256}$) d) $P = \frac{1}{238}$ ($P = \frac{9}{238}$)

e) $P = \frac{1 + 36}{\binom{12}{6}} = \frac{37}{924} = 0,04 = 4 \%$

Diese Aufgabe ist schon in ähnlicher Form in Engel (1973) enthalten.

Die Besprechung dieser Aufgabe, d. h. die Diskussion der Lösungen, gehört für mich zu dem Mühsamsten, was ich im Stochastikunterricht erlebt habe. Bei kaum einer anderen Fragestellung habe ich soviel reden, argumentieren, richtigstellen müssen, damit die Lösungen akzeptiert wurden. Es gibt eine erstaunliche Vielfalt falscher Lösungen, und die notwendige Auseinandersetzung mit den falschen Denkansätzen kann Lehrern und Schülern die Stochastik gründlich verleiden, wenn man nicht zu einer restlosen Aufklärung kommt - und dies kann unvertretbar viel Unterrichtszeit kosten. Ganz abgesehen davon, daß bei dieser Diskussion das eigentliche Ziel der Aufgabe, Einsicht in das Hypothesentesten, Verständnis für die negative Testlogik, Auswirkungen von Fehlern 1. und 2. Art in der Realität, leicht verloren geht.

Die Schwierigkeiten liegen m. E. nicht so sehr in der Kombinatorik, sondern sie beruhen darauf, daß die Nullhypothese (H) nicht hinreichend beachtet bzw. nicht in den Mittelpunkt der Überlegungen gestellt wird:

Der Experte ist keiner und weiß dies auch und rät bloß, d. h. der gesuchte P-Wert ist die Wahrscheinlichkeit, durch bloßes Raten eine (fast) fehlerfreie Identifizierung der Gläser zu erzielen.

Die Schwierigkeiten mit der Nullhypothese sollen anhand der Aufgaben b, c und d erörtert werden:

b) Es werden ihm je 4 Gläser A- und B-Wein in zufällig permutierter Reihenfolge vorgesetzt. Er identifiziert alle 8 (nur 6) Gläser richtig. Berechne den P-Wert. Kann er genau 7 Gläser richtig identifizieren?

Die Berechnung

$$P(\text{alle 8 Gläser richtig} | H) = \frac{1}{\binom{8}{4}} = \frac{1}{70}$$

paßt überhaupt nicht zu der Vorstellung, daß der zu testende Weinkenner an jedem Glas riecht, probiert und ausspuckt, um nicht durch den Alkoholgenuß eine Bewußtseinstörung zu erleiden, und der eventuell zwischendurch ein Stück Weißbrot kaut, um seine Geschmacksnerven wieder zu neutralisieren.

Wenn das tatsächliche Verkosten des Weines von dem erwähnten, ahnungslosen Laien vorgenommen wird, ist folgende Lösung einleuchtender

$$P(\text{alle 8 Gläser richtig} | H) = \frac{4}{8} \cdot \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{70}$$

d. h. der Geschmackstest wird als 8-Stufenprozeß angesehen: Das erste A-Glas wird mit der Wahrscheinlichkeit $\frac{4}{8}$ richtig geraten, das erste B-Glas mit $\frac{4}{7}$, das zweite A-Glas mit $\frac{3}{6}$ usw.

Man kann auch eine bestimmte Anordnung, z. B. AABABBBA, vorgeben und nach einer fehlerfreien Identifizierung fragen:

$$\begin{array}{cccccccc} A & A & B & A & B & B & B & A \\ \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} & = & \frac{1}{70} \end{array}$$

Möglich wäre auch, daß der ahnungslose Experte aufhört, wenn er die 4 A-Gläser identifiziert hat, denn der Rest sind B-Gläser:

$$P(\text{alle 8 Gläser richtig} | H) = \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{70}$$

Die anschaulich vorstellbare Situation der konkreten Expertenverkosterei macht es schwierig, eine völlige Ahnungslosigkeit des Experten anzunehmen. Die reale Situation wird wohl so vorgestellt, daß der Experte etwas - aber nicht genügend - weiß, daß er um seinen guten Ruf bzw. seine Anstellung kämpft und daher die

Gläser wirklich durchprobiert, nicht aber auf gut Glück eine der $\binom{8}{4} = 70$ mögli-

chen Gläserkonfigurationen benennt. Allenfalls könnte man sich vorstellen, daß der Experte erst einmal kommentarlos alle Gläser durchprobiert, um dann auf eine bestimmte Gläserkonfiguration zu tippen. Eine simultane Erfassung aller Gläser, z. B. bei einer optischen Prüfung, käme dieser Art der Berechnung entgegen.

