Eine Aufgabensequenz zum statistischen Hypothesentesten (2. Teil)

von Raphael Diepgen , Bochum

Didaktischer Kommentar

Aufgabe 1 führt ein in das inferenzstatistische Grundproblem, und zwar sofort aus der Sicht der statistischen Entscheidungstheorie; sie hat in diesem Angang ganz zentrale Bedeutung. Denn diese Aufgabe beinhaltet, festgemacht an einem motivierenden, weil anwendungsnahen Problem aus dem Wirtschaftsleben, explizit alle Determinanten, die überhaupt die statistische Entscheidungssituation charakterisieren. Es geht in dieser Situation um immer wieder notwendige Entscheidungen zwischen zwei statistischen Hypothesen über die Wahrscheinlichkeitsverteilung in einer unbekannten Population aufgrund von Beobachtungen in einer Zufallsstichprobe - in Aufgabe 1 um die einfachen statistischen Hypothesen "Kiste stammt aus Produktionsanlage B, d.h. $p = 0.5$" und "Kiste stammt aus Produktionsanlage A, d.h. $p = 0.25$" - . Die Suche nach einem sinnvollen Entscheidungsverfahren für solche Situationen hat zu berücksichtigen die a-priori-Wahrscheinlichkeiten beider Hypothesen, die Folgekosten der möglichen Fehlentscheidungen un die Kosten für die Erhebung der Stichprobe. Mathematisch gesehen ist damit die Suche nach einem optimalen Entscheidungsverfahren ein für die Schüler prinzipiell aus der Erwartung-mal-Nutzen-Rechnung bekanntes Optimierungsproblem: Es soll unter möglichen Verhaltensstrategien für stochastische Situationen die Strategie bestimmt werden, die "auf lange Sicht" den Nutzen maximiert, d.h. den Verlusterwartungswert minimiert.

Die Reflexion der Vorgaben von Aufgabe 1 führt ganz schnell zu Entscheidungsverfahren der Form "Prüfe n Stücke; sind mehr als k_n mangelhaft, so verkaufe die Kiste als Qualität B, sonst als Qualität A". Dabei ist aber von Anfang an deutlich, daß die Beschränkung auf nur zwei Entscheidungsalternativen und die nicht sukzessive Stichprobenziehung möglicherweise sinnvollere Strategien ausschließt; an das Bewußtsein dieser Einschränkung soll dann später die Betrachtung der sequentiellen Testverfahren anknüpfen.

Nach der Formulierung der Form solcher Entscheidungsverfahren dürfte die Berechnung der nur noch von n und k_n abhängigen Verlusterwartungswerte und deren Optimierung keine besonderen mathematischen Schwierigkeiten bereiten, da hier nur die Binomialverteilung und einfachste Regeln der Wahrscheinlichkeitsrechnung (Definition der bedingten Wahrscheinlichkeit und Additivität des Erwartungswertes) eingehen.

Ihre besondere didaktische Bedeutung hat die Aufgabe 1 vor allem darin, daß sie von Anfang an den Schülern deutlich macht, wovon der Nutzen eines statistischen Testes prinzipiell abhängt: von den a-priori-Wahrscheinlichkeiten der Hypothesen, den Fehlerkosten und den Stichprobenkosten. Diese Abhägigkeit besteht natürlich auch in den Situationen, in denen diese Parameter nicht quantifiziert werden können, in denen also die Suche nach Entscheidungsverfahren mit minimalem Verlusterwartungswert, also nach dem BAYES-Prinzip, nicht möglich ist, so daß ersatzweise nur Prinzipien der Testkonstruktion verbleiben, die in einem eingeschränkten Sinne vernünftig sind, wie etwa die signifikanzstatistische Beschränkung der bedingten Fehlerwahrscheinlichkeit erster Art. Erst das Bewußtsein dieser Abhängigkeit erlaubt eine realistische Einschätzung der Bedeutung statistischen Hypothesentestens und (damit) einen angemessenen Umgang mit ihm. Erst vor diesem Hintergrund wird etwa deutlich, daß der übliche Signifikanztest durchaus nicht notwendig eine optimale Entscheidungsstrategie bedeutet.

Diese Abhängigkeit des Nutzens eines statistischen Testes von den Problemparametern Hypothesenwahrscheinlichkeiten, Fehlerkosten und Stichprobenkosten sollte bei Aufgabe 1 durch die Variation dieser Problemparameter gegenüber dem Aufgabentext im Unterricht unterstrichen werden. Der Nutzen statistischen Testens überhaupt wird anschaulich durch den Vergleich der Verlusterwartungswerte bei einer Teststrategie gegenüber denen der a-priori-Strategien "Verkaufe auf jeden Fall als Qualität A" und "Verkaufe in jedem Falle als Qualität B", d.h. gegenüber Entscheidungsverfahren ohne Stichprobenziehung. Die gegenläufige Beziehung zwischen den Fehlerwahrscheinlichkeiten erster Art und zweiter Art wird bei der Lösung von Aufgabe 1 gleichsam von selbst deutlich, dann nämlich, wenn die Schüler bei der Suche nach einem optimalen k_n für eine vorgegebene Stichprobengröße n erleben, wie eine Veränderung dieses k_n entweder die Fehlerwahrscheinlichkeit erster Art erniedrigt , aber leider die Fehlerwahrscheinlichkeit zweiter Art erhöht oder umgekehrt. Hier erleben die Schüler bereits bei der Lösung des Einstiegsproblems die für die gesamte Teststatistik fundamentale Gegenläufigkeit beider Fehlerwahrscheinlichkeiten, die ihnen im signifikanzorientierten Angang gänzlich verborgen bleibt und im testtheoretisch orientierten Angang erst in einer späteren Unterrichtsphase - nämlich bei der Gegenüberstellung von Tests mit verschiedenen Signifikanzniveaus - im nachhinein und ohne direkten Problembezug vorgeführt wird.

Aufgabe 1 erlaubt natürlich auch schon die Diskussion weiterer Prinzipien von Entscheidungsverfahren, insbesondere des Minimax-Prinzips oder des Prinzips der Beschränkung der Fehlerwahrscheinlichkeit erster Art, also des Signifikanzprinzips. Dabei wird konkret am Beispiel die praktische Bedeutung dieser verschiedenen Prinzipien sichtbar. So führt etwa zum Signifikanzprinzip die Vorstellung, eine vorsichtige Geschäftsführung wolle angesichts von Liquiditätsproblemen des Unternehmens auf jeden Fall die Wahrscheinlichkeit eines hohen Verlustes von 10 000 DM bei fälschlichem Verkauf als Qualität A begrenzen. Klar bleibt den Schülern aber, daß "in the long run" allein die Testkonstruktion nach dem BAYES-Prinzip, das heißt die Minimierung des von Hypothesenwahrscheinlichkeiten und Fehler- wie Stichprobenkosten abhängigen Verlusterwartungswertes optimal ist.

Ein letzter Aspekt von Aufgabe 1 : Sie erlaubt auch eine Lösung mit Hilfe des den Schülern vertrauten BAYES-Theorems. Dieser Satz von BAYES gestattet die Berechnung von neuen Hypotehesenwahrscheinlichkeiten nach der Erhebung einer Stichprobe aufgrund der gewonnenen Daten, und die Verrechnung dieser a-posteriori-Hypothesenwahrscheinlichkeiten mit den Verlusten bei Verkauf als Qualität A oder B liefert dann das Entscheidungskriterium. Im Vergleich zu dieser an Bekanntes anknüpfenden Lösung über das BAYES-Theorem erscheint dann die oben skizzierte Testkonstruktion nach dem BAYES-Prinzip als Verfahrensverbesserung. Denn nur dort läßt sich die Frage beantworten, wie groß denn die Stichprobe gewählt werden soll. Das BAYES-Prinzip liefert ein Verfahren, das auch noch die Stichprobenziehung selbst variabel betrachtet, während der Satz von BAYES nur anwendbar ist nach der Ziehung einer Stichprobe.

Aufgabe 2 konfrontiert die Schüler wieder mit einem Alternatvtestproblem ähnlich wie Aufgabe 1. Der Versuch einer Lösung nach dem BAYES-Prinzip stößt aber sofort auf das Fehlen von Angaben zu den Hypothesenwahrscheinlichkeiten. Dies ist dann Anlaß für eine Diskussion, ob und gegebenenfalls über welche Überlegungen und Heuristiken eine subjektive Schätzung dieser fehlenden Hypothesenwahrscheinlichkeiten sinnvoll ist, eine Diskussion, die schließlich auch die Grundlagenproblematik eines subjektiven Wahrscheinlichkeitsbegriffes tangiert. Angesichts der Fraglich-

keit der subjektiven Schätzung der Hypothesenwahrscheinlichkeiten stellt sich dann die Frage, wie denn auch ohne Bezugnahme auf Hypothesenwahrscheinlichkeiten, aber in Berücksichtigung der angegebenen Fehlerkosten ein sinnvolles Entscheidungsverfahren konstruiert werden kann. Die am ehesten noch befriedigende Antwort auf diese Frage ist das Minimax-Prinzip. Die Durchführung der Testkonstruktion nach dem Minimax-Prinzip - mathematisch wegen der zugrundeliegenden Binomialverteilung wieder ohne besondere Schwierigkeiten - macht den Schülern wenigstens im nachhinein auch klar, warum in der Aufgabenstellung im Unterschied zu Aufgabe 1 der Stichprobenumfang fest vorgegeben und Stichprobenkosten nicht genannt wurden: Denn wegen der fehlenden Hypothesenwahrscheinlichkeiten kann das Minimax-Prinzip Stichprobenkosten nicht verrechnen und damit erst recht nicht den Stichprobenumfang optimieren. Neben der Diskussion um den Begriff subjektiver Wahrscheinlichkeit mit ihren weitreichenden grundlagen- und wissenschaftstheoretischen Implikationen ist das Ziel von Aufgabe 2 vor allem die Erarbeitung des Minimax-Prinzips, und zwar als eines gegenüber dem eigentlich optimalen BAYES-Prinzip nur "zweitbesten" Ersatzprinzips, geboren aus der Not der Unkenntnis der Hypothesenwahrscheinlichkeiten.