Zu dieser Aufgabe gibt es viele falsche Lösungen, und wohl jeder Lehrer von Stochastikkursen hat leidvolle Erfahrungen mit mühsamen Diskussionen über falsche Denkansätze. Auch richtige Lösungen können schwer zu durchschauen sein. Bei der folgenden Lösung, die wohl durch die Lösung von Aufgabe a ange-regt wurde, werden die Gläserfüllungen nach der Stellung des letzten A-Glases klassifiziert und die Wahrscheinlichkeit einer richtigen Identifizierung der 4 A-Gläser für jede Klasse einzeln bestimmt:

Gläserfüllung AAAABBBB AAABABBB AAABBABB AAABBBAB

$$\begin{array}{l} \text{Wahr-} \\ \text{scheinlich-} \\ \text{keit} \end{array} \quad 2 \cdot \frac{1}{70} \cdot \left(\frac{1}{2}\right)^4 + 2 \cdot \frac{\binom{4}{1}}{70} \cdot \left(\frac{1}{2}\right)^5 + 2 \cdot \frac{\binom{5}{2}}{70} \cdot \left(\frac{1}{2}\right)^6 + 2 \cdot \frac{\binom{6}{3}}{70} \cdot \left(\frac{1}{2}\right)^7 = \frac{1}{70}$$

Während der Faktor 2 noch leicht durch Vertauschung von A und B zu erklären ist, kann man bei den anderen Zahlen nur hoffen, daß der Problemlöser auch verständlich erklären kann, was er richtig gedacht hat.

Bei solchen Stochastikaufgaben in der Oberstufe wiederholen sich die Schwierigkeiten, die das Sachrechnen in der Mittelstufe begleiten: Zum einen muß bei einer Textaufgabe von den vielfältigen Bezügen zur Wirklichkeit abstrahiert werden, um den mathematischen Kern des Sachverhaltes herausarbeiten zu können; zum anderen müssen die Ergebnisse der mathematischen Modellierung für die reale Ausgangssituation interpretiert werden.

Soll der Hypothesentest zu inhaltlich bedeutsamen Aussagen führen, dürfen die Bezüge zur Realsituation nicht verloren gehen. Die Beibehaltung der realen Situation der Einzelglasprüfung führt dann aber zu ungemein aufwendigen kombinatorischen Überlegungen, bei denen es schwer ist, keinen Fehler zu machen.

Beim zweiten Teil der Aufgabe führt der Versuch, die Wahrscheinlichkeit über eine Einzelglasprüfung zu berechnen, meist zu keinem Ergebnis. Günstiger ist folgende Überlegung: Wäre beispielsweise AABABBBBA die dem Experten vorgelegte Anordnung, so gibt es 16 Gläserstellungen, bei denen ein A mit einem B vertauscht ist, bzw. der Experte kann auf $\binom{4}{1} \cdot \binom{4}{1}$ Arten ein A-Glas mit einem B-Glas verwechseln und umgekehrt:

$$P(\text{genau 6 richtig} | H) = P(\text{genau 2 Fehler} | H) = \frac{16}{70}.$$

c) Bei jedem der 8 Gläser wird durch Werfen einer L-Münze entschieden, ob es durch A- oder B-Wein gefüllt wird. Er identifiziert 8 (7, 6) Gläser richtig. Berechne den P-Wert.

Das unter b beschriebene unterschiedlich interpretierbare Expertenverhalten zeigt sich auch bei der Berechnung der Wahrscheinlichkeit

$$P(\text{alle 8 richtig} | H) = \left\{ \frac{\binom{1}{2}^8}{\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \dots + \binom{8}{8}} \right\} = \frac{1}{256}.$$

Bei der ersten Lösung ist die Identifizierung der Gläser ein 8-Stufenprozeß; bei der zweiten werden alle möglichen Gläserzusammenstellungen danach eingeteilt, ob eins, zwei, drei, ..., acht Gläser mit A-Wein gefüllt sind und eine der $\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \dots + \binom{8}{8} = 2^8 = 256$ Möglichkeiten wird geraten. Die erste Lösung ist einfacher, d. h. die Pfadregeln zur Berechnung einer Wahrscheinlichkeit durch Zerlegung eines Zufallsversuchs in einzelne leicht berechenbare Ereignisse, sind hier eine sehr erfolgreiche Methode. Dies gilt für die beiden anderen Teilaufgaben nicht:

$$P(\text{genau 7 richtig} | H) = P(\text{genau ein Glas falsch} | H) = \frac{8}{256},$$

d. h. von den 256 möglichen Gläseraufstellungen liegt genau eine vor, und 8 der 256 unterscheiden sich von der Richtigen an genau einer Stelle.