Erst **Aufgabe 3** bringt das typisch signifikanzstatistische Entscheidungsproblem, bei dem der herkömmliche signifikanzorientierte wie testtheoretisch orientierte Statistikunterricht beginnt, ein Entscheidungsproblem also ohne Angabe von Hypothesenwahrscheinlichkeiten, Fehler- und Stichprobenkosten. Damit scheiden für die Schüler sofort offensichtlich BAYES- und Minimax-Prinzip zur Testkonstruktion aus. Außerdem zeigt die Übersetzung der inhaltlichen Fragestellung in statistische Hypothesen den Schülern sehr schnell, daß es nun nicht mehr um die Entscheidung zwischen zwei einfachen Hypothesen geht, sondern um die

Entscheidung zwischen einer einfachen (Null-)Hypothese und ihrer Negation, das heißt einer zusammengesetzten (Alternativ-)Hypothese. Während aber im herkömmlichen Unterricht hier die Perspektive der Schüler auf den Fehler erster Art beschränkt und allenfalls im nachhinein, nämlich beim testtheoretisch orientierten Angang, auch der Fehler zweiter Art betrachtet wird, haben die Schüler in diesem entscheidungstheoretisch orientierten Angang bei Aufgabe 3 von Anfang an das Problem im Auge, wie sich sinnvoll auch für diese neue Situation ein Entscheidungsverfahren formulieren läßt, das natürlich die Fehler beider Art berücksichtigt; denn auch bei der vorangegangenen Beschäftigung mit Aufgabe 1 und 2 wurden ganz selbstverständlich beide Fehlermöglichkeiten ins Kalkül gezogen.

Dieses Problem führt die Schüler sicherlich zu dem naheliegenden Gedanken, bei der Testkonstruktion sinnvollerweise wenn schon nicht mehr explizit die Kosten von Fehlentscheidungen, so doch zumindest die - wegen fehlender Hypothesenwahrscheinlichkeiten leider nur - bedingten Wahrscheinlichkeiten der Fehlentscheidungen zu berücksichtigen, und zwar möglicherweise je nach praktischen Folgen für beide Fehlerarten mit unterschiedlicher "Gewichtung". Dies wiederum ist der natürliche Anlaß, für die Entscheidungsverfahren der einzig plausiblen Form "Werden mehr als $10 + k$ oder weniger als $10 - k$ Patienten geheilt, so verwerfe die Nullhypothese; anderenfalls behalte sie bei" für verschiedene k sowohl die bedingte Fehlerwahrscheinlichkeit erster Art, als auch die von der "wahren" Heilungswahrscheinlichkeit abhängigen bedingten Fehlerwahrscheinlichkeiten zweiter Art zu berechnen und zu diskutieren. Die Betrachtung der ß-Fehlerfunktion zeigt dabei für diese plausiblen Entscheidungsverfahren die erfreuliche Eigenschaft, daß die bedingte Fehlerwahrscheinlichkeit zweiter Art desto geringer ist, je mehr die

"wahre" Heilungswahrscheinlichkeit von den 50% der Nullhypothese abweicht; dies ist eine ganz wichtige Erkenntnis über die Natur des Signifikanztestes. Das Erlebnis der Gegenläufigkeit beider Fehlerwahrscheinlichkeiten in Aufgabe 1 und 2 hat hier nun die Konsequenz, daß den Schülern natürlich von Anfang an das Dilemma bewußt ist , nicht beide bedingten Fehlerwahrscheinlichkeiten zugleich verkleinern zu können. Naheliegend ist jetzt daher der verbleibende Signifikanzgedanke, die bedingte Fehlerwahrscheinlichkeit erster Art durch eine Grenze, das Signifikanzniveau, zu beschränken. Aber dieser Gedanke gebirt sofort ein schlechtes Gewissen, ist es doch für die Schüler nach Aufgabe 1 und 2 inzwischen ganz selbstverständlich, daß die bedingte Fehlerwahrscheinlichkeit erster Art eben nur einer unter mehreren relevanten Aspekten der statistischen Entscheidungssituation ist, und häufig nicht einmal der wichtigste. Dieses Bewußtsein führt die Schüler von selbst insbesondere zu der Frage nach der praktischen Bedeutung beider Fehlerarten in der betreffenden Situation, eine Frage, die sich in Aufgabe 3 sehr deutlich illustrieren läßt durch die Gegenüberstellung der beiden Annahmen, die Krankheit sei nur eine relativ harmlose Erkältungskrankheit oder sie sei eine bislang tödlich verlaufende Krankheit. Die Erörterung dieser Frage schließlich verweist auf für den sinnvollen Umgang mit Signifikanztests eminent wichtige Kriterien für die Wahl des Signifikanzniveaus.

Nach dieser problemorientierten Erarbeitung des Signifikanztestes sollte im Anschluß zum Vergleich auch die ß-Fehlerfunktion für einen gegenüber der ursprünglichen Aufgabenstellung erheblich vergrößerten Stichprobenumfang berechnet und diskutiert werden: Diese Betrachtung zeigt den Schülern, daß bei großen Stichprobenumfängen - also bei hoher Testgüte - mit hoher Wahrscheinlichkeit die Nullhypothese verworfen wird, auch dann, wenn die Differenz des "wahren" Parameters vom hypothetischen Parameter der Nullhypothese nur sehr klein und in praktischer oder wissenschaftlicher Hinsicht möglicherweise völlig unbedeutend ist. Konkret geht es hier am Beispiel dann um die Frage, ob ein Testverfahren sehr sinnvoll ist, wenn es mir die Nullhypothese, daß die Heilungswahrscheinlichkeit 50% betrage, mit hoher Wahrscheinlichkeit schon dann verwirft, wenn die Heilungswahrscheinlichkeit in Wirklichkeit nur unwesentlich besser oder schlechter ist. Diese Frage zeigt sehr schön die Problematik von Signifikanztests bei großen Stichproben und vor allen Dingen, daß statistische Signifikanz alleine noch nichts über praktische Bedeutsamkeit aussagt. Für einen kritischen Umgang mit Signifikanztests und vor allen Dingen mit "signifikanten Ergebnissen" erscheint mir diese anschließende Diskussion sehr wichtig.