Der Wechsel von den Pfadregeln bei

$$P(\text{kein Fehler} | H) = \left(\frac{1}{2}\right)^8$$

zu kombinatorischen Überlegungen bei

$$P(\text{genau 6 richtig} | H) = P(\text{genau 2 Fehler} | H) = \frac{\binom{8}{2}}{2^8}$$

setzt große Flexibilität beim Gebrauch stochastischer Methoden voraus!

d) Der Test c ist unfair. Sollte die Münze 8mal 0 oder 1 zeigen, so verliert der Experte den Vorteil des Vergleichs. Daher werden ihm mindestens zwei Gläser jeder Sorte zugebilligt, indem Münzenwürfe mit mindestens 7 Einsen oder

Nullen verworfen werden. Berechne den P-Wert bei richtiger Identifizierung von 8 (7) Gläsern.

Die Berechnung der Wahrscheinlichkeiten vereinfachen sich, wenn man die konkrete Testsituation völlig vergißt und nur danach fragt, wieviel Gläserordnungen bei dem vorgegebenen Versuchsplan möglich sind:

$$P(8 \text{ Gläser richtig} | H) = \left\{ \frac{\frac{1}{\binom{8}{2} + \binom{8}{3} + \binom{8}{4} + \binom{8}{5} + \binom{8}{6}}}{\frac{1}{256 - \binom{8}{0} - \binom{8}{1} - \binom{8}{7} - \binom{8}{8}}} \right\} = \frac{1}{238}$$

Die bei Engel angegebene Lösung der zweiten Teilaufgabe

$$P(\text{höchstens 7 richtig} | H) = \frac{9}{238} = 3,78 \% \text{ bzw.}$$

$$P(\text{genau 7 richtig} | H) = \frac{8}{238} = 3,36 \% \text{ ist nicht richtig.}$$

Bei etwas anspruchsvolleren Stochastikaufgaben wie dieser, mit vorgegebener falscher Lösung, kann man bei der Suche nach der richtigen erleben, um wieviel schwieriger Stochastikaufgaben sein können, im Vergleich zu Analysis- oder Lineare-Algebra-Aufgaben, bzw. welche stochastische Unsicherheiten bei einem selber vorhanden sind.

Ein möglicher Lösungsweg ist folgender:

$$\text{Es gibt } \binom{8}{2} + \binom{8}{3} + \binom{8}{4} + \binom{8}{5} + \binom{8}{6} = 238 \text{ verschiedene Gläserfüllungen.}$$

$$\text{Darunter sind } \binom{8}{5} + \binom{8}{6} = 56 \text{ mit genau 2 von einer Sorte}$$

$$\text{und } \binom{8}{2} + \binom{8}{3} = 182 \text{ mit mehr als 2 von jeder Sorte.}$$

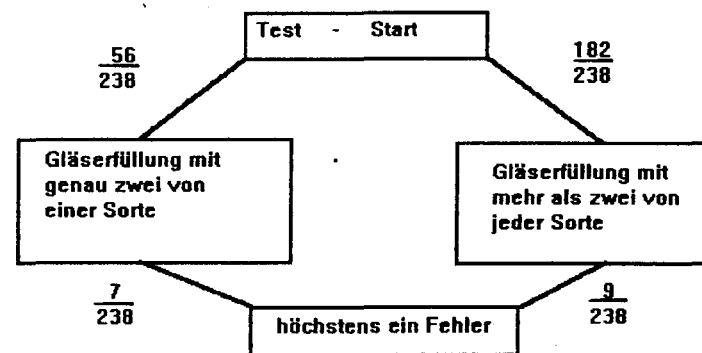
Im ersten Fall können auf 6 Arten 7 Gläser richtig erraten werden, im zweiten Fall 8. Also ist die mittlere Anzahl geratener Gläserfüllungen, die sich von der richtigen Füllung um genau ein falsch geratenes Glas unterscheiden

$$6 \cdot \frac{56}{238} + 8 \cdot \frac{182}{238} = \frac{1792}{238} = 7,53 \text{ (statt 8).}$$

$$\text{Also } P(\text{genau 7 richtig} | H) = \frac{1792}{238 \cdot 238} = \frac{7,53}{238} = 3,16\% \text{ (statt 3,36\%).}$$