Insgesamt dient die Plazierung des Signifikanztestes erst in Aufgabe 3 nach der entscheidungstheoretischen Behandlung der Alternativtests in Aufgabe 1 und 2 dazu, daß die Schüler durch die konsequente Berücksichtigung des Fehlers zweiter Art von Anfang an ein tieferes Verständnis der Logik des Signifikanztestes samt seiner Probleme gewinnen.

Aufgabe 4 erweitert die Anwendbarkeit des Signifikanztestes. Nachdem die Logik des Signifikanztestes in Aufgabe 3 an einer zweiseitigen Fragestellung entwickelt wurde, zielt nun Aufgabe 4 auf den einseitigen Signifikanztest. Diese Erweiterung des signifikanzstatistischen Instrumentariums geschieht am motivierenden Beispiel der "Wahlhochrechnung", also an einem Beispiel aus dem Erfahrungshorizont der allermeisten Schüler. Hier wird ein bislang unverstandenes Phänomen wenigstens ansatzweise verständlich. Auch bei dieser Aufgabe läßt sich die Wahl des Signifikanzniveaus in praktischer Hinsicht ausführlich diskutieren. Neu im Vergleich zu den vorangegangenen Aufgaben ist hier auch die Berechnung der Prüfverteilung: Die hier benutzte Approximation der

Binomialverteilung durch die Normalverteilung erlaubt die Anwendung des Binomialtestes auch für große Stichproben, wie sie vor allem im politisch interessanten Bereich der Umfrageforschung häufig vorkommen. Auch bei dieser Aufgabe ist das Studium der ß-Fehlerfunktion wieder aufschlußreich: Sie relativiert nämlich die Skepsis gegenüber dem Signifikanztest bei großer Stichprobe, wie sie sich gerade im Anschluß an Aufgabe 3 aufdrängte. Im Gegensatz dazu nämlich ist es hier bei Aufgabe 4 sogar sehr erwünscht, daß auch kleinste Differenzen des "wahren" Parameters zu den hypothetischen 50% mit hoher Wahrscheinlichkeit erkannt werden, bedeuten doch auch schon 50,1% den Gewinn der Wahl.

Aufgabe 5 hinterfragt die Selbstverständlichkeit, mit der bei der bisherigen Konstruktion der Signifikanztests der Ablehnungsbereich jeweils ans Ende der Prüfverteilung gelegt wurde. Hier ist es wichtig, daß die Schüler das intuitive Gefühl, der in Aufgabe 5 vorgeschlagene einelementige Ablehnungsbereich sei unsinnig, obwohl er der Forderung des Signifikanzniveaus genüge, in eine statistisch stringente Argumentation überführen : Dieser einelementige Ablehnungsbereich ist deshalb unvernünftig, weil er zu einer durchweg hohen bedingten Fehlerwahrscheinlichkeit zweiter Art führt. Damit holen die Schüler in dieser Aufgabe eine explizite Begründung nach für einen bislang nur plausibel gemachten Schritt der Testkonstruktion, eine Begründung, die paradigmatisch ist für die Denkweise der klassischen NEYMAN-PEARSON-Statistik.

Darüber hinaus bietet Aufgabe 5 Gelegenheit zu der Einsicht, daß es der Idee eines statistischen Entscheidungsverfahrens, das sich an bedingten Fehlerwahrscheinlichkeiten, das heißt an bedingten Wahrscheinlichkeiten für das Auftreten bestimmter Daten orientiert, grundsätzlich widerspricht, erst nach der Betrachtung der Daten das Entscheidungsverfahren festzulegen; denn die Beobachtung von Daten verändert trivialerweise die Wahrscheinlichkeit von Daten. Wählt man nämlich erst nach der Inspektion der Daten das Entscheidungsverfahren, also den Test, so ist diese Wahl selbst natürlich ein Teil des Entscheidungsverhaltens. Das Entscheidungsverhalten ist also nicht mehr durch den Signifikanztest selbst festgelegt. Die Fehlerwahrscheinlichkeit dieses Entscheidungsverhaltens ist damit nicht mehr das Signifikanzniveau, also auch nicht beschränkt. Aufgabe 5 beugt damit der weitverbreiteten Unsitte statistischer Praxis vor, in völliger Verkennung der signifikanzstatistischen Logik konkrete Tests erst nach der Inspektion der Daten festzulegen.

Aufgabe 6 schließlich löst endlich das Thema Hypothesentesten von der Beschränkung auf die Binomialverteilung, die in den Aufgaben 1 bis 5 ihrer mathematischen Einfachheit wegen die Konzentration auf die von den Besonderheiten der jeweiligen Prüfverteilung völlig unabhängige Logik des Hypothesentestens schlechthin erlaubte. Aufgabe 6a können die Schüler in Anwendung des Gelernten mit Hilfe der Binomialverteilung als Vorzeichentest selbst lösen; das Beibehalten der Nullhypothese im Vorzeichentest ist aber angesichts der angegebenen Ergebnisse wenig befriedigend, weil dabei offensichtlich nicht berücksichtigt wird, daß vor allem die Größen der Differenzen zwischen beiden Versuchsdurchgängen für die Annahme eines Effektes des Leistungsdrucks sprechen. Es schließt sich also die Suche nach einem Testverfahren an, das auch die Größe der Differenzen berücksichtigt. Nach der dabei entscheidenden Idee, von Zahlen zu ihren Rangplätzen überzugehen, führen relativ einfache kombinatorische Überlegungen zum Vorzeichenrangtest. Dieselbe Idee macht dann in Aufgabe 6b die Entwicklung des WILCOXON-Rangtestes sehr naheliegend, der ein neues Anwendungsfeld, nämlich den Vergleich von unabhängigen Stichproben, eröffnet. Schließlich könnte bei Aufgabe 6 auf die Schwierigkeit verwiesen werden, bei diesen nonparametrischen Tests den Fehler 2.Art in den

Griff zu bekommen; die relativ simple Optimierung des Fehlers 2.Art, wie sie gerade Aufgabe 5 hervorhob, ist hier nicht mehr möglich.

Insgesamt gewinnen die Schüler an Aufgabe 6 einen ersten Eindruck, daß sich Signifikanztests - unabhängig von der Binomialverteilung - für verschiedenste Fragestellungen konstruieren lassen, und daß sich auch Signifikanztests für ein und dieselbe Fragestellung hinsichtlich der Information unterscheiden, die sie dem Ergebnis entnehmen. Damit wird angedeutet erstens die Fülle von Testverfahren, die heute Interferenzstatistik für den Praktiker bedeutet, und zweitens das Problem der möglichst sinnvollen Informationsextraktion aus Daten, dem heute ein Großteil der Arbeit von Statistikern gewidmet ist.

Die Schüler sollten vor Beginn dieser entscheidungstheoretisch orientierten Unetrrichtsreihe über Hypothesentesten

- Übung haben in der elemtaren Wahrscheinlichkeitsrechnung, insbesondere im Umgang mit bedingten Wahrscheinlichkeiten,
- übliche Vorkenntnisse aus der Kombinatorik mitbringen (Anzahl von geordneten/ungeordneten Proben vom Umfang k aus einer n-elementigen Menge mit/ohne Wiederholung),
- die KOLMOGOROVsche Axiomatik gründlich kennen,
- vertraut sein mit dem BAYES-Theorem,
- Routine haben in der Handhabung der Binomialverteilung einschließlich entsprechender Tafeln,
- möglichst auch die hypergeometrische Verteilung sowie die Approximierbarkeit der Binomialverteilung durch die Normalverteilung kennen,
- vertraut sein mit der Denkweise der Erwartung-mal-Nutzen-Rechnung.

Erleichtert wird die Unterrichtsarbeit, wenn auch die Konzepte Zufallsvariable, Verteilungsfunktion und Erwartungswert vorhanden sind.

Lösungsskizze zu Aufgabe 1

Hypothesen über die Wahrscheinlichkeit p defekter Artikel:

H_0: p = 0.5 , d.h. Qualität B liegt vor.
H_1: p = 0.25 , d.h. Qualität A liegt vor.

A-priori-Wahrscheinlichkeiten der Hypothesen:

$$p(H_0) = p(H_1) = 0.5$$

Fehlentscheidungskosten

für Fehler 1. Art ("H_1" und H_0) : K_1 = 10 000 DM
für Fehler 2. Art ("H_0" und H_1) : K_2 = 1 000 DM
Prüfkosten pro Artikel: K_3 = 5 DM

A-priori-Entscheidungsverfahren ohne Prüfung:

1. Verkaufe ungeprüft als Qualität A.
   Verlusterwartungswert:

$$\begin{aligned} E(V) &= P(\text{"}H_1\text{" und } H_0) \cdot K_1 \\ &= P(H_0)\ \ P_{H_0}(\text{"}H_1\text{"}) \cdot K_1 \\ &= 0.5 \cdot 1 \cdot 10000 \\ &= 5000 \end{aligned}$$

(Dabei bedeutet "H_1" die Entscheidung für H_1, während H_0 ohne Anführungszeichen bedeutet, daß in Wirklichkeit H_0 wahr ist. Und umgekehrt.)

2. Verkaufe ungeprüft als Qualität B.
   Verlusterwartungswert:

$$\begin{aligned} E(V) &= P(\text{"}H_0\text{" und } H_1) \cdot K_2 \\ &= P(H_1) \cdot P_{H_1}(\text{"}H_0\text{"}) \cdot K_2 \\ &= 0.5 \cdot 1 \cdot 1000 \\ &= 500 \end{aligned}$$

Entscheidungsverfahren mit Prüfung, also als Hypothesentest: Prüfe eine Zufallsstichprobe von n Stücken. Falls mehr als k_n Stücke mangelhaft sind, entscheide für H_0. Sonst entscheide für H_1.

Prüfgröße ist also die Anzahl X mangelhafter Stücke in der Stichprobe. X ist annähernd B(n;p)-verteilt.