Bei solchen Aufgaben ist es sehr hilfreich, kompetente Gesprächspartner zu haben. Horst Sewerin hat neben dem Training der deutschen Mathematik-Olympiademannschaft noch Zeit gefunden, mit großer Geduld diese Expertenaufgabe mit mir zu diskutieren. Er hat folgende Lösung gefunden:

$$\begin{aligned} P(\text{mindestens 7 richtig} | H) &= P(\text{höchstens ein Fehler} | H) \\ &= \frac{56}{238} \cdot \left(\frac{1}{238} + \frac{6}{238} \right) + \frac{182}{238} \cdot \left(\frac{1}{238} + \frac{8}{238} \right) \\ &= \frac{145}{4046} = 3,58\% \text{ (statt 3,78\%)} \end{aligned}$$



Auch bei dieser Lösung besteht die Schwierigkeit, den Schülern das Diagramm als Abbild der konkreten Expertenverkosterei schmackhaft zu machen.

Zu den beschriebenen Schwierigkeiten bei der Bearbeitung dieser Aufgaben kommt noch hinzu, daß der Lehrer den sehr knapp gehaltenen Aufgabentext einbetten muß in eine reale Situation, um einen Hypothesentest plausibel erscheinen zu lassen:

Zum Beispiel will eine Winzergenossenschaft Experten fürs Weinverkosten anstellen und unterzieht die Bewerber einem Einstellungseignungstest, der bestanden werden muß, um in die engere Wahl, z. B. für einen zweiten verschärften Test, zu kommen. Die Nullhypothese geht dann davon aus, daß sich auch ahnungslose bzw. völlig unfähige "Experten" bewerben. Diese Annahme, daß es nur Experten oder Laien, aber keine Zwischenstufen gibt, entspricht allerdings nicht der Realität. (Das tertium non datur des indirekten Beweises übertragen auf das Hypothesentesten erscheint in dieser Situation unlogisch.)

Man könnte versuchen, diese Schwierigkeit durch eine andere Einkleidung zu vermeiden. Zum Beispiel darf einem neugegründeten Expertenverein für Ge-

schmacksprüfungen nur beitreten, wer einen Eignungstest besteht. Da die Mitgliedschaft sehr begehrt ist, bewerben sich auch Laien. Die Wahrscheinlichkeit $\frac{1}{70}$ (Aufgabe a) bedeutet dann, daß einer von 70 Laien beim Eignungstest zufällig durch bloßes Raten keinen Fehler macht und damit in der Verein aufgenommen wird.

Welche Einkleidung man auch immer wählt, die Schwierigkeit bei der Berechnung der Wahrscheinlichkeiten völlig von der konkreten Testsituation abzusehen, bleibt bestehen.

Nur wenn die Aufgabe von allen Realitätsbezügen entkleidet so gestellt wird, daß nach der Wahrscheinlichkeit gefragt wird, eine bestimmte (A, B)-Folge zu erraten, dürfte es einfacher werden, dann aber ist dies eine kombinatorische Aufgabe und keine Fragestellung zum Testen von Hypothesen.

Anmerkung

Es wird immer wieder vorgeschlagen, auch Paradoxien im Stochastikunterricht zu behandeln. Wenn dies nur unter Berufung auf allgemeine didaktische Prinzipien geschieht und nicht aufgrund konkreter erfolgreicher Unterrichtserfahrung, kann ich - nach eigenen leidvollen Erfahrungen - nur davor warnen. Der theoretische Bildungswert steht in keinem Verhältnis zur praktischen Verwirrung und möglichen Verunsicherung beim späteren Aufgabenlösen. Ich hatte einmal das Bertrandsche Paradoxon in einem Stochastikkurs vor den Grundaufgaben der Statistik behandelt und mußte mir später bei der Expertenaufgabe zur Rechtfertigung auch völlig falscher Lösungen sagen lassen, daß die unterschiedlichen Wahrscheinlichkeiten dadurch zustande kämen, daß das "zufällige" Raten des Experten genauso wie das "zufällige" Zeichnen einer Sehne im Kreis zur Gruppe der mehrdeutigen Probleme gehöre.

Literatur

Engel, A.: Stochastik. Stuttgart, 1987.

Engel, A.: Wahrscheinlichkeitsrechnung und Statistik. Band 1. Stuttgart, 1973.