Testkonstruktion nach dem BAYES-Prinzip bedeutet nun die Bestimmung von n und k_n so, daß der Verlusterwartungswert dieser Entscheidungsstrategie minimal wird. Der Verlusterwartungswert berechnet sich als:

$$\begin{aligned} V(n;k_n) &= p(\text{"}H_1\text{" und } H_0)) \cdot K_1 + P(\text{"}H_0\text{" und } H_1) \cdot K_2 + n \cdot K_3 \\ &= P(H_0) \cdot P_{H_0}(\text{"}H_1\text{"}) \cdot K_1 + P(H_1) \cdot P_{H_1}(\text{"}H_0\text{"}) \cdot K_2 + n \cdot K_3 \\ &= 0.5 \cdot P_{H_0}(X \leq k_n) \cdot 10000 + 0.5 \cdot P_{H_1}(x > k_n) \cdot 1000 + 5n \\ &= 5000\ F_{n;0.5}(k_n) + 500\ (1 - F_{n;0.25}(k_n)) + 5n\ , \end{aligned}$$

wobei $F_{n;p}$ die kumulative Verteilungsfunktion der binomialverteilten Prüfgröße X in Abhängigkeit von n und p ist, nachzuschlagen in den entsprechenden Binomialtabellen.

Die im Unterricht vorhandenen Binomialtabellen für n = 5, 10, 15, 20, 50 und 100 ergaben als Minimum V(50;16) = 337.50 DM. Optimaler Test im Sinne des BAYES-Prinzips wäre demnach ein Test mit Stichprobengröße 50 und Annahmegrenze 16. Relativ leichte Überlegungen erlauben übrigens die Bestimmung des exakten Minimums unabhängig von der Beschränkung auf wenige Binomialtabellen mit nur bestimmten n.

Zur möglichen Lösung mit dem BAYES-Theorem:

Angenommen, man hätte eine Zufallsstichprobe von n Stücken überprüft und dabei X = d mangelhafte Stücke festgestellt. Dann folgt aus dem Satz von BAYES:

$$P_{X=d}(H_0) = \frac{P(H_0)\cdot P_{H_0}(X = d)}{P(H_0)\cdot P_{H_0}(X=d) + P(H_1)\cdot P_{H_1}(X=d)}$$

$$= \frac{0.5\ (F_{n;0.5}(d) - F_{n;0.5}(d-1))}{0.5(F_{n;0.5}(d)-F_{n;0.5}(d-1)) + 0.5(F_{n;0.25}(d)-F_{n;0.25}(d-1))}$$

Man berechne nun mit Hilfe dieser a-posteriori-Wahrscheinlichkeiten von H_0 und H_1 (ergibt sich als Ergänzung zu 1) den Verlusterwartungswert für den Verkauf als Qualität A, also $10000\cdot P_{X=d}(H_0)$, und für den Verkauf als Qualität B, also $1000\cdot P_{X=d}(H_1)$. Man entscheide so, daß der zu erwartende Verlust das Minimum dieser beiden Möglichkeiten ist.

Lösungsskizze zu Aufgabe 2

Hypothesen über den Ausschußanteil p

$$H_0 : p = 0.1$$
$$H_1 : p = 0.2$$

Keine Hypothesenwahrscheinlichkeiten; feste Stichprobengröße n = 50.

Fehlentscheidungskosten

für Fehler 1. Art: K_1 = 15 DM pro Uhr

für Fehler 2. Art: K_2 = 12 DM pro Uhr

Entscheidungsverfahren: Falls von den 50 geprüften Quarzen mehr als k die Qualitätsnorm nicht erfüllen, entscheide für H_1. Sonst entscheide für H_0.

Prüfgröße ist also die Anzahl X der nicht normgerechten Quarze in der Zufallsstichprobe. X ist angenähert B(50;p)-verteilt.

Die Testkonstruktion nach dem Minimax-Prinzip erfolgt so: Berechne für jede Grenze k jeweils den zu erwartenden

Verlust unter der Bedingung, daß die Nullhypothese gilt, also den Ausdruck

$$P_{H_0}(\text{"}H_1\text{"})\cdot K_1 = 15\ (1 - F_{50;0.1}(k)),$$

und den zu erwartenden Verlust unter der Bedingung, daß die Alternativhypothese gilt, also den Ausdruck

$$P_{H_1}(\text{"}H_0\text{"})\cdot K_2 = 12\ F_{50;0.2}(k).$$

Bestimme dann für jedes k das Maximum dieser beiden Ausdrücke. Wähle als Grenze jenes k, für welches dieses Maximum minimal ist.

Lösungsskizze zu Aufgabe 3

Hypothesen über die Heilungswahrscheinlichkeit p der neuen Therapie:

$$H_0 : p = 0.5$$
$$H_1 : p \neq 0.5$$

Keine Hypothesenwahrscheinlichkeiten, keine Fehlerkosten angegeben; Stichprobengröße n = 20 vorgegeben.

Entscheidungsverfahren (zweiseitiger Test):
Verwerfe H_0 und entscheide für H_1, falls - bezogen auf die B(20;p)-verteilte Anzahl X der in der Stichprobe geheilten Patienten - $|X - 10| > k$ gilt. Sonst behalte H_0 bei.

Konstruktionsprinzip eines Signifikanztestes ist nun die Begrenzung der bedingten Fehlerwahrscheinlichkeit 1. Art, also von $P_{H_0}(\text{"}H_1\text{"})$, durch ein Signifikanzniveau α.

Mit Hilfe der Binomialtabellen ergibt sich aus

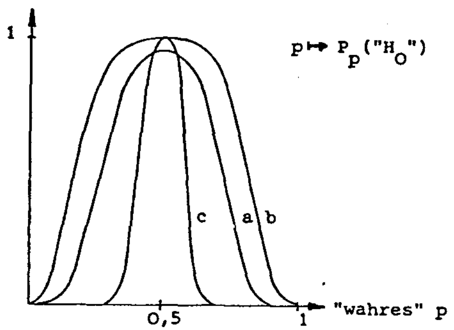
$$\begin{aligned} P_{H_0}(\text{"}H_1\text{"}) &= P_{H_0}(|X - 10| > k) \\ &= 2\, P_{H_0}(X < 10 - k) \\ &= 2\, F_{20;0.5}(10 - k - 1) \\ &\leq \alpha \end{aligned}$$

als jeweils kleinste diese Bedingung erfüllende Grenze für $\alpha = 0.05$ bzw. $\alpha = 0.01$ dann $k = 4$ bzw. $k = 6$.

Wichtig ist im Unterricht die Interpretation der zugehörigen ß-Fehlerfunktionen

$$p \longmapsto P_p(\text{"}H_0\text{"}) = 1 - 2\, F_{20;p}(10 - k - 1)).$$

Angegeben sind diese Funktionen in der folgenden Graphik unter a für den Signifikanztest zum 5%-Niveau bei $n = 20$ sowie unter b für den Test zum Niveau von 1% bei $n = 20$. Zum Vergleich ist auch die ß-Fehlerfunktion eines Signifikanztestes zum 1%-Niveau angegeben, aber bei großer Stichprobe mit $n = 200$ (unter c).



Lösungsskizze zu Aufgabe 4

Hypothesen über die Wahrscheinlichkeit p, daß ein Wähler für die Regierung votiert:

$$H_0 : p = 0.5$$
$$H_1 : p > 0.5$$

Keine Hypothesenwahrscheinlichkeiten, keine Fehlerkosten bekannt.

Vorgegebene Stichprobengröße $n = 1\,000$.

Entscheidungsverfahren (einseitiger Signifikanztest):
Falls mehr als $500 + k$ Wähler in der Zufallsstichprobe für die Regierung stimmen, verwerfe H_0 und entscheide für H_1. Sonst bleibe bei H_0.

Prüfgröße ist hier also die Anzahl X der Stimmen für die Regierung in der Stichprobe. X ist $B(1000;p)$-verteilt und daher kaum noch exakt zu berechnen. Zur Konstruktion des Signifikanztestes wird daher die Approximation durch die Standard-Normalverteilung benutzt. Denn

$$\frac{X - 1000\,p}{\sqrt{1000\,p\,(1-p)}}$$

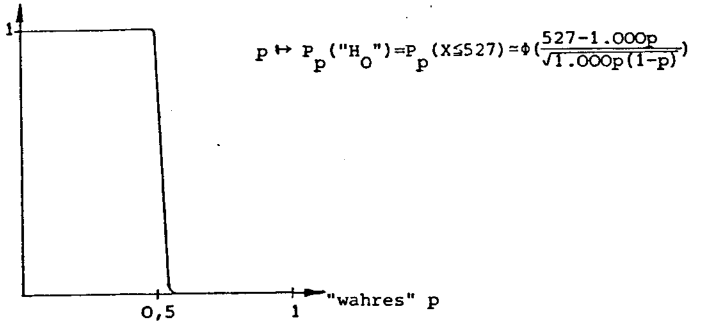
ist annähernd standard-normalverteilt. Mit Hilfe einer Tabelle über die kumulative Verteilungsfunktion Φ ergibt sich also aus

$$\begin{aligned} P_{H_0}(\text{"}H_1\text{"}) &= P_{H_0}(x > 500 + k) \\ &= P_{H_0}\left(\frac{x-500}{\sqrt{250}} > \frac{k}{\sqrt{250}}\right) \end{aligned}$$

$$= 1 - \Phi\left(\frac{k}{15.81}\right)$$

$$\leq \alpha$$

für a = 0.05 die Grenze k = 27 . Auch hier ist wieder die Inetrpretation der ß-Fehlerfunktion gerade im Vergleich zu Aufgabe 3 interessant. Ihren ungefähren Verlauf zeigt die folgende Graphik.



Lösungsskizze zu Aufgabe 5

Hypothesen über die Wahrscheinlichkeit p für das Erreichen des Lernziels:

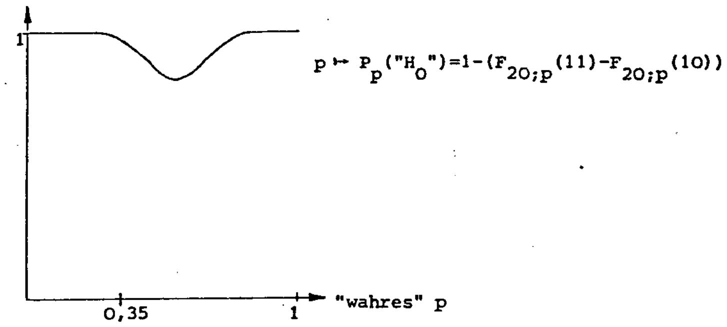
$$H_0 : p = 0.35$$
$$H_1 : p > 0.35$$

Abgesehen von der erst nachträglichen Konstruktion des Signifikanztestes - nämlich erst nach Inspektion der Daten - stellt sich hier folgendes Problem: Zwar erfüllt das vorgeschlagene Entscheidungsverfahren das Signifikanzkriterium,

denn seine bedingte Fehlerwahrscheinlichkeit 1. Art liegt tatsächlich, wie man leicht mit Hilfe einer Binomialtabelle nachrechnet, unter 5%. Denn bezogen auf die B(20;p)-verteilte Anzahl X der das Lernziel erreichenden Schüler der Stichprobe gilt:

$$\begin{aligned} P_{H_0}(\text{"}H_1\text{"}) &= P_{H_0}(X = 11) \\ &= F_{20;0.35}(11) - F_{20;0.35}(10) \\ &= 0.03359 \end{aligned}$$

Aber dieses Entscheidungsverfahren führt zu einer ungünstigen ß-Fehlerfunktion, also zu einer durchweg hohen bedingten Fehlerwahrscheinlichkeit 2. Art, wie die folgende zugehörige Graphik zeigt.



Lösungsskizze zu Aufgabe 6

Für die etwas aufwendigere Darstellung der Details sei verwiesen auf den Beitrag von STRICK (1982). Hier seien nur die Grundideen ganz grob angedeutet.

Zu (a)

Man berechne zunächst für jede der 9 Versuchspersonen die Differenz der Meßwerte aus den beiden Versuchsdurchgängen. (Es sei dabei vorausgesetzt, daß diese Differenzen - etwa wegen der hohen Präzision des Meßinstrumentes - praktisch nicht Null werden können.)

Erstens läßt sich nun ein zweiseitiger Signifikanztest anwenden über die Wahrscheinlichkeit p einer positiven Differenz mit der Nullhypothese H_0: $p = 0.5$ - also nach Muster von Aufgabe 3 -, denn bei Unwirksamkeit des Leistungsdruckes würde man Leistungsverbesserungen und -verschlechterungen im zweiten Durchgang gegenüber dem ersten für gleichwahrscheinlich halten, weil sie nur auf zufälligen Schwankungen beruhen. Dieser sogenannte Vorzeichentest berücksichtigt also lediglich das Vorzeichen der Differenzen.

Zweitens läßt sich im Vorzeichenrangtest auch Information über die Größe der Differenzen berücksichtigen: Man bringt die Beträge der Differenzen in eine Rangreihe und ordnet jedem der 9 Rangplätze das Vorzeichen der zugehörigen Differenz zu. Es gibt theoretisch $2^9 = 512$ Möglichkeiten, Plus- und Minuszeichen auf 9 Plätze zu verteilen. Hat nun der Leistungsdruck keine Auswirkung auf die Leistung, so müßte - da Leistungsveränderung nur vom Zufall abhängig - jede dieser 512 Möglichkeiten gleichwahrscheinlich sein. Diese Gleichwahrscheinlichkeitsannahme ist genau die Nullhypothese des Vorzeichenrangtestes. Für eine Wirkung des Leistungsdruckes sprechen nun alle die Verteilungen der

Vorzeichen auf die 9 Rangplätze, für die die Summe aus den Rängen mit positivem Vorzeichen entweder sehr hoch oder sehr niedrig ist, wie man sich leicht klar macht. Als Verwerfungsbereich faßt man also alle diese Verteilungen zusammen, und zwar so viele, daß die Summe ihrer Wahrscheinlichkeiten unter der Nullhypothese, jeweils 1/512 ,gerade unter dem gewählten Signifikanzniveau bleibt. Bei den angegebenen Daten in der Aufgabe wird der Vorzeichenrangtest signifikant, der Vorzeichentest jedoch nicht.

Zu (b)

Zum Rangtest nach WILCOXON führen ähnliche Überlegungen: Man bringe die Meßwerte beider Versuchsgruppen in eine gemeinsame Rangreihe und ordne jedem der $5 + 6 = 11$ Rangplätze ein A bzw. B zu, je nachdem ob der Meßwert auf diesem Rangplatz der Gruppe A oder B entstammt. Theoretisch gibt es

$$\binom{11}{5} = \binom{11}{6} = 462$$

Möglichkeiten für die Verteilung der 5 Buchstaben A und der 6 Buchstaben B auf die 11 Rangplätze. Gegen diese Annahme sprechen nun solche Verteilungen, in denen die Summe aus den Rängen mit einem zugeordneten Buchstaben B entweder sehr hoch oder sehr niedrig ist. Analog zu (a) faßt man diese Verteilungen, jede mit einer Wahrscheinlichkeit von 1/462 unter der Nullhypothese, zum Verwerfungsbereich zusammen, und zwar wieder so viele, daß ihre Gesamtwahrscheinlichkeit gerade unter dem Signifikanzniveau liegt. Auch der Rangtest wird bei den in der Aufgabe angeführten Daten signifikant, allerdings in anderer Richtung als der Vorzeichenrangtest, was im Unterricht Anlaß geben sollte zur Diskussion der offensichtlich verschiedenen Populationen, denen die Zufallsstichproben unter (a) und (b) entstammen.

Die Schwierigkeit übrigens, hier den Fehler 2.Art in den Griff zu bekommen, wird deutlich, wenn man bedenkt, daß jede denkbare Abweichung von der Gleichverteilung der 512 bzw. 462 Möglichkeiten einen solchen Fehler bedeutet. Eine solche Ungleichverteilung wird dann durch 511 bzw. 461 Wahrscheinlichkeitsparameter beschrieben. Die Wahrscheinlichkeit des Fehlers 2.Art ist damit aber Funktion von 511 bzw. 461 Veränderlichen.

Literatur

STRICK, K.H.: Parameterfreie Verfahren im Stochastik-Unterricht .- MNU 35(1982), S.138-